

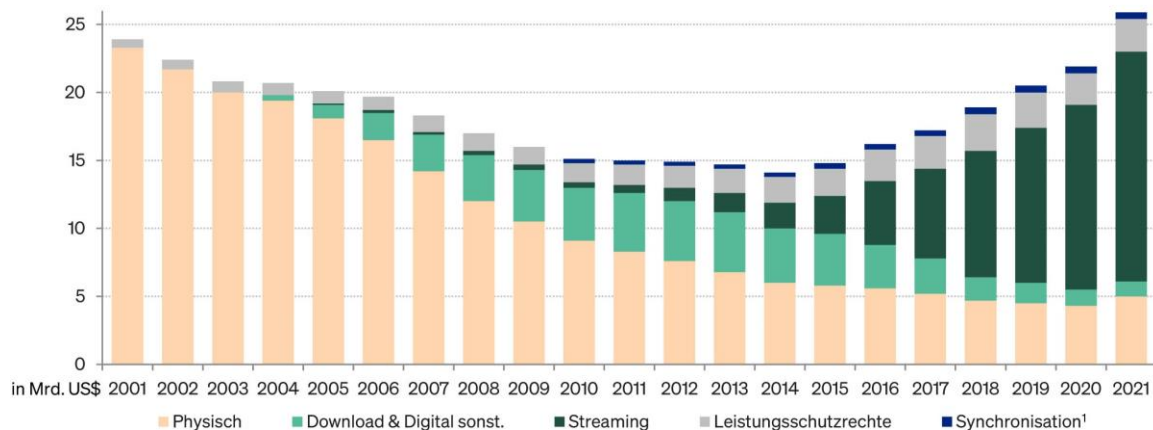
Projektarbeit Social Network Analysis
Analyse eines Spotify Artist Feature Collaboration Netzwerks

Florian Kenner – 4796728 – WWI2021F

I) Einleitung und Forschungsfrage

Allein in Deutschland lagen die Umsätze der Musikbranche im ersten Halbjahr 2022 bei 967 Mio. Euro. Der Anteil von Musik-Streaming lag bei 710 Mio. Euro, bis 2027 wird ein jährliches Umsatzwachstum von rund 3,9% erwartet¹.

Gesamtumsatz aus Verkäufen physischer Tonträger und digitaler Musik weltweit:



Quelle: IFPI Stand August 2022

Weltweit sind ähnliche Entwicklungen zu beobachten, der Anteil des Umsatzes durch digitale Produkte nimmt stark zu, wobei insbesondere Streaming eine immer größer werdende Rolle einnimmt.

Diese Arbeit widmet sich der Analyse der Kollaborationen von Künstlern der Jahre 2013 bis 2022 der Plattform Spotify. Im ersten Teil der Analyse sollen allgemeine Fragen beantwortet werden wie:

- Welche Genres sind am häufigsten in den wöchentlichen Spotify Charts vertreten?
- Welche Genres besitzen auf Spotify die größte Reichweite?
- Wie beeinflusst die Größe der Künstler, gemessen an der Anzahl der Follower die Wahl der Kollaborationen?

Dazu wurden Netzwerkdaten des Streamingdienstes Spotify ausgewertet, welche alle Songs von mehreren Künstlern der wöchentlichen Spotify Charts enthalten.

Der zweite Analyse Teil widmet sich dem Vergleich der internationalen und der deutschen Musikbranche. Generelle Leitfragen sind:

- Wie unterscheiden sich die internationale und die deutsche Musikbranche?
- Wie sind hier die größten Künstler untereinander vernetzt?

¹ Statista / Musikindustrie – Kennzahlen, August 2022

I.I) Einführung des Datensatzes

Die Datengrundlage für diese Analyse wurde von Levi Szamek erstellt. Vom 29.09.2013 bis zum 10.10.2022 wurden wöchentlich die weltweiten Spotify-Charts über kword.net aufgezeichnet. Außerdem wurden alle Kollaborationen dieser in den Charts auftretenden Künstlern hinzugefügt. Zusätzlich wurden Daten zu allen auftretenden Künstlern von Spotify selbst abgefragt und abgespeichert². Somit wurden Daten von 300386 Kollaborationen von 156422 Künstlern gesammelt.

Der Datensatz ist aus zwei csv-Dateien, den „edges“ und des „nodes“ aufgebaut. Die Edges stellen eine Kollaboration zweier Künstler dar, also ein Song. Dies ist so umgesetzt, dass jeder Künstler eine eindeutige „Spotify ID“ besitzt, über welche er identifiziert wird und mit deren Hilfe weitere Daten über den Künstler abgefragt werden können. Diese IDs sind in den Spalten „id_0“ und „id_1“ der Edges zu finden. In Verbindung mit dem nodes Datensatz kann jeder ID ein konkreter Künstler zugeordnet werden. Um dies zu ermöglichen, beinhalten die Nodes jeweils die Spotify ID und den Künstlernamen. Außerdem sind die Spotify Follower Zahlen, die Popularität, die Genres und die Chart Plazierungen der Künstler angegeben. Die Popularität ist ein von Spotify selbst gesetzter Maßstab für die Beliebtheit der Künstler. Diese werden auf einer Skala von 0-100 eingestuft, bei 100 ist die Popularität am größten. Die Einordnung der Künstler in Genres stammt auch von Spotify, dabei kann ein Künstler mehreren Genres angehören, teilweise fehlt diese Information aber auch. Die Chart Platzierungen geben die höchsten Platzierungen in den Spotify Charts nach Land an, diese Informationen stammen auch von kword.net.

Nicht enthaltene Informationen sind somit beispielsweise um welchen konkreten Song es sich bei der Kollaboration handelt, wann dieser erschien und ob und welche Chartplatzierung er erreichte.

Eine Richtung oder Gewichtung existiert nicht, darüber hinaus existiert jede Edge nur einmal, falls ein Künstlerpaar mehrfach kollaborierte lässt sich dies nicht erkennen.

Im nachfolgenden Teil wird die tatsächliche Datenanalyse durchgeführt. Hierbei wird jede Forschungsfrage nacheinander betrachtet. Für jede Forschungsfrage werden benötigte Attribute aufgezählt und im Anschluss wird das Verfahren zur Datenmanipulation beschrieben. Darauf folgt die Präsentation des Ergebnisses, sowie eine Interpretation.

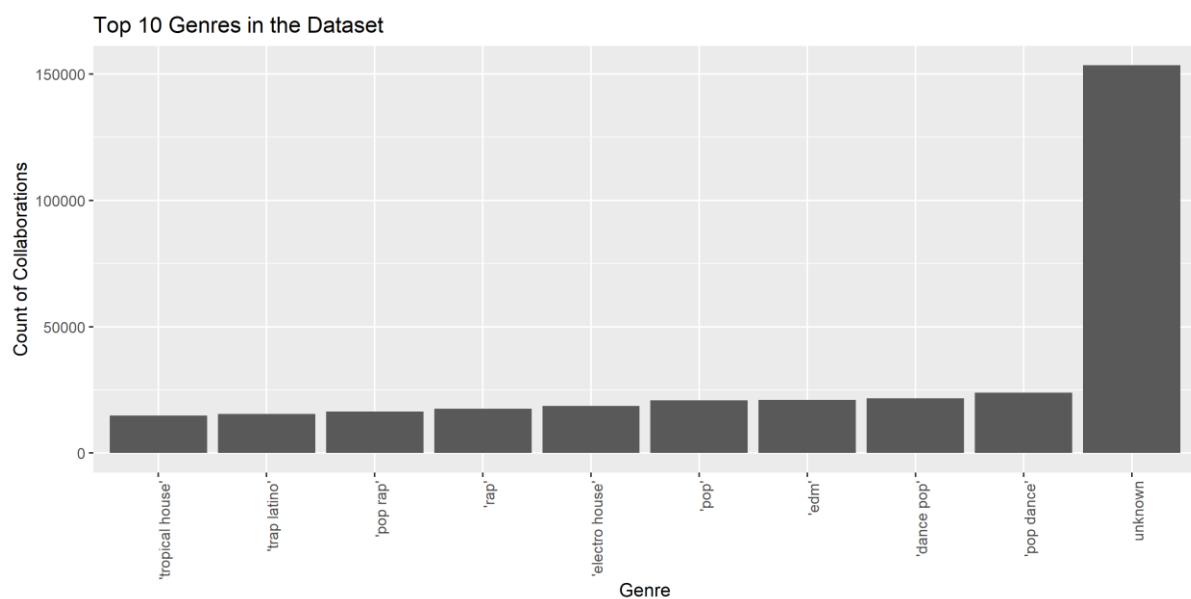
²Datensatz: <https://www.kaggle.com/datasets/jfreyberg/spotify-artist-feature-collaboration-network>, letzter Zugriff am 19.11.2022

II) Analyse der Netzwerkdaten

1.) Welche Genres sind am häufigsten in den wöchentlichen Spotify Charts vertreten?

Aus dem Kontext des Datensatzes ergibt sich, dass in den Edges nicht nur Kollaborationen enthalten sind, die es tatsächlich in die Spotify Charts geschafft haben, sondern auch alle Kooperationen der Künstler, die in den Charts waren. Somit lässt sich im Nachhinein nicht mehr nachvollziehen, welche Kollaborationen tatsächlich in den Charts waren und welche nicht. Folglich muss die Forschungsfrage leicht abgeändert werden in „Welche Genres sind am häufigsten im Datensatz enthalten?“

Hierfür wurde jeder einzelnen Kollaboration die Summe der Genres der einzelnen Künstler zugeordnet. Falls Künstler sich in gleichen Genres befinden, werden diese doppelt gewichtet. Daraufhin wurde das Auftreten jedes Genres gezählt, mit dem folgenden Ergebnis:

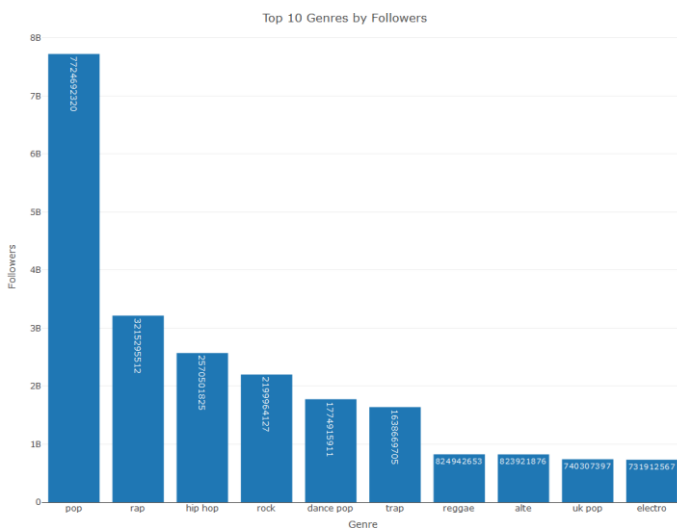


Wie zu erkennen ist, besteht der Großteil der Kollaborationen aus einem oder beiden Künstlern, welche keinem Genre zugeordnet werden können. Darauf folgen „pop dance“, „dance pop“ und „edm“. Diese fehlende Zuordnung ist auf Spotify zurückzuführen, in den Nodes ist für 103133 Künstler kein Genre angegeben. Die zweiten Plätze lassen sich sowohl durch die generelle Popularität dieser Genres als auch durch deren Ähnlichkeit erklären. Insgesamt existieren 121 Künstler, die diesen beiden Genres zugeordnet werden können. Dass die Summe allein der 10 häufigsten Genres die Anzahl der Songs weit überschreitet, ist auf die durchschnittliche Zuordnung von 4,9 Genres pro Künstler zurückzuführen.

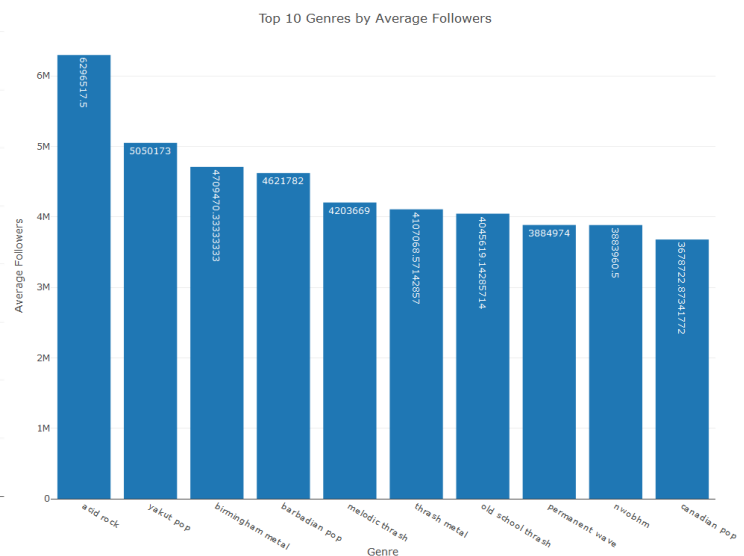
2.) Welche Genres besitzen die größte Reichweite?

Für die Beantwortung dieser Forschungsfrage wurden lediglich die Daten aus den Nodes benötigt, im Speziellen die Genres und die Follower der Künstler. Zu beachten ist, dass es sich hier um keinen vollständigen und auch keinen komplett repräsentativen Auszug handelt, was auf die Erstellung des Datensatzes zurückzuführen ist, wie unter I.I) erläutert wurde. Dennoch ist diese Berechnung aussagekräftig zwar nicht für die gesamte Musikbranche, sondern für die erfolgreichsten Künstler und deren Kollaborationen aus dem betrachteten Zeitraum.

Zur Berechnung wurde ein neues Dataframe angelegt, welches in der ersten Spalte jedes einzelne Genres und in der zweiten Spalte die Summe der Follower der Künstler, welche in dieses Genre eingestuft wurden, enthält. Dazu wurde aus den Genres der Künstler jeder einzigartige Wert genutzt, um so eine Liste aller Genres zu erhalten. Im zweiten Schritt wurde für jedes Genre nach allen Künstlern gefiltert, die in dieses fallen und dann die Summe der Follower dieser berechnet, mit dem folgenden Ergebnis:



Genres nach absoluten Follower Zahlen



Genres nach durchschnittlichen Follower Zahlen

Um dieses Ergebnis besser einordnen zu können, ist es wichtig zu berücksichtigen, dass keine Informationen über den einzelnen Follower vorliegen. Dieser kann mehreren Künstlern aus einem Genre folgen und wird somit mehrfach in die Reichweite dieses Genre eingerechnet. Eine alternative und aussagekräftigere Herangehensweise ist nicht die Summe aller Follower eines Genres zu bilden, sondern die durchschnittliche Anzahl der Follower der Künstler dieses Genres. Im Vergleich dieser Berechnungsweisen fällt auf, dass ein komplett anderes Ergebnis erscheint. Ursache muss sein, dass den Genres in der rechten Grafik deutlich weniger aber im Schnitt größere Künstler angehören. Auffällig ist auch, dass die Unterschiede der absoluten Follower Zahlen deutlich größer sind als die der relativen. Wie sich nun die tatsächliche Reichweite der Genres berechnen lässt bleibt eine offene Frage.

3.) Wie beeinflusst die Größe der Künstler, gemessen an der Anzahl der Follower die Wahl der Kollaborationen?

Diese Forschungsfrage zielt darauf ab, ob gemessen an der Anzahl der Follower beispielsweise große Künstler Präferenzen für die Anzahl der Follower bei der Wahl des Kollaborations-partners haben. Diese Frage lässt sich leicht durch die Korrelation der Follower Zahlen beantworten. Hierfür wurde in den Edges für jede Kollaboration die Anzahl der Follower der beiden Künstler in eine neue Spalte eingefügt. Im nächsten Schritt wurde die Korrelation dieser Spalten berechnet mit einem Ergebnis von **0,12**. Da dieses nicht signifikant von 0 abweicht, lässt sich interpretieren, dass es kaum einen Zusammenhang zwischen den Reichweiten der Künstler in diesem Datensatz gibt.

II.I) Vergleich der deutschen und der internationalen Musikszene

Dieser Teil widmet sich nun dem Vergleich der deutschen und der internationalen Musikszene, mit dem Ziel, Unterschiede und Gemeinsamkeiten herauszuarbeiten. Wichtig hier ist es zu beachten, dass die deutsche Musikszene ein Teil der internationalen Musikszene wird. Im Folgenden werden lediglich die größten Künstler (wieder gemessen an der Anzahl der Follower) der beiden Szenen betrachtet. Dies geschieht aus den folgenden Gründen: einerseits sind sowohl die deutsche Szene als auch die internationale Szene so groß, dass kaum differenzierte Aussagen getroffen werden können. Andererseits werden so die deutschen Künstler nicht in der internationalen Szene betrachtet, da kein deutscher Künstler zu den weltweit größten gehört. Die Einstufung als deutsch erfolgt durch die Genres der Künstler, falls diese die Schlagworte „deutsch“ oder „german“ beinhalten wurde ein Künstler als deutsch eingeordnet.

4.) Wie unterscheiden sich die internationale und die deutsche Musikbranche?

Um dieser Frage nachzugehen wurden unterschiedliche Kennzahlen berechnet. Diese sind die durchschnittliche Anzahl der Genres pro Künstler, die durchschnittliche Anzahl der Follower, Popularität, sowie die Anzahl der Kollaborationen im Betrachtungszeitraum.

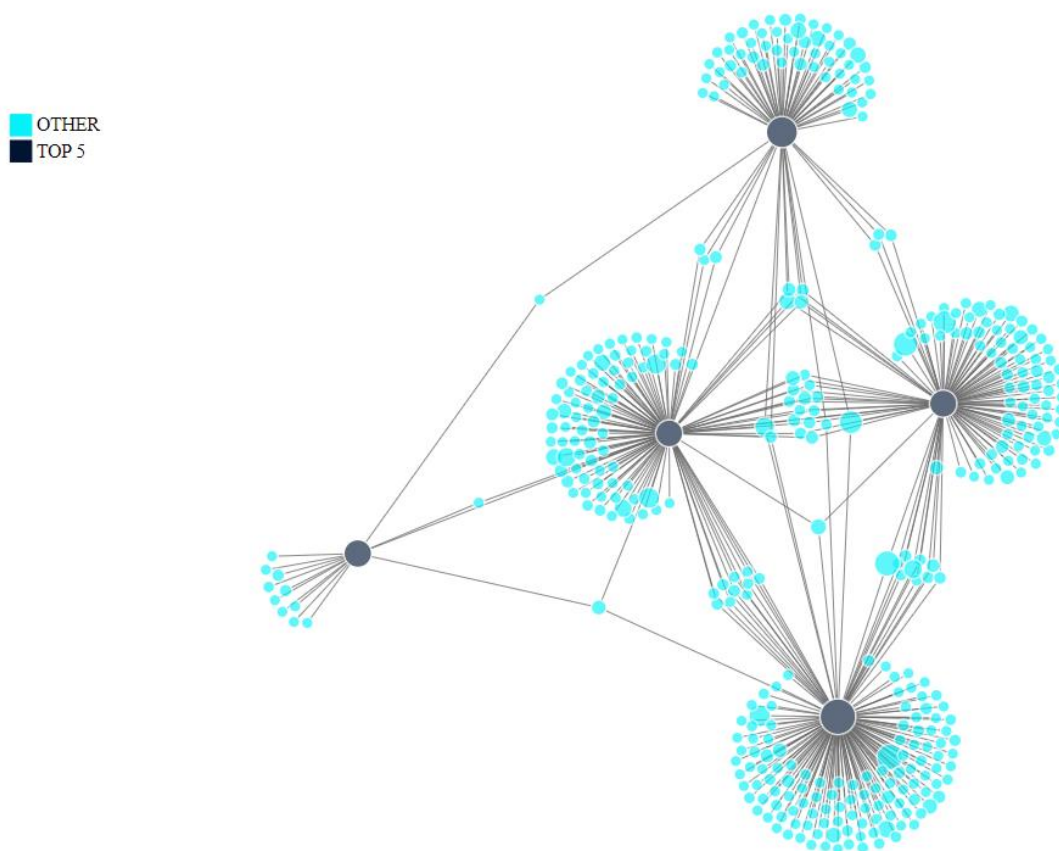
	international	deutsch
Follower	77.210.356	5.463.776
Popularität	90,2	73,6
Genres	2,8	6
Kollaborationen	457	197

Die unterschiedlichen Follower Zahlen waren zu erwarten, interessant hierbei ist der Faktor 14, um welchen die größten Künstler mehr Follower haben als die größten deutschen Künstler. Ob hier ein Zusammenhang mit dem Anteil des Umsatzes der

deutschen Musik-Streaming Branche an der gesamten Musik-Streaming Branche besteht, benötigt weiterer Untersuchungen. Außerdem sticht heraus, dass die Popularität der deutschen Künstler nicht direkt in Relation zu den Followern steht. Aus den Daten aller Künstler geht hervor, dass Follower und Popularität nur schwach mit einem Koeffizienten von rund **0,23** korrelieren. Informationen, wie die Popularität von Spotify festgelegt wird sind nicht gegeben. Darüber hinaus gehören die deutschen Künstler im Schnitt mehr als doppelt so vielen Genres an, wie die internationalen Künstler, wobei die internationalen Künstler über doppelt so viele Kollaborationen veröffentlichten.

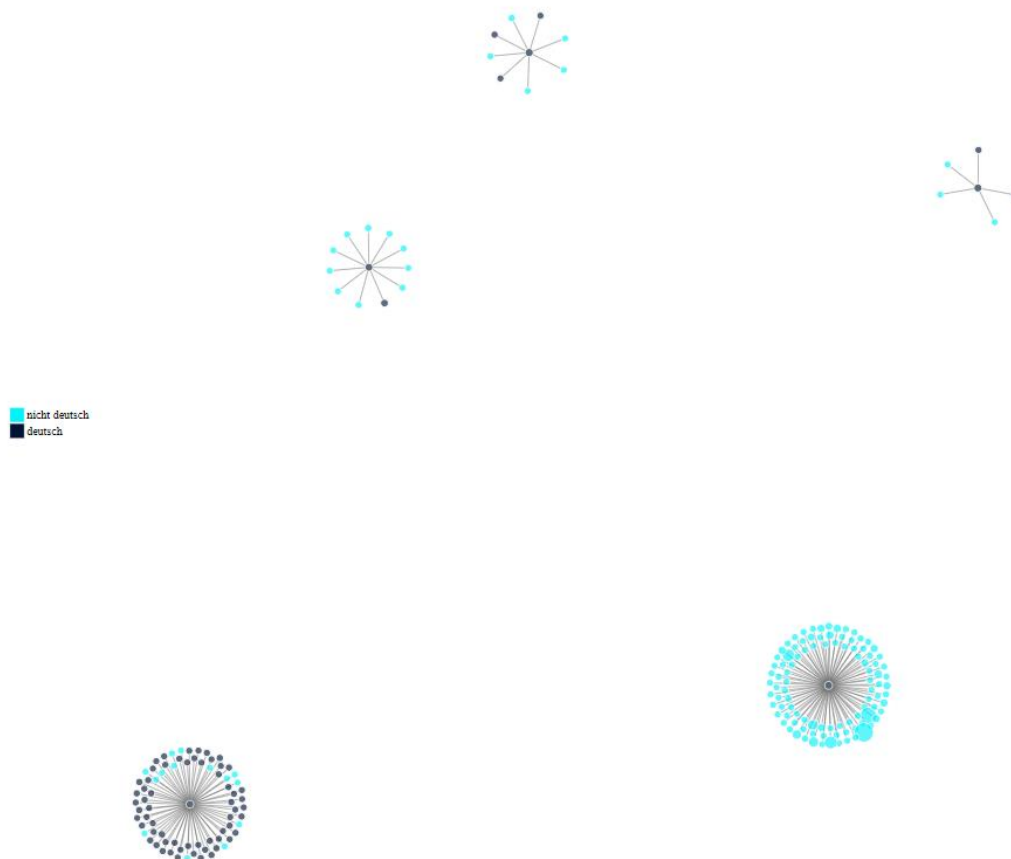
5.) Wie sind hier die größten Künstler untereinander vernetzt?

Wie bereits bei der vorigen Frage soll hier der Fokus auf dem Vergleich der deutschen Szene und der internationalen Szene liegen. Dazu wurden zuerst Netzwerk Visualisierungen jeweils der fünf größten Künstler und deren Features erstellt. Um visuell leichter unterscheiden zu können, welches die ursprünglichen Künstler und welches die Features sind wurde eine weitere Spalte in den Nodes hinzugefügt, welche aussagt, ob ein Künstler zu den fünf größten gehört. Mithilfe der Bibliothek NetworkD3 wurden die Visualisierungen erstellt. Die Größe der Knoten ist abhängig von der Follower Zahl



Netzwerk der fünf international größten Künstler

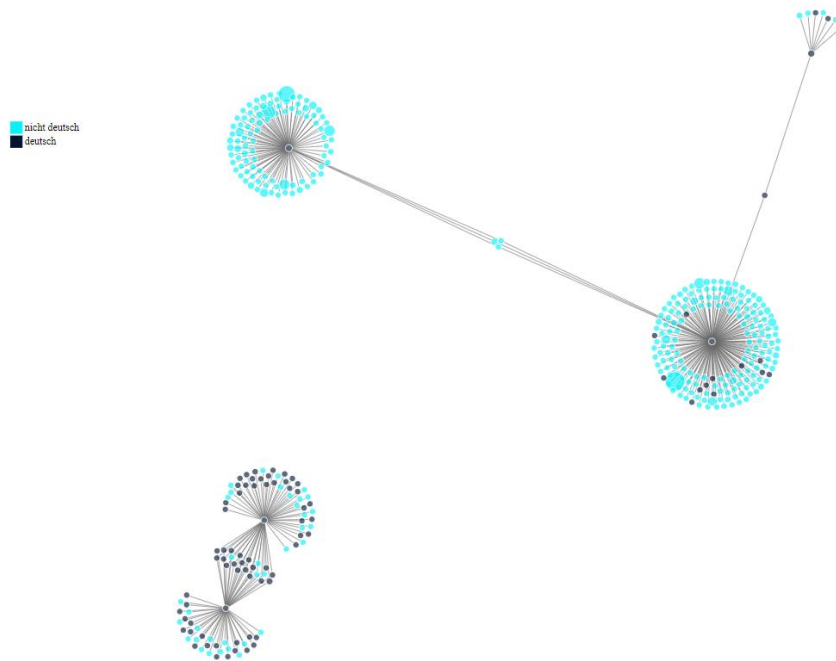
Wie zu erkennen ist, bilden diese Künstler ein einziges großes Netzwerk, es existiert nur eine Komponente. Außerdem existieren auch direkte Kollaborationen zwischen den Top 5 Künstlern. Am zentralsten ist hier Justin Bieber, welcher Kollaborationen mit allen anderen vier Künstlern hatte. Darüber hinaus hat jeder Top 5 Künstler eine Gruppe von Künstlern, mit denen lediglich er arbeitete. Außerdem entstehen Schnittmengen von zwei oder drei Künstlern. Sowohl die geringste Zentralität im Sinne des Grades als auch die wenigsten Kollaborationen hat Billie Eilish. Eine mögliche Erklärung ist, dass ihr erster Song im Jahr 2016 veröffentlicht wurde und somit weit nach denen der anderen Top 5 Künstler.



Netzwerk der fünf größten deutschen Künstler

Die deutsche Szene stellt einen sehr starken Kontrast dar. Das Netzwerk besteht aus 5 Komponenten, keine einzige Kollaboration der deutschen Top 5 Künstler existiert. Eine mögliche Erklärung, direkt aus dem Datensatz selbst ist die deutlich höhere Anzahl der Genres der einzelnen Künstler. Die Zahl der Kollaborationen variiert stark zwischen den einzelnen Künstlern, außerdem ist bemerkenswert, dass der Künstler Capital Bra überwiegend mit anderen deutschen Künstlern kollaboriert, wohingegen der Künstler Zedd ausschließlich mit nicht-deutschen Künstlern Songs veröffentlicht hat. Diese Auswahl der Künstler enthält als einen Top 5 Künstler Ludwig van Beethoven, welcher lange vor den anderen Künstlern lebte, somit ist infrage zu stellen, wie sinnvoll ein Vergleich dieses Datensatzes, mit dem der internationalen Szene ist.

Daher wurde eine weitere Visualisierung erstellt, in dieser ist das ausschlaggebende Kriterium für die Auswahl der deutschen Künstler anstatt der Follower Zahl die Popularität, mit dem folgenden Ergebnis.



Netzwerk der fünf populärsten deutschen Künstler

Aus diesen Visualisierungen geht hervor, dass sich die deutsche Musikszene auch bei unterschiedlichen Auswahlkriterien stark von der internationalen Szene unterscheidet. Die Vernetzung ist deutlich geringer. Diese Veränderung der Auswahl der Kriterien hat zur Folge, dass die Künstler Ludwig van Beethoven, Scorpions und Capital Bra durch die Künstler Robin Schulz, RAF Camora und Bonez MC ersetzt wurden. Dadurch besteht das Netzwerk nur noch aus zwei Komponenten. Die Künstler Robin Schulz und Zedd sind international stark vernetzt und teilen sich drei Kollaborationspartner. Die Künstler RAF Camora und Bonez MC, welche Teil des Deutsch Rap Genres sind bilden die zweite Komponente. Beide haben Features sowohl mit deutschen als auch mit nicht deutschen Künstlern und teilen sich eine große Schnittmenge. Die Künstler Zedd und Robin Schulz, welche ausschließlich oder überwiegend mit nicht deutschen Künstlern zusammenarbeiten stellen eine Verbindung zu den weltweiten Top 5 Künstlern dar.

III) Fazit

Zusammenfassend konnte der erste Analyse Teil beantwortet werden. Sowohl die größten als auch die am häufigsten vertretenen Genres im Datensatz wurden bestimmt. Hier entsteht neuer Raum für Recherche über diesen Datensatz hinaus, sind die hier gezeigten Ergebnisse repräsentativ für die gesamte Streaming-Branche? Aus den Analysen entstand auch die Frage, was eine sinnvolle Definition für die Reichweite eines Genres in diesem Kontext ist, und wie sich diese berechnen lässt. Sind Unterschiede von Plattform zu Plattform, also beispielsweise zwischen Spotify und Apple Music festzustellen? Auch die Frage, ob große Künstler tendenziell eher mit anderen großen Künstlern zusammenarbeiten, konnte erfolgreich mit nein beantwortet werden.

Im zweiten Analyseteil mussten im Vorhinein einige Einschränkungen gemacht werden, um differenzierte Aussagen treffen zu können. Es wurden nur Ausschnitte der deutschen und der internationalen Musikszene betrachtet, welche dennoch für sich aussagekräftig sind. Es wurde erfolgreich dargestellt, wo die Hauptunterschiede liegen. Offene Fragen sind hier beispielsweise, wie Spotify die Popularität von Künstlern berechnet. Die deutsche Szene unterscheidet sich stark von der internationalen, vor allem ausgezeichnet durch eine schwächere Vernetzung der größten Künstler.

Anmerkungen:

Die Visualisierungen der beiden Netzwerke befinden sich als interaktive html Datei in der Abgabe

In dieser Arbeit wurde nicht gegendert, mit der Wortwahl des generischen Maskulinums sind alle Geschlechter gemeint und niemand soll diskriminiert werden

Der Code-Block in der .Rmd Datei zur Beantwortung der 2. Forschungsfrage benötigt zum Ausführen auf meiner Hardware ca. 4h, funktioniert aber.

Das Ausführen von worker.Rmd erstellt alle Grafiken aus dem Analyse-Teil und speichert diese als .png oder .html Dateien ab.