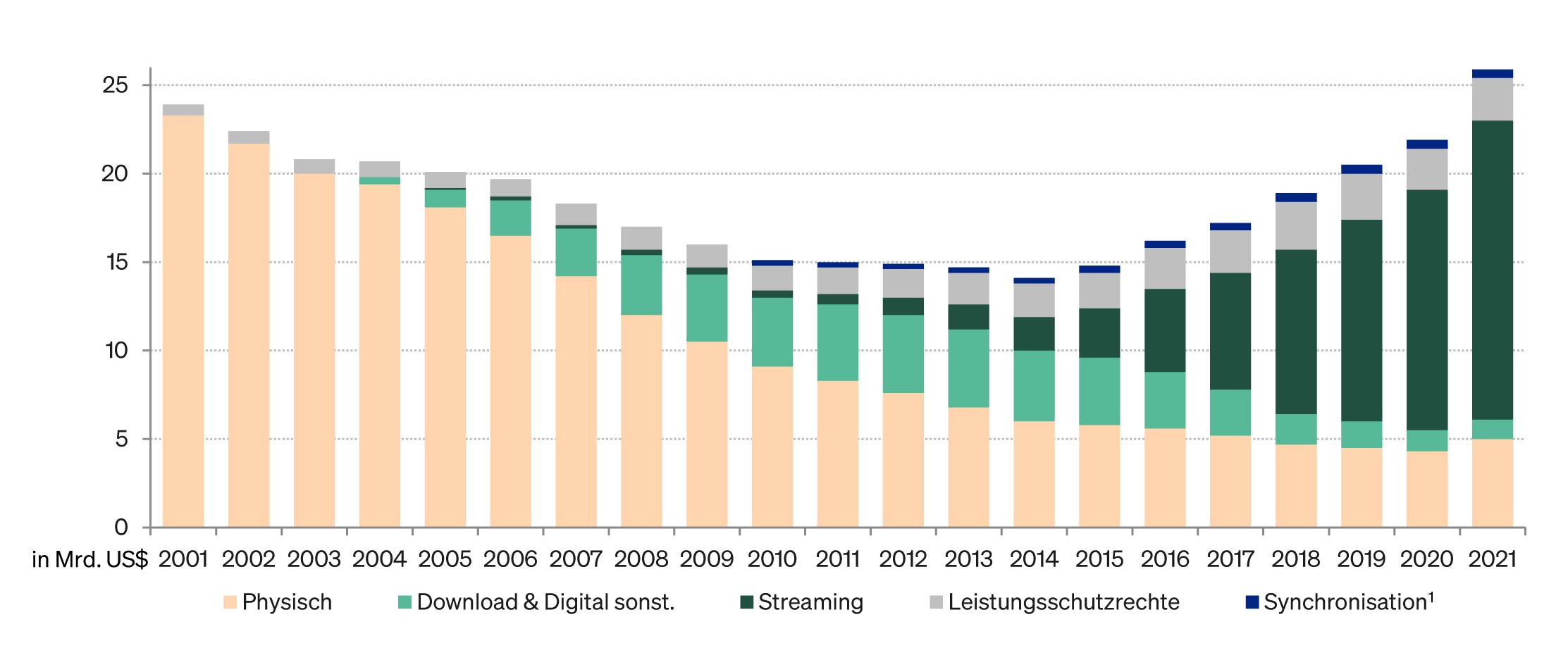
**Projektarbeit Social Network Analysis**

**Analyse eines Spotify Artist Feature Collaboration Netzwerks**

Florian Kenner – 4796728 – WWI2021F

1. **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% erwartet[[1]](#footnote-1).

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 größere Rolle einnimmt.

Diese Arbeit widmet sich der Analyse der Kollaborationen von Künstlern der Jahre 2013 bis 2022. 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 Größte 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 ausgewertet.

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?

**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 kworb.net aufgezeichnet. Außerdem wurden alle Kollaborationen dieseer in den Charts auftretenden Künstlern hinzugefügt. Zusätzlich wurden Daten zu allen auftretenden Künstlern von Spotify selbst abgefragt und abgespeichert[[2]](#footnote-2). 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 mithilfe welcher 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 können jeder ID konkrete 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 Einordung der Künstler in Genres stammt auch von Spotify, dabei kann ein Künstler mehreren Genres zugehören, 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 kworb.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.

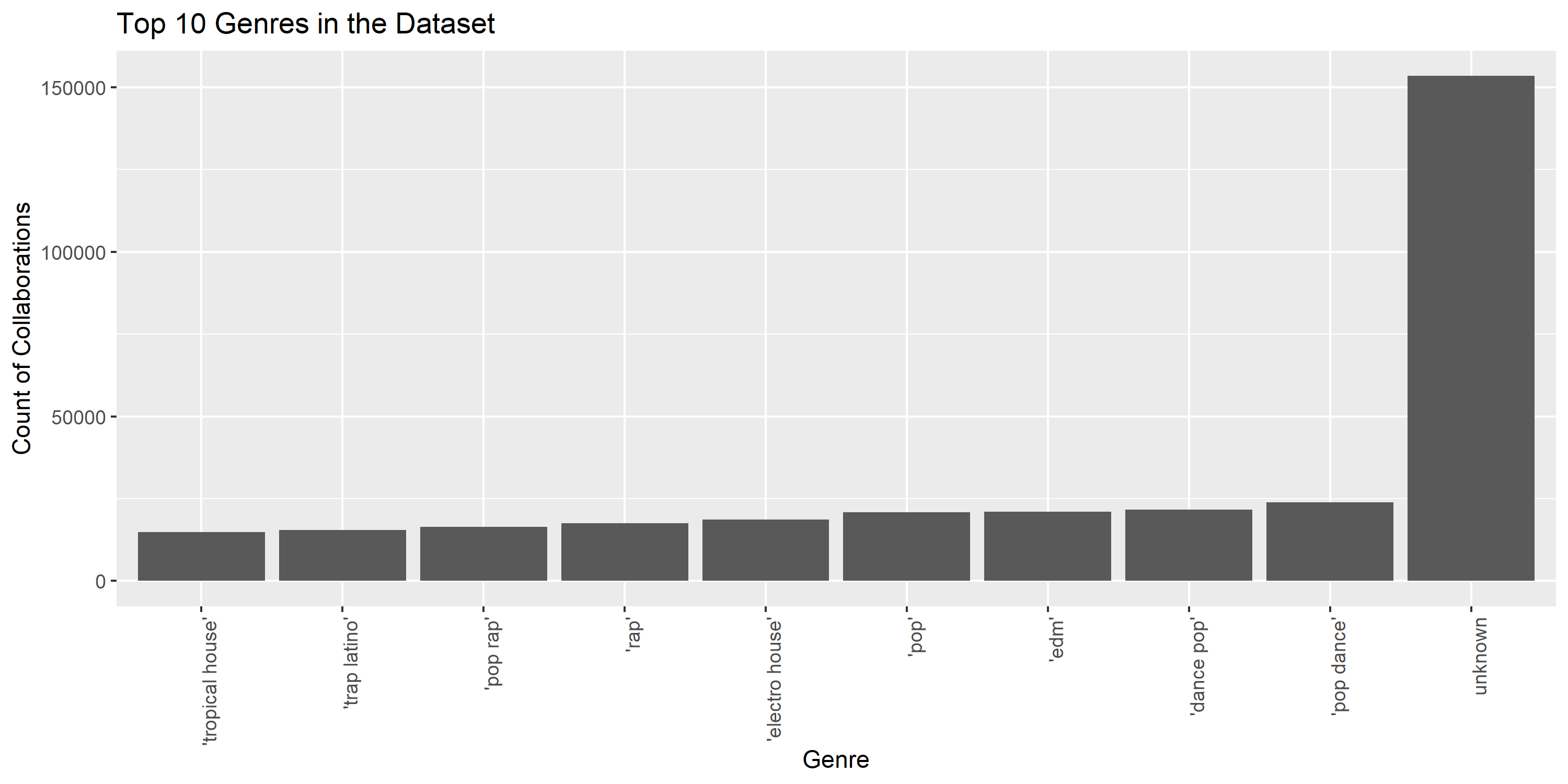
Im nachfolgenden Teil wird die tatsächliche Datenanalyse durchgeführt. Hierbei wird jede Forschungsfrage nacheinander betrachtet. Für jede Forschungsfrage werden benötigte Attribute aufgezählt und im Anschluss wir das Verfahren zur Datenmanipulation beschrieben. Darauf folgt die Präsentation des Ergebnisses,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

1. **Analyse der Netzwerkdaten**

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 der Großteil der Kollaborationen mit 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

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, was auf Erstellung des Datensatzes zurückzuführen ist, welche 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 einzelnes Genre 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:

Um dieses Ergebnis besser einzuordnen zu können ist 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.

Wie beeinflusst die Größte 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 Followerzahlen 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.

1. Statista / Musikindustrie – Kennzahlen, August 2022 [↑](#footnote-ref-1)
2. Datensatz: <https://www.kaggle.com/datasets/jfreyberg/spotify-artist-feature-collaboration-network>, letzter Zugriff am 19.11.2022 [↑](#footnote-ref-2)