A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

Examining Bilateral Music Flows in the Digital Age:

The Role of Distance in International Music Demand

SUNNIVA GIULIA HERMANN

Work project carried out under the supervision of:

Qiwei Han

16-12-2022

Table of Contents

Figures

Tables

Appendices

I. **Abstract**

Examining Bilateral Music Flows in the Digital Age: The Role of Distance in International Music Demand

This study explores the relationship between international music demand and consumer preferences in the music streaming era. An updated distance measure between countries to account for their music flow is proposed, considering the impact of digitization. The analysis shows that while traditional distance measures such as geographical proximity and shared language remain essential, cultural, and social distance are also significant. A composite measure is suggested to be more appropriate for accounting for consumer preferences in music streaming. The study provides insights into how the rise of digital music streaming has changed the pattern of international music demand.

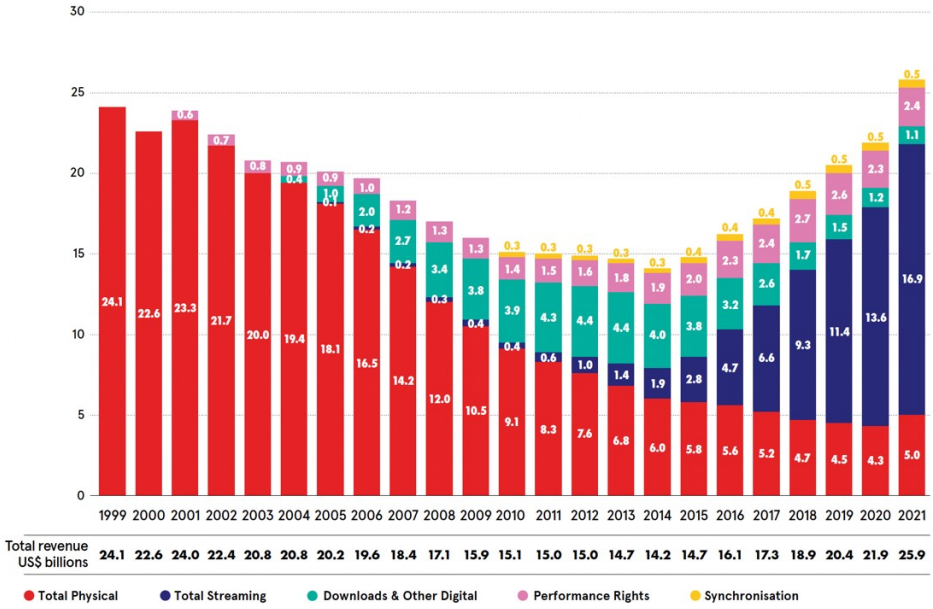
Keywords: Music Streaming, International Music Demand, Consumer Preferences, Spotify, Bilateral Music Flow, Gravity Model

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

Digitization is arguably the most significant change the music market has seen in recent decades. Streaming is the dominant revenue format in the global music market and accounted for 65.0% of music revenues in 2021, with 47.3% coming from paid subscriptions. In total, there are around 523 million users of paid subscription accounts on music streaming services such as Spotify, Apple Music, and others (IFPI 2022). Figure 1 shows global music industry revenues from 1999 to 2021, illustrating how dramatically fast the lucrative transition from analog to digital music distribution has occurred. As a result, online streaming platforms have become the most important method of music consumption (Soares Araujo, Pinheiro De Cristo, and Giusti 2019).

Figure 1: The global recorded music industry's revenues from 1999-2021 (in US$ Billions). (IFPI 2022)



The widespread adoption of digital technologies in the music industry has significantly impacted how music is produced, distributed, and consumed. This shift has reduced the costs associated with these processes, making it more accessible and cost-effective for creators and consumers (Waldfogel 2017; Graham et al. 2004). With the rise of services like Spotify and Apple Music, artists can easily upload their music and make it available to a global audience without incurring the significant costs associated with physical distribution. This digital format

allows for easy duplication and distribution at almost zero marginal cost, facilitating the global dissemination of music. This has expanded the range of music available to consumers and significantly increased their choice. In addition, streaming platforms have further reduced the cost of accessing music, allowing users to listen to millions of songs worldwide for a flat monthly fee. The discovery of new music from different countries is thus facilitated, and it is easier for individuals to enjoy music in a flexible, personalized, and cost-effective way. The digitization of music has significantly lowered consumer obstacles to accessing and consuming music worldwide. Barriers to the internationalization of music have been removed, making it easier for musicians and consumers to connect (Ferreira and Waldfogel 2013; Way, Garcia-Gathright, and Cramer 2020; Datta, Knox, and Bronnenberg 2018).

Despite the vast amount of music available on streaming services at no cost, consumers continue to exhibit preferences for specific songs and artists (Prior 2013). In this study, we explore the relationship between international music demand and consumer preferences at the country level, treating international music demand as a form of international trade. Previous research has used the gravity model of international trade to quantify bilateral music flows, considering geographical proximity and shared language. However, the rise of music streaming has made it possible to access music from anywhere in the world at zero cost, leading us to question whether traditional distance measures are still relevant (Ferreira and Waldfogel 2013). In particular, the increasing use of online platforms for music consumption may have introduced new factors, such as social connectedness, influencing the music flow between countries (Kim, Suh, and Lee 2014; Araujo, Cristo, and Giusti 2017; Dewan, Ho, and Ramaprasad 2017). Our analysis builds on previous research by proposing an updated distance measure that considers the impact of digitization on the music market. This paper aims to use a composite distance measure to account for consumer preferences in music streaming. We aim to determine whether traditional measures, such as geographical proximity and shared language, are still relevant in

the age of digital music streaming or whether other factors, such as social, cultural, or economic factors, have become more important in describing the pattern of international music demand. We investigate the moderating effects of the different distance measures on bilateral music flows and hypothesize that the rise of digital music streaming has changed the importance of individual measures. Furthermore, we examine how distance measures are affected by the COVID-19 pandemic.

To investigate these questions further, we analyze Spotify's daily top 200 songs from 96 countries from 2017 to 2021. Our analysis uses a panel data gravity model with dyadic country-pair data. The results show that while geographic distance and shared language remain essential factors in quantifying bilateral music flows, cultural and social distance are also significant independent variables. Therefore, a composite distance measure is more appropriate for accounting for consumer preferences in music streaming. This added significance may be due to the increasing cultural divergence of popular music and the influence of social networks on music demand (Bello and Garcia 2020; Araujo, Cristo, and Giusti 2017; Kim, Suh, and Lee 2014). In addition, the results show that the COVID-19 pandemic has strengthened the effect of all distance measures on bilateral music flows. As a result, people are less inclined to listen to music from other parts of the world with which they do not feel connected (Yeung 2020).

Our research provides a deeper understanding of the trends and patterns in the international music trade within the global streaming market. Spotify can use these results to develop strategies that promote the music flow between countries that are not as connected and encourage listeners to discover music from more diverse and distant parts of the world.

The structure of this thesis consists of seven sections. The introduction is followed by a review of the relevant literature and the development of the hypotheses. Subsequently, a description of the data sources used for the empirical analysis is given. The fourth section provides an overview of the research methodology and explains the model specification. Then, the analysis

results are presented, and the hypotheses are tested. Next, two robustness tests are performed to assess the validity of the results. Finally, the findings are discussed regarding their implications, limitations, and prospects.

2. Literature Review

In this section, we establish the theoretical foundation for our analysis that guides our research and provides a basis for interpreting our findings. We review previous relevant research and derive hypotheses tested in our analysis.

2.1. Relevant Research Areas

Two areas of research will be explored in more detail. First, we will investigate the factors that determine song popularity in general. Second, we will explore how music streaming has affected the music flow between countries and changed how music is consumed. By examining these two research areas, we aim to provide a more nuanced understanding of music consumption and the spatial dimensions of bilateral music flow in the era of music streaming.

The scientific literature on predicting song popularity in music streaming is extensive, with many studies using machine learning methods (Soares Araujo, Pinheiro De Cristo, and Giusti 2019). Table 1 illustrates the conceptual breadth of this research by providing an overview of the critical determinants affecting song popularity. However, it represents only a fraction of the literature on this topic. In this paper, we account for a song's popularity by looking at its number of streams and rankings on Spotify.

Table 1. Determinants discussed in the literature that influence Song Popularity.

Determinant	Influence on Song Popularity	Author(s)
Social Networks	The results indicate that the number of positive tweets about a song and artist can effectively predict the song's popularity.	Kim, Suh, and Lee 2014; Araujo, Cristo, and Giusti 2017
Acoustic features of the song	The song's popularity dynamics correlate with its acoustic characteristics, such as danceability or happiness.	Georgieva, Suta, and Burton 2018; Interiano et al. 2018
Algorithmic Recommendations	The coherence of the songs a user listens to can be affected by algorithmic recommendations.	Anderson et al. 2020
Length of a song	To create an attention-grabbing song that will be played on the radio, it should have a certain length.	McKinney Kelsey 2015
Spotify's playlists inclusion decisions	Including songs in the top 100 charts increase song discovery by 11-13%.	Sim, Park, et al. 2022

While previous research has identified several factors that may influence a song's popularity, our study focuses on the spatial dimensions of music flows between countries and their impact on bilateral music consumption. Before the widespread adoption of music streaming, music flow was primarily limited to physical formats, such as CDs and vinyl records. These physical formats were only available in the countries where they were produced and had to be transported to other countries to be distributed there. That made it difficult for music to flow freely between countries, and it often took a long time for music to reach audiences in other parts of the world (Alexander 1994; Graham et al. 2004).

In contrast, digital music streaming has made it much easier for music to flow between countries. Anyone with an internet connection can easily access and listen to music, regardless of location. Music streaming has removed many barriers that previously limited the music flow, allowing music to be more easily shared and discovered by audiences in other parts of the world (Ferreira and Waldfogel 2013; Way, Garcia-Gathright, and Cramer 2020; Datta, Knox, and Bronnenberg 2018). Research has shown that the widespread adoption of music streaming has led to increased music flow between countries. In their analysis, Datta, Knox, and Bronnenberg (2018) found that streaming has increased overall music consumption and greater musical diversity. It has facilitated the globalization of the music industry, as music from different countries and cultures can be more easily shared and discovered by listeners worldwide.

Leading to a more diverse range of music available to listeners, it allowed musicians from different parts of the world to gain exposure to audiences in other countries. In addition, to facilitate the music flow between countries, digital music streaming has also changed how music is consumed. With streaming services, people can listen to a broader range of music from different countries and cultures and discover new music more easily. This has likely increased the flow between countries, as people are more willing to listen to and share music from other parts of the world (Gomez-Herrera, Martens, and Waldfogel 2014; Datta, Knox, and Bronnenberg 2018; Waldfogel 2017).

The effects of globalization on the music industry have been a topic of debate in the literature, with some arguing that it has overall positive effects and others arguing it has adverse effects. On the positive side, globalization can expand consumers' choices, introduce artists to new audiences, and inspire new styles of music (Cowen 2009; Magu 2015). However, it can also lead to wealthier countries dominating global trade and displacing local culture with global products. Ferreira and Waldfogel (2013) studied the potential negative impact of cultural products from large economies, such as the United States, on the local cultural products of smaller economies. Gravity estimates showed that there had been a clear trend towards domestic music in recent years, contradicting the notion that large countries dominate the market. Furthermore, their analysis revealed that this trend has increased over the past decade. In addition, Way, Garcia-Gathright, and Cramer (2020) also discovered that the preference for local content among Spotify users increased from 2014 to 2019.

The mentioned findings in the literature show that music streaming significantly impacts the flow of music between countries.

2.2. Hypotheses Development

The following hypotheses were developed based on the findings from the literature review that the widespread adoption of music streaming has fundamentally changed the international music

trade. The literature adapted the gravity model of international trade to quantify music flow, which is discussed in more detail later. It states that the music trade is more significant between geographically close countries with a common language (Way, Garcia-Gathright, and Cramer 2020). Given the effects of the adoption of music streaming just described, previous studies considering only these two variables may have overlooked significant factors affecting the music flow. This gap in previous research may lead to an incomplete understanding of music trade patterns. This work differs from other studies investigating music flows in that we base the gravity model on a new composite distance measure from which our hypotheses are derived. The first hypothesis is to test whether the previously considered distance measure is still significant after adding more independent variables to the model.

H1. Music flow is negatively associated with geographical distance.

The second hypothesis posits that music flow is related to cultural factors. Music is a cultural phenomenon, and different cultures may have distinct musical preferences. Previous research has studied the relationship between cultural proximity and trade and found that cultural proximity positively affects trade by lowering trade costs (Disdier et al. 2010). To further explore the relationship between music flow and cultural factors, two sub-hypotheses are proposed:

H2. Music flow is associated with cultural factors.

H2a. Music flow is positively associated with a common official language.

H2b. Music flow is negatively associated with cultural distance.

The third hypothesis is based on the idea that social connections between people in different countries can influence their music exchange. For example, through social media, individuals can discover new music from different parts of the world and share it with their friends and followers (Dewan, Ho, and Ramaprasad 2017). This can lead to an increase in the popularity of certain songs or artists and, subsequently, in music flows between countries.

H3. Music flow is negatively associated with social distance.

The final hypothesis is that the economic relationship between countries can influence their trade. Therefore, we assume that countries with solid economic relationships will be more likely to trade music and other cultural products with each other.

H4. Music flow is negatively associated with economic distance.

3. Data

The primary data source for our analysis is the music streaming service Spotify. In the third quarter of 2022, Spotify reported a peak of 456 million monthly active users. Furthermore, in terms of paying subscribers, Spotify was the market leader among music streaming providers worldwide in Q1 2021, with a market share of 32% (Statista 2022a; Statista 2022b). Given this dominant position, it is plausible to analyze Spotify's data to obtain generic results for the music streaming industry.

In this study, we used Spotify's top 200 daily charts, which show the most played songs per country and their total worldwide streams. We have the song title, artist name, number of streams, unique ID, and more for each song. We scraped this data from January 1, 2017, to June 26, 2021, and concatenated the daily charts into a single dataset. An observation in the dataset represents a song on a particular country's chart on a given day. To obtain the desired dyadic data of country pairs, we utilized the International Standard Recording Code (ISRC) assigned to each song to extract the country code of the songs' origin country for each song. As part of the clean-up process, some relatively small duplicates in the ISRC codes were corrected to avoid artificially inflating the number of countries in the dataset. For example, the breakdown of some countries was combined into a single designation, such as QM and QZ, to US. We then aggregated the music flow of each country pair on an annual basis, resulting in a dataset where each observation represents the unidirectional music flow from a country of origin j to the destination country i for one year. If we include observations where the origin country equals

the destination country, the final dataset contains 5616 observations with 69 unique consuming countries and 60 unique exporting ones. The dataset includes a total of 671,516,285,867 streams, with 61.11% originating from the United States. Analysis of the dataset and examination of the network graph in Appendix 1 reveals the dominant position of the United States in the global music market, which occupies a central position in the network graph. Furthermore, the music of the top 10 artists in the dataset is primarily produced in the United States, as shown in Appendix 2. This indicates the significant influence of the country on global music flows.

4. Research Method

This chapter presents the methodology for addressing the research question and testing the hypotheses. First, it explains the underlying empirical model and derives the formula for the analysis. Then, the dataset was extended with additional distance measures between the countries to test the hypotheses in the next section. This chapter is the foundation for the subsequent interpretation and analysis of the results.

4.1. Empirical Model Specification

A gravity model is a popular approach to evaluating and predicting economic variables, especially bilateral trade flows (Kabir, Salim, and Al-Mawali 2017). It is the most prevalent illustration of spatial interaction modeling. The *law of universal gravity*, a part of Newtonian physics, is the foundation for the gravity model. According to it, the force of attraction between two bodies is affected by their size and proximity. Economists have adopted this concept to describe bilateral trade flows between countries, the gravity model of international trade. By analogy with flows of goods between countries, the model shows that trade increases with the size and proximity of trading partners (Porojan 2001). Because of its strong theoretical foundations and empirical efficacy in forecasting bilateral trade flows between nations of

diverse commodities under varied circumstances, it has acquired significant application in economics. Gravity models emphasize the spatial relationship between origins and destinations and have already been employed in literature to explore bilateral music flows in the streaming era. Since international music demand is also a form of international trade, it is reasonable to investigate the determinants of music flows with the help of a gravity model. In this paper, we use the existing extensive trade literature and advocacy of the gravity model to quantify the extent of various distance measures in global music flows. In their simplest form, gravity models of international trade follow this equation:

$$(1)\ trade_{ij} = G * \frac{MiMj}{dist\_geoij}$$

$Trade_{ij}$ is the estimate of bilateral trade flows between the place of origin i and destination j. The volume of spatial interaction between the two countries can be expressed in terms of some constant gravitational force G times the product of their economic masses M_i and M_j (typically approximated using their GDP) divided by the geographic distance $dist\_geo_{ij}$ from one another (Way, Garcia-Gathright, and Cramer 2020). This formula implies that the larger the economies are, the more they trade; the larger one economy is, the more goods it can sell or buy from the other, and vice versa. Depending on whether they are closer together or farther apart, trade is easier or more challenging to conduct. Thus, spatial interaction is directly related to economic mass and, inversely, to the distance between them.

This model provides a foundation for the magnitude of trade flows that can reasonably be anticipated between countries based on their relative economic size and proximity. However, it represents only a simple form, as many factors beyond those two variables shape trade partnerships. When examining the music flow of country pairs in the streaming era, we assume that adding multiple variables gives a better result. We augment equation (1) with two dummy variables to account for the effects of shared language and home bias. $Common\_lng_{ij}$ takes the value of 1, if country i and j have the same official language and 0 otherwise. $Home\_bias_{ij}$ takes

the value of 1 if country i equals country j, and 0 if they are not. This inclusion allows us to test the impact of the two variables on bilateral music flows.

Moreover, we believe that a range of distance measures reflects a better result regarding the volume of music flows between country pairs than just simple geographic distance. Thus, the distance variable is decomposed into the geographic, cultural, social, and economic distance. For econometric analysis, the multiplicative model (1) can be estimated in linear regression by taking the natural logarithm of both sides. To bring our results in line with the literature, we fitted gravity models of logarithmized consumption using ordinary least squares regression (Gomez-Herrera, Martens, and Waldfogel 2014; Ferreira and Waldfogel 2013; Way, Garcia-Gathright, and Cramer 2020). This results in the following:

$$(2)\ log(trade_{ij}) = \beta 0 + \beta 1\ log(M_i) + \beta 2\ log(M_j) + \beta 3\ log(dist\_geo_{ij}) + \beta 4\ log(dist\_eco_{ij}) + \beta 5\ log(dist\_cultural_{ij}) + \beta 6\ log(dist\_social_{ij}) + \beta 7\ common\_lng_{ij} + \beta 8\ home\_bias_{ij} + \mu_i + \mu_j + \gamma_{ij}$$

In this equation, $log(trade_{ij})$ represents the logarithm of the trade flow between two countries during a given year. This trade flow is influenced by the size of the source and destination economies (M_i and M_j), as well as the logarithm of various measures of distance between the two countries, including geographical (in km), economic, cultural, and social. Moreover, whether they share an official language $common\_lng_{ij}$, if there is domestic consumption $home\_bias_{ij}$, origin and destination-year fixed effects μ_i and μ_j, and ε_{ij} as an error term.

To apply the Spotify dataset to the Gravity Model, we used Python's Gravity Modeling Environment (GME) package, a collection of tools that can be used for gravity trade analysis. The estimation strategy used in the GME module is based on the Poisson Pseudo Maximum Likelihood (PPML) method. This regression analysis is used to analyze count data, such as the number of objects in a gravitational system. The advantage of using the GME module with the PPML estimator is that it is more efficient than linear regression for count data, producing more

precise estimates with a smaller sample size. Moreover, it can handle data with zeros, a common problem in count data. Linear regression tends to produce biased estimates when there are zeros in the data, but PPML regression can account for this issue (GME Package Documentation n.d.).

4.2. Variable Operationalization

The dependent variable in our gravity model, *lnStreams*, is measured as the logarithm of the cumulative sum of streams in the top 200 charts of country i from origin country j on an annual basis. A higher value of this variable indicates that consumers in country i have a greater preference for songs from country j, which drives their music trade.

The control variables used are as follows:

- *common_lng*$_{ij}$*:* this dummy variable indicates whether two countries have the same official language. It takes a value of 1 if they do and 0 if they do not. To obtain the language data of the countries, we used a Python module called countryinfo. A positive and significant estimate of *common_lng*$_{ij}$ supports H2a.
- *lndist_geo*$_{ij}$: represents the log of the geographic distance (in km) between the two countries capitals. This data was extracted from CEPII's Gravity database (Conte 2022). A negative and significant estimate of *lndist_geo*$_{ij}$ supports H1.

To test our hypotheses, we employ several moderating variables in our empirical analysis:

- *lndist_social*$_{ij}$: We draw on Meta's Social Connectedness Index (SCI) to determine the social distance between countries. The SCI uses Meta friendships to measure the connectedness between two regions, and a higher SCI score indicates a higher level of social connectedness (Bailey et al. n.d.). To bring the effect of the variable in line with the other variables, we took the inverse of the index (in log). A negative and significant estimate of *lndist_social*$_{ij}$ supports H3.

- *lndist_cultural*$_{ij}$: Hofstede's cultural dimensions theory is used to measure cultural distance between countries. Hofstede identifies six comparison criteria - uncertainty avoidance, power distance, individualism, masculinity, indulgence, and long-term orientation - representing similarities and differences between cultures. We supplement our Spotify dataset with Hofstede's culture dimensions dataset (Geert Hofstede. n.d.). We extract the dimension for both countries in each country pair and combine the cultural dimensions into an index using Yang et al.'s (2022) formula: $CD_{od} = \sqrt{\sum_{i=1}^{6}\{(I_{io} - I_{id})^2 \,/\, V_i\}}$

  Again, we used the log of this variable in our final dataset. A negative and significant estimate of *lndist_cultural*$_{ij}$ supports H2b.
- *lndist_eco*$_{ij}$: We followed previous literature to reflect the economic distance of country pairs. We took the inverse of the average percentage of their music flow relative to their total international music trade (Yang et al. 2022). This variable is dynamic, so we use its lagged value to capture the time effect on the economic relationship. For example, for the 2021 music flow, we use the 2020 trade flow. We obtain the trade flow data from the CEPII Gravity database (Conte 2022). However, since this dataset only contains data up to 2019, we manually create the trade flow dataset for 2020 from the Comtrade website. A negative and significant estimate of *lndist_eco*$_{ij}$ supports H4.
- *home_bias*$_{ij}$: this dummy variable is 1 when country i is equal to country j, meaning it listens to its own music. Otherwise, the variable is 0. It captures the tendency of people to listen to music from their own country rather than from other countries.

The descriptive statistics for the variables discussed above can be found in Appendix 3 better to understand the distribution and characteristics of the data.

5. Empirical Results

In the following, we analyze the gravity model results to test the established hypotheses. Next, we perform two robustness checks on the results to assess their credibility and draw conclusions about the relationships between the variables.

5.1. Presentation of Results

Table 2 reports the gravity estimation results based on equation (2). Each model (1)-(7) includes fixed effects for each origin and destination country by year to capture the country size and consequently avoid size bias. Different models comprise varying observations, ranging from 5,355 to 4,375, due to the absence of data for specific distance measures.

Table 2. Gravity estimation results based on music flows, 2017-2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
lndist_geo	-0.051***	-0.012***	-0.046***	-0.035***	-0.047***	-0.015***	-0.013***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.005)	(0.005)
common_lng	0.116***	0.071***	0.102***	0.105***	0.129***	0.066***	0.055***
	(0.008)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
lndist_social		-0.043***				-0.041***	-0.041***
		(0.003)				(0.003)	(0.003)
lndist_cultural			-0.027***				-0.032***
			(0.005)				(0.005)
lndist_eco				-0.012***		0.001	0.003
				(0.003)		(0.003)	(0.003)
home_bias					0.119***	0.146**	
					(0.012)	(0.060)	
Observations	5,355	5,035	5,050	4,869	5,616	4,570	4,375
Country FE	yes	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes	yes
Pseudo-R ²	0.117	0.119	0.118	0.117	0.117	0.119	0.119

Notes: Dependent variable Model (1)-(7) log Sum of Streams top 200 charts. Models where home_bias is included, use all country pairs, the remaining ones only those where country i≠j. *** 0.001 significance level; ** 0.01 significance level; * 0.05 significance level. Country-pair-based clustered standard errors are presented in parentheses.

Model (1) describes the baseline model for our analysis with the two control variables, geographical distance, and common language, which are included in every subsequent model. Models (2)-(5) each add a distance measure to examine its discrete effect on our dependent variable. Model (6)-(7) combine all independent variables in one model to see how they interact with each other, (6) excluding cultural distance, and (7) excluding home bias. Due to the collinearity of those two variables, they cannot be represented together in a model because when

home_bias is 1, *dist_cultural* is 0. This relationship is not apparent for any of the other independent variables.

The first variable, geographic distance, remains negatively significant at the 0.001 level across all models. The model (7) coefficient is -0.013***, so a 1% increase in this variable decreases the music flow by 1.3%. H1 is thus supported. The pattern of the common official language behaves similarly but in the inverse direction. A country pair speaking the same language increases their music flows by about 5.5%. The coefficient remains positively significant at the 0.001 level across all models, confirming H2a. Continuing the trends found in previous studies (Ferreira and Waldfogel 2013; Way, Garcia-Gathright, and Cramer 2020), geographic distance and common language remain significant in explaining bilateral music flows, even when other distance measures are added. One question that might arise is why geographic distance plays a significant role in determining bilateral music flow in music streaming. As noted in previous chapters, music streaming has reduced the marginal cost of music distribution and consumption to almost zero and removed barriers to the internationalization of music (Ferreira and Waldfogel 2013). Indeed, Way et al. (2020) found a decreasing effect of geographic distance on music streams over time. However, it should be considered that the costs of trading digital music are not limited exclusively to production or distribution costs but may include other costs. There are other barriers to trade, such as costs associated with cultural differences (Gomez-Herrera, Martens, and Waldfogel 2014). Therefore, the geographic distance can also represent other frictions, such as information friction due to language barriers (Bailey et al. 2021). This could explain why people are less likely to listen to songs produced far away from them, indicating the continued importance of geographic distance in music streaming.

Looking at the other results in Table 2, it quickly becomes apparent that the effect of economic distance is highly dependent on the other variables with which it is integrated into a model. For example, in the model (4), the variable is significantly negatively associated with music flow.

However, once it is combined with the other distance measures in models (6) and (7), it is no longer significant for the result. Therefore, H4 could only be partially confirmed.

Moreover, H2b is supported as the results indicate that music flow is negatively associated with cultural distance. Model (7) shows that a 1% increase in this variable leads to a 3.2% decrease in bilateral music flow. The explanation for this could lie in observing the increasing cultural divergence of popular music. Bello and Garcia‘s (2020) analysis discovered an upward trend in the diversity of music consumption, leading to increasing differentiation in countries' music charts. The authors attribute this to a segmentation of the music market, as groups of countries with similar tastes increasingly converge while differentiating themselves from others. These clusters seem to be strongly determined by geographical and cultural distance. Therefore, it can be inferred that music charts have diversified due to cultural differences between countries, making it essential to consider this factor when quantifying bilateral music flows.

Similarly, it is observable that music flow is negatively significantly associated with social distance. The dependent variable decreases by about 4.1% for every 1% increase in the variable, concluding that H3 also states the truth. It is worth noting that when comparing models (2)-(5), after adding social distance to our baseline model, the absolute values of *dist_geo* and *common_lng* decrease the most. Likewise, Bailey et al. (2021) discovered that once social connectedness is controlled for, the estimated effects of geographic distance on trade significantly decrease. The authors propose that social connectedness, as measured by Meta's Social Connectedness Index, helps to alleviate information bottlenecks between countries. Although only slightly, R^2 increases the most when adding social distance to the model compared to the other independent variables. These findings suggest that social connectedness plays a crucial role in shaping musical flows.

The reason for this significant impact could be demographics. Most Spotify users belong to the Millennial and Gen Z age groups, and therefore it can be inferred that the Spotify top 200 charts

primarily reflect their musical preferences. These young adults are highly connected through social media, which enables them to interact and share information, including cultural assets such as music. Previous research has already investigated the impact of peer influences on an individual's music consumption (Dewan and Ramaprasad 2014; Kaimann, Tanneberg, and Cox 2021; Kim, Suh, and Lee 2014). Dewan, Ho, and Ramaprasad (2017) studied the impact of social influence on the favoriting behavior of users in an online music community. This community allowed users to favorite songs and track the behavior of their friends and the entire community. Their analysis identified strong evidence of the so-called *proximity influence* in music consumption, which arises from the favoritism behavior of immediate friends within the social network. The authors speculate that this may be due to the *bandwagon effect*, a psychological phenomenon in which people respond primarily to what others are doing. In digital streaming services, users copy other users' consumption behavior and listen to a particular song only because they see other users doing the same. In addition, Kim, Suh, and Lee (2014) found that the music-listening behavior of Twitter users, as measured by daily tweets about specific songs, strongly correlates with general music trends and plays a significant role in understanding consumer behavior. This demonstrates the influence of peers on music consumption in the era of music streaming and could be a plausible explanation for the significance of social distance in our model. However, it is essential to mention that consumers' listening habits cannot be directly controlled; thus, this is only one possible approach to interpreting the results.

To explore this further with our example of Spotify data, we draw on Spotify's viral charts, which track and rank the songs that attract the most attention on social media. As viral charts capture social buzz, they are more diverse than the top charts and include more country pairs. Using Spotify's viral charts in our gravity model, we expect to observe a significant and potentially even more vital effect of distance measures on the dependent variable due to social

connectedness. Since the dataset does not provide information on the number of streams but only on ranks, we will use the logarithm of the survival time in days that individual tracks from country j spent in the viral charts of country i on an annual basis as the new dependent variable.

Table 3. Gravity estimation results based on viral music flows, 2017-2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
lndist_geo	-0.132***	-0.041***	-0.125***	-0.091***	-0.126***	-0.040***	-0.033***
	(0.004)	(0.006)	(0.004)	(0.007)	(0.004)	(0.007)	(0.007)
common_lng	0.305***	0.197***	0.270***	0.279***	0.310***	0.187***	0.163***
	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
lndist_social		-0.094***				-0.093***	-0.093***
		(0.004)				(0.005)	(0.005)
lndist_cultural			-0.056***				-0.056***
			(0.008)				(0.008)
lndist_eco				-0.037***		-0.006	-0.006
				(0.005)		(0.005)	(0.005)
home_bias					0.176***	0.057	
					(0.023)	(0.065)	
Observations	8,570	8,034	8,058	7,845	8,846	7,340	7,000
Country FE	yes	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes	yes
Pseudo-R ²	0.165	0.169	0.165	0.166	0.165	0.169	0.170

Notes: Dependent variable top Viral Charts Survival Days. Models, where *home_bias* is included use all country pairs, the remaining ones only those where country i≠j. *** 0.001 significance level; ** 0.01 significance level; * 0.05 significance level. Country-pair-based clustered standard errors are presented in parentheses.

Table 3 shows similar results, with the same distance measures remaining significant. Only *home_bias* is no longer significant in model (6). One explanation could be that *dist_social* already absorbs the effect of *home_bias*. That geographic proximity is representative of social connectedness between individuals has already been confirmed in the literature, indicating that distant countries tend to have a higher social distance (Bailey et al. 2021). However, it is impossible to determine whether the distance measures are more pronounced in this case compared to the top charts. Two different data samples are used, so the absolute values of the coefficients of the estimations cannot be directly compared. It is important to note that the mechanism underlying the viral charts differs from that of the top charts. The viral charts reflect the songs people are talking about on social media, while the top charts show the most popular songs in terms of streams. The results show that the distance measures still constrain individuals willing to share songs on social media.

After testing all hypotheses, the second part of the study examines whether the results are still valid when the effects of the COVID-19 pandemic are considered. The pandemic has undoubtedly changed consumer behavior and influenced cultural goods consumption. Previous research has shown that the COVID-19 pandemic has also changed how people consume music, leading to a significant decrease in the consumption of streaming music in many countries (Denk et al. 2022; Yeung 2020; Sim, Cho et al. 2022). In the following, we investigate whether this change in music consumption has affected the importance of the different distance measures. In order to account for the effect of the pandemic on our results, we re-estimated the models from Table 2 by adding an interaction term for the pandemic to each model. This interaction term is represented by a dummy variable multiplied by each distance measure to determine the direct effect of the pandemic on each measure. This allows us to see how the pandemic has affected the significance of the different distance measures.

Table 4. Gravity estimation results based on music flows COVID-19 interaction terms, 2017-2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
lndist_geo	-0.053***	-0.012**	-0.051***	-0.031***	-0.036***	-0.014*	-0.013*
**covid*	(0.004)	(0.005)	(0.004)	(0.007)	(0.004)	(0.007)	(0.007)
common_lng	0.134***	0.116***	0.131***	0.128***	0.147***	0.115***	0.113***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
lndist_social		-0.044***				-0.038***	-0.038***
**covid*		(0.004)				(0.004)	(0.004)
lndist_cultural			-0.017**				-0.017*
**covid*			(0.007)				(0.009)
lndist_eco				-0.015***		-0.001	0.000
**covid*				(0.005)		(0.005)	(0.005)
home_bias					0.185***	0.406***	
					(0.012)	(0.063)	
Observations	5,355	5,035	5,355	4,869	5,616	4,570	4,569
Country FE	yes	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes	yes
Pseudo-R ²	0.115	0.116	0.115	0.115	0.115	0.116	0.116

Notes: Dependent variable log Sum of Streams with covid interaction terms, meaning that each distance measure is multiplied by a covid dummy variable. Models, where *home_bias* is included use all country pairs, the remaining ones only those where country i≠j. *** 0.001 significance level; ** 0.01 significance level; * 0.05 significance level. Country-pair-based clustered standard errors are presented in parentheses.

The results clearly show that COVID-19 can be seen as a catalyst, so to speak, that further strengthens all distance measures. In particular, the coefficients of *common_lng*, *dist_social*, and *home_bias* impact the result at the 0.001 significance level. For instance, in the case of

social distance, this means that songs from low-connected countries are heard even less frequently after the pandemic's onset. This variable was already negatively associated before, but after the onset of the pandemic, this effect was amplified again. These three significant variables can be interpreted as a preference for locally produced content. The increasing preference for locally produced music has already been demonstrated in the literature (Ferreira and Waldfogel 2013; Way, Garcia-Gathright, and Cramer 2020). The substantial increase may be due to the pandemic strengthening the preference for locally produced music even further. This strengthening may have been exacerbated by the fact that people could not travel as much during the pandemic as they could before (Yang et al. 2022), resulting in fewer cultural goods being exchanged. Additionally, previous research has discovered that individuals tend to listen to more nostalgic music following the outbreak of the pandemic, which many individuals associate with music from their home country (Yeung 2020). This finding is consistent with research indicating that people use nostalgia to cope with challenging situations (Sedikides, Wildschut, and Baden 2004). The gravity model results show that COVID-19 strengthened distance measures and increased the importance of nostalgic and native songs. The pandemic caused people to prefer songs close to them and culturally similar, which led to a further decrease in the music trade from countries that are already distant.

5.2. Robustness Check

Two robustness tests are performed to check whether the obtained results are robust by re-running the models listed in Table 2 with minor modifications. The first robustness check (Appendix 4) uses the same dataset as in Table 2, with the new dependent variable being the logarithm of the chart survival days. This variable measures the cumulative amount of time in days that individual songs from country j remained in the top 200 charts of country i on an annual basis. The results show that the overall trends are consistent across the different models. The music flow is negatively associated with geographical, cultural, and social distance

(supporting H1, H2b, H3) and positively associated with a common official language (supporting H2a). Combining all independent variables, economic distance remains insignificant in both robustness checks (contradicting H4). In model (6), *home_bias* is no longer significant, which could be caused by the fact that *dist_social* already absorbs its effect, as explained in the previous section (Bailey et al. 2021).

The second robustness check (Appendix 5) used the same dataset as the first but filtered by position to obtain a dataset with only the top 50 charts. In this case, the dependent variable was again the chart survival days. Although the literature discusses that the top charts are becoming more diversified, Bello and Garcia (2020) point out that the differences between songs within the top 10 are the smallest. They suggest that the artists whose songs reach the top of the charts have a more similar style. From this, it can be inferred that while there is a higher proportion of unique artists at the top of the charts, the music they create is relatively similar. Thus, we assume that the top 50 charts are more similar than the top 200 charts, and therefore specific distance measures may lose their importance for quantifying bilateral music flow. A look at the gravity estimation results in Appendix 5 confirms this assumption. *Common_lng* is no longer significant for the result, rejecting H2a. We can conclude that the songs in the top 50 charts have predominantly the same lyric language. The Economist's (2022) study examines the top 100 Spotify tracks in 70 countries over five years and finds that while the dominance of English is declining, it is still the dominant lyrical language. Geographic distance remains negatively significantly associated with the music flow, supporting H1. Compared to Table 2, the significance level changed from 0.001 to 0.01. H2a and H3 also stay true, as cultural and social distance remain negatively significant even when considering only the top 50 charts, reflecting these variables' dominant influence.

As can be seen, the main empirical results were confirmed using two robustness checks.

6. Discussion

We will first examine the implications of our results. Then, we will briefly consider the limitations of our findings to provide a more balanced and nuanced view.

The results of this study have implications for both Spotify and the broader music industry. From a business perspective, Spotify is the primary platform for generating revenue in the music streaming industry. Artists increasingly rely on platforms like Spotify to sell their music, so their revenue is tied to the number of streams they receive (Graham et al. 2004). However, as our results show, music flow between countries is influenced by several factors, including geographic, cultural, and social distance. This means that artists from smaller countries may get fewer streams, even if their music is just as good as that of artists from larger countries. To address this problem, Spotify could consider strategies to promote the music flow between countries that are not firmly connected and encourage listeners to discover music from more distant and dissimilar parts of the world. One option is recommending more diverse songs to listeners or promoting artists from distant countries on the platform.

In the post-pandemic world, it will be even more important to understand music market behavior and take action to support the internationalization of music. As our results show, the pandemic has further exacerbated the influences of social and cultural differences between countries on their music flows. As a result, people are less inclined to listen to music from other parts of the world with which they do not feel connected (Yeung 2020). The music flow from countries that are already socially disconnected is thus further reduced by the pandemic. This suggests that after the pandemic, people will have less opportunity to be exposed to music from countries that are very different from theirs. It may take years for people to recover from this effect. By promoting diverse music on the platform, Spotify could help counteract this trend and support music flow between countries.

There are a few potential limitations of the results that should be considered. Firstly, the data used in the study are drawn from Spotify's daily top 200 songs from 96 countries, which may only represent some of the international music markets. Additionally, Spotify is only available in some countries, which could limit the generalizability of the findings to the broader population of music consumers. Furthermore, using platform-specific data may introduce bias into the analysis due to potential differences between the users of the platform and the general population. For instance, research has shown that Spotify users are disproportionately young and male compared to the general population of their countries (Datta, Knox, and Bronnenberg 2018). Secondly, the study focuses on the moderating effects of distance measures on bilateral music flows. However, it does not consider other factors that may influence music consumption, such as the popularity of specific genres. The popularity of artists or genres may play a significant role in shaping music flows between countries. According to research conducted by Park et al. (2019) regional clusters exhibit variations in the acoustic characteristics of the music preferred by their respective populations.

Furthermore, scientists have already studied the algorithmic effects on the diversity of consumption on Spotify and found that algorithmically controlled recommendations are associated with lower diversity in consumption (Anderson et al. 2020). These factors could limit the study's ability to comprehensively understand the relationship between international music demand and consumer preferences. These limitations should be considered when interpreting the results of the study.

7. Conclusion

In conclusion, this study has found that a composite distance measure that considers cultural and social factors is more effective in understanding consumer preferences than the previous measure that mainly considers only geographic and linguistic proximity. This suggests that digitization has changed the importance of individual distance measures in the music streaming

market. Furthermore, our results indicate that while geographic distance and common language remain significant, cultural and social distance between countries also significantly impact the music flow. Specifically, we found that geographic, cultural, and social distance are all negatively associated with music flow, while common official language and home bias are positively associated with music flow. These findings add to our understanding and further studies of the factors that shape consumer consumption patterns in the digital age and highlight the need for a broader range of distance measures when studying the gravity model for bilateral music flow. Likewise, they have implications for music streaming platforms and can help develop strategies to promote the internationalization of music in the digital age.

Further research is needed to examine the impact of other factors on the flow of music between countries, such as the availability of music streaming platforms or the popularity of specific genres in different countries. Additionally, this study only considered a limited number of cultural and social factors. Other factors, such as shared historical or political experiences, could also impact the music trade between countries. Therefore, future studies could consider a broader range of cultural and social factors to provide a more comprehensive understanding of their impact on the music flow.

However, measuring consumers' listening habits is always challenging due to the numerous factors that can influence them. As such, the findings of this study should be interpreted in the context of the limitations of the data used.

II. **References**

Anderson, Ashton, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. 2020. "Algorithmic Effects on the Diversity of Consumption on Spotify." In *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020*, 2155–65.

Araujo, Carlos V S, Marco A P Cristo, and Rafael Giusti. 2017. "Predicting Music Popularity on Streaming Platforms." *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*, 149–56.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel. 2021. "International Trade and Social Connectedness." *Journal of International Economics* 129 (March).

Bailey, R. Cao, T. Kuchler, J. Stroebel and A. Wong. n.d. "Social Connectedness: Measurements, Determinants, and Effects." Journal of Economic Perspectives, 32(3):259–80, 2018b. and the Facebook Data for Good Program, Social Connectedness Index (SCI). Accessed December 1, 2022. https://dataforgood.facebook.com/.

Bello, Pablo, and David Garcia. 2020. "Cultural Divergence in Popular Music: The Increasing Diversity of Music Consumption on Spotify across Countries." *Humanities and Social Sciences Communications* 8 (1).

Conte, M., P. Cotterlaz and T. Mayer. 2022. "The CEPII Gravity Database." N°2022-05, July 2022.

Cowen, T. 2009. "Creative Destruction: How Globalization is Changing the World's Cultures." Princeton University Press

Datta, Hannes, George Knox, and Bart J. Bronnenberg. 2018. "Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery." *Marketing Science* 37 (1): 5–21.

Denk, Janis, Alexa Burmester, Michael Kandziora, and Michel Clement. 2022. "The Impact of COVID-19 on Music Consumption and Music Spending." *PLoS ONE* 17 (May 5).

Dewan, Sanjeev, Yi Jen Ho, and Jui Ramaprasad. 2017. "Popularity or Proximity: Characterizing the Nature of Social Influence in an Online Music Community." *Information Systems Research* 28 (1): 117–36.

Dewan, Sanjeev, and Jui Ramaprasad. 2014. "Social Media, Traditional Media, and Music Sales." *Source: MIS Quarterly* 38 (1): 101–22.

Disdier, Anne Célia, Silvio H.T. Tai, Lionel Fontagné, and Thierry Mayer. 2010. "Bilateral Trade of Cultural Goods." *Review of World Economics* 145 (4): 575–95.

Economist. 2022. "What Spotify Data Show about the Decline of English." https://www.economist.com/interactive/graphic-detail/2022/01/29/what-spotify-data-show-about-the-decline-of-english.

Ferreira, Fernando, and Joel Waldfogel. 2013. "Pop Internationalism: Has Half a Century of World Music Trade Displaced Local Culture?" *The Economic Journal* 123 (569): 634–64.

Geert Hofstede. n.d. „Dimension data matrix" Accessed November 15, 2022. https://geerthofstede.com/research-and-vsm/dimension-data-matrix/.

Georgieva, Elena, Marcella Suta, and Nicholas Burton. 2018. "Hitpredict: Predicting Hit Songs Using Spotify Data."

GME Package Documentation. n.d. "Methodology - GME Package Documentation," Accessed December 1, 2022. https://www.usitc.gov/data/gravity/gme_docs/estimate_technical/.

Gomez-Herrera, Estrella, Bertin Martens, and Joel Waldfogel. 2014. "What's Going On? Digitization and Global Music Trade Patterns since 2006." *SSRN Electronic Journal.*

Graham, Gary, Bernard Burnes, Gerard J. Lewis, and Janet Langer. 2004. "The Transformation of the Music Industry Supply Chain: A Major Label Perspective." *International Journal of Operations and Production Management*.

IFPI. 2022. "GLOBAL MUSIC REPORT STATE OF THE INDUSTRY."

Interiano, Myra, Kamyar Kazemi, Lijia Wang, Jienian Yang, Zhaoxia Yu, and Natalia L. Komarova. 2018. "Musical Trends and Predictability of Success in Contemporary Songs in and out of the Top Charts." *Royal Society Open Science* 5 (5).

Kabir, Mahfuz, Ruhul Salim, and Nasser Al-Mawali. 2017. "The Gravity Model and Trade Flows: Recent Developments in Econometric Modeling and Empirical Evidence." *Economic Analysis and Policy* 56 (December): 60–71.

Kaimann, Daniel, Ilka Tanneberg, and Joe Cox. 2021. "'I Will Survive': Online Streaming and the Chart Survival of Music Tracks." *Managerial and Decision Economics* 42 (1): 3–20.

Kim, Yekyung, Bongwon Suh, and Kyogu Lee. 2014. "Nowplaying the Future Billboard: Mining Music Listening Behaviors of Twitter Users for Hit Song Prediction." In *SoMeRA 2014 - Proceedings of the 1st ACM International Workshop on Social Media Retrieval and Analysis, Co-Located with SIGIR 2014*, 51–55. Association for Computing Machinery, Inc.

Magu, S. 2015. "Reconceptualizing cultural globalization: Connecting the "cultural global" and the "cultural local"." *Social Sciences* (4):630–645.

McKinney Kelsey. 2015. “A Hit Song Is Usually 3 to 5 Minutes Long. Here’s Why. - Vox.” *Vox*, 8–12. https://www.vox.com/2014/8/18/6003271/why-are-songs-3-minutes-long.

Park, Minsu, Jennifer Thom, Sarah Mennicken, Henriette Cramer, and Michael Macy. 2019. “Global Music Streaming Data Reveal Diurnal and Seasonal Patterns of Affective Preference.” *Nature Human Behaviour*. Nature Publishing Group.

Porojan, A. 2001. “Trade Flows and Spatial Effects: The Gravity Model Revisited.” *Open Economies Review* 12: 265–80.

Prior, Nick. 2013. “Bourdieu and the Sociology of Music Consumption: A Critical Assessment of Recent Developments.” *Sociology Compass* 7 (3): 181–93.

Sim, Jaeung, Daegon Cho, Youngdeok Hwang, and Rahul Telang. 2022. “Frontiers: Virus Shook the Streaming Star: Estimating the COVID-19 Impact on Music Consumption.” *Marketing Science* 41 (1): 19–32.

Sim, Jaeung, Jea Gon Park, Daegon Cho, Michael D. Smith, and Jaemin Jung. 2022. “Bestseller Lists and Product Discovery in the Subscription-Based Market: Evidence from Music Streaming.” *Journal of Economic Behavior and Organization* 194 (February): 550–67.

Soares Araujo, Carlos Vicente, Marco Antonio Pinheiro De Cristo, and Rafael Giusti. 2019. “Predicting Music Popularity Using Music Charts.” *Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019*, 859–64.

Statista. 2022. “Number of Spotify monthly active users (MAUs) worldwide from 1st quarter 2015 to 3rd quarter 2022” Accessed November 20, 2022. https://www.statista.com/statistics/367739/spotify-global-mau/.

Statista. 2022. “Share of music streaming subscribers worldwide in the 1st quarter of 2021, by company” Accessed November 20, 2022. https://www.statista.com/statistics/653926/music-streaming-service-subscriber-share/.

Waldfogel, Joel. 2017. “How Digitization Has Created a Golden Age of Music, Movies, Books, and Television.” *Journal of Economic Perspectives*. American Economic Association.

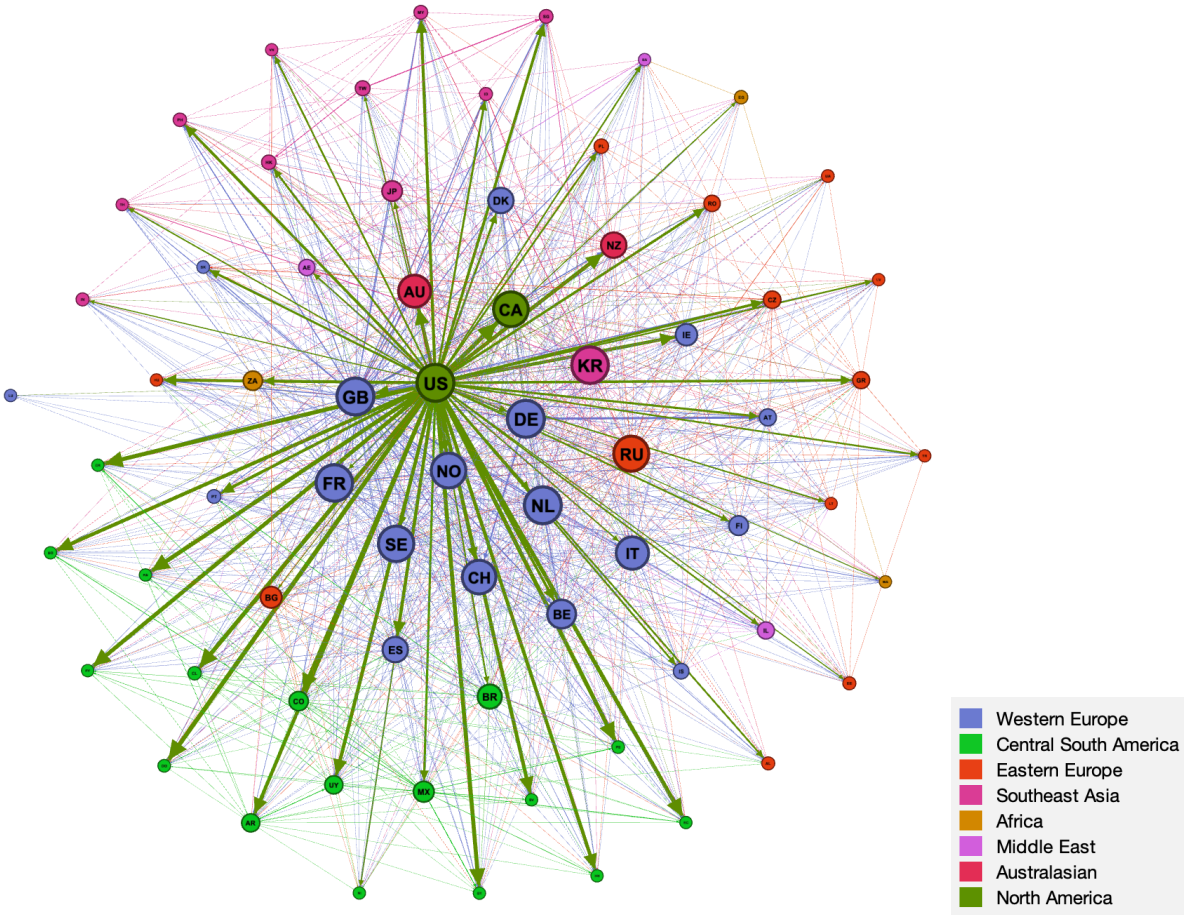
Way, Samuel F, Jean Garcia-Gathright, and Henriette Cramer. 2020. “Local Trends in Global Music Streaming.” In *Fourteenth International AAAI Conference on Web and Social Media*, 2020:705–14.

Yang, Yang, Linjia Zhang, Laurie Wu, and Zhenlong Li. 2022. “Does Distance Still Matter? Moderating Effects of Distance Measures on the Relationship Between Pandemic Severity and Bilateral Tourism Demand.” *Journal of Travel Research*.

Yeung, Timothy Yu-cheong. 2020. “Did the COVID-19 Pandemic Trigger Nostalgia? Evidence of Music Consumption on Spotify.” *Covid Economics* (44): 154-185.

Appendix

Appendix 1. Network Graph based on Spotify's top 200 music flows, 2017-2021.



The network graph was generated using Gephi and only included the top 75% quantile of country pairs based on their survival days in the top 200 charts.

Appendix 2. Country of Song Production of Top 10 Artists of Total Music Streams, 2017-2021.

Artist	Country of Production	Total Streams
Post Malone	US	17748080008
Ed Sheeran	GB	15373631355
Drake	US	11830261104
XXXTENTACION	US	9783243148
Ariana Grande	US	8219158055
Bad Bunny	US	7541015081
Billie Eilish	US	7347984152
Juice WRLD	US	6845549235
Dua Lipa	GB	6745292050
The Chainsmokers	US	6684229741

Appendix 3. Descriptive Statistics for variables used in the Gravity Model.

Variable	Observation	Mean	SD	Min	Max
lnchart_survival	5616	4.870725	2.847150	0.000000	11.108994
lnStreams	5616	13.863668	3.400648	6.909753	24.787480
lndist_geo	5616	8.125494	1.263286	2.257588	9.884789
common_lng	5616	0.184829	0.388193	0.000000	1.000000
lndist_social	5290	-8.733850	2.049287	-19.038242	-4.812184
lndist_cultural	5050	1.206954	0.469718	-2.108513	2.067747
lndist_eco	4870	4.856724	1.720720	0.508237	10.958730
home_bias	5616	0.046474	0.210529	0.000000	1.000000

Appendix 4. Robustness Check 1; Gravity estimation results based on music flows, 2017-2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
lndist_geo	-0.132***	-0.029***	-0.120***	-0.090***	-0.123***	-0.040***	-0.032**
	(0.007)	(0.010)	(0.007)	(0.013)	(0.007)	(0.013)	(0.013)
common_lng	0.314***	0.202***	0.277***	0.281***	0.357***	0.184***	0.149***
	(0.021)	(0.021)	(0.023)	(0.022)	(0.020)	(0.022)	(0.023)
lndist_social		-0.111***				-0.108***	-0.110***
		(0.008)				(0.009)	(0.009)
lndist_cultural			-0.078***				-0.092***
			(0.014)				(0.015)
lndist_eco				-0.035***		0.006	0.009
				(0.009)		(0.009)	(0.009)
home_bias					0.292***	0.309	
					(0.030)	(0.190)	
Observations	5,344	5,024	5,039	4,858	5,611	4,559	4,364
Country FE	yes	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes	yes
Pseudo-R^2	0.234	0.238	0.234	0.234	0.233	0.238	0.239

Notes: Dependent variable log Top 200 Chart Survival Days. Models, where *home_bias* is included use all country pairs, the remaining ones only those where country i≠j. *** 0.001 significance level; ** 0.01 significance level; * 0.05 significance level. Country-pair-based clustered standard errors are presented in parentheses.

Appendix 5. Robustness Check 2; Gravity estimation results based on music flows, 2017-2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
lndist_geo	-0.133***	-0.055***	-0.120***	-0.093***	-0.117***	-0.060***	-0.049***
	(0.008)	(0.012)	(0.008)	(0.015)	(0.008)	(0.016)	(0.017)
common_lng	0.126***	0.045*	0.093***	0.109***	0.193***	0.037	-0.004
	(0.025)	(0.025)	(0.025)	(0.025)	(0.022)	(0.026)	(0.027)
lndist_social		-0.082***				-0.083***	-0.089***
		(0.009)				(0.010)	(0.011)
lndist_cultural			-0.081***				-0.093***
			(0.018)				(0.020)
lndist_eco				-0.030***		0.008	0.010
				(0.011)		(0.011)	(0.012)
home_bias					0.298***	0.660**	
					(0.032)	(0.323)	
Observations	3,038	2,858	2,869	2,750	3,286	2,584	2,474
Country FE	yes	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes	yes
Pseudo-R ²	0.209	0.212	0.210	0.209	0.209	0.212	0.212

Notes: Dependent variable log Top 50 Chart Survival Days. Models, where *home_bias* is included use all country pairs, the remaining ones only those where country i≠j. *** 0.001 significance level; ** 0.01 significance level; * 0.05 significance level. Country-pair-based clustered standard errors are presented in parentheses.