As People Sing, So Do They Live

*Daniel Crisostomo-Wainstock**

University of Oxford

May 17, 2026

Abstract

Music is among the most pervasive forms of human expression and historically constituted a central component of social life across human societies. Yet, despite the growing recognition of culture as an important force shaping economic and social outcomes, music has remained largely absent from empirical work in economics. This paper contributes to filling this gap by introducing a new source of cross-cultural data: the Cantometrics catalogue of traditional songs. This dataset contains information on the stylistic and structural characteristics of musical traditions across hundreds of indigenous societies worldwide, providing unusually broad and systematic coverage of expressive cultural variation across human societies. Combining Cantometrics with historical ethnographic data, I show that musical traits systematically predict ethnographic characteristics spanning social organization, economic practices, and political institutions. Moreover, societies that are closer in musical space are also closer in broader ethnographic space. These findings suggest that musical traditions co-evolve with other dimensions of culture and reflect deeper patterns of cultural variation across populations. Consistent with this interpretation, the paper further documents that ethnic homelands characterized by greater musical similarity exhibit substantially stronger contemporary social and economic connectedness. Taken together, the results illustrate the potential of traditional songs as measurable objects of economic analysis and suggest that cultural proximity plays an important role in shaping patterns of connectedness across subnational regions worldwide.

Keywords: Cultural Evolution, Traditional Songs, Cultural Distance, Connectedness
JEL codes: O10, Z10, Z13

*Daniel Crisostomo-Wainstock: daniel.crisostomowainstock@queens.ox.ac.uk. I am especially grateful to Anna Lomax Wood for her generosity in sharing her deep knowledge of ethnomusicology and the Cantometrics project, and for numerous illuminating conversations about the origins, structure, and interpretation of the data. I would also like to thank Oded Galor, Sam Passmore, and Patrick Savage. I also thank Justin Weiss for excellent research assistance.

"No musical style has 'its own terms': its terms are the terms of its society and culture."

— John Blacking

1. Introduction

Music is among the most pervasive forms of human expression. Across societies and throughout history, traditional songs have accompanied rituals, labor, worship, conflict, courtship, and collective celebration. Long before mass literacy or modern communication technologies, music constituted a central mechanism through which communities transmitted memory, values, identities, and shared forms of behavior across generations. Musical traditions therefore emerged as enduring components of social and cultural life.

Despite the growing recognition of culture as an important force shaping economic development, music has remained largely absent from empirical work in economics.[1] This paper contributes to filling this gap by introducing Cantometrics as a systematic source of cross-cultural data on traditional music. Developed by the ethnomusicologists Alan Lomax and Victor Grauer (Lomax, 1962; Wood et al., 2022), Cantometrics codes the structure and performance style of traditional songs across more than 800 indigenous societies worldwide. The dataset records a rich set of characteristics describing rhythm, melody, vocal coordination, group organization, and vocal production, thereby providing a high-dimensional representation of musical traditions across human societies.

The paper then studies whether musical traditions systematically co-vary with broader dimensions of cultural organization. Building on longstanding ideas in anthropology and ethnomusicology that musical styles are embedded within wider social and cultural systems (Blacking, 1973; Lomax, 2017), I combine Cantometrics with historical ethnographic data to examine this relationship systematically across societies. The analysis shows that musical traits strongly predict ethnographic characteristics spanning social organization, economic practices, and political institutions. Moreover, societies that are closer in musical space are also closer in broader ethnographic space. Together, these results provide quantitative evidence that musical traditions co-evolve with other dimensions of culture and encode meaningful information about broader patterns of cultural variation across populations.

The paper further examines whether musical distance predicts contemporary social and economic connectedness across groups. Combining Cantometrics with data on ethnic homelands, Facebook friendship networks, and bilateral business interactions, I show that groups

[1] Notable exception includes Kampanelis et al. (2023), who examine the relationship between Aboriginal songlines and patterns of economic development in Australia.

with more similar musical traditions exhibit substantially stronger contemporary connectedness. This relationship emerges consistently across specifications, including those incorporating country-pair fixed effects and alternative measures of cultural similarity such as linguistic, folkloric, and religious distance. The results suggest that musical traditions contain persistent information about cultural proximity across populations that remains informative alongside other manifestations of long-run cultural relatedness.

Taken together, the findings suggest that traditional songs contain persistent information about cultural proximity that remains reflected in contemporary patterns of interaction across societies. More broadly, the paper illustrates how systematically coded forms of expressive culture can provide a useful empirical window into processes of cultural evolution and long-run social connectedness.

2. Data and Measurement

This section introduces the Cantometrics dataset and provides a detailed description of its construction, content, and scope. The goal is to establish the dataset as a systematic and credible source of cross-cultural information, comparable in spirit to widely used ethnographic datasets, while emphasizing its distinct focus on expressive cultural behavior.

2.1. Origins and Purpose

Cantometrics was developed by ethnomusicologists Alan Lomax and Victor Grauer with the objective of systematically classifying traditional songs across human societies (Lomax, 1962, 2017; Wood et al., 2022). The project emerged from Lomax's broader research program documenting indigenous music through extensive fieldwork and archival collection, and from the premise that musical style constitutes a pervasive and informative form of cultural expression rather than an arbitrary aesthetic choice. Crucially, the dataset focuses exclusively on *traditional* songs—musical forms that emerged locally and were transmitted prior to large-scale globalization and industrialized cultural production. Developed through a large interdisciplinary effort involving ethnomusicologists, anthropologists, statisticians, and computer programmers, Cantometrics transformed qualitative recordings of musical performances into systematically comparable cross-cultural measures. From the perspective of economics, the resulting data can be understood as a structured measure of expressive behavior, analogous in spirit to how the Ethnographic Atlas captures variation in social organization (Murdock, 1967) and the Folklore catalogue captures variation in narratives (Berezkin, 2015; Michalopoulos and Xue, 2021).[2]

[2]More broadly, Cantometrics follows a data-generating logic shared by other foundational cross-cultural datasets. Ethnographers conduct in-depth fieldwork on individual societies, producing detailed qualitative

2.2. What the Data Measure

Cantometrics encodes the formal properties of musical performance along a large set of dimensions that together characterize how songs are produced within a society. The coding framework is explicitly multidimensional, capturing in an integrated way the organization of performers, the structure of the musical signal, and the stylistic features of vocal delivery. The dataset includes measures describing how performances are organized, such as whether singing is centered around a dominant soloist or a group, the extent to which contributions are independent or tightly interlocked, and how vocal and instrumental parts relate to one another. These organizational dimensions are complemented by detailed measures of musical structure, including melodic properties such as shape, range, interval size, and phrase organization, as well as rhythmic properties such as overall rhythmic schemes and the relationships across simultaneous parts.

The coding further captures the degree of synchronization and integration across performers, both tonally and rhythmically, alongside a rich set of features describing vocal production itself, including tempo, volume, ornamentation (e.g., melisma, glissando, tremolo), pitch register, nasality, rasp, and articulation. Importantly, all variables are constructed from listening to recorded performances and do not rely on knowledge of language or meaning. While some measures refer to properties of the text, these are evaluated purely on acoustic grounds. As a result, Cantometrics provides a structured representation of *how* music is performed rather than *what* is being communicated, yielding a high-dimensional profile of recurrent, group-level patterns of expressive behavior that is well suited for comparative analysis across societies.[3]

2.3. Units of Observation and Coverage

The primary unit of observation in Cantometrics is the ethnic group, each associated with one or more recorded performances of traditional songs from which the Cantometric traits are coded. These groups can be linked to geographic coordinates and, in many cases, matched to standard ethnographic datasets such as the Ethnographic Atlas, thereby enabling

accounts, which are subsequently codified into structured variables by researchers such as Murdock for the Ethnographic Atlas and Berezkin for the Folklore Catalogue. Cantometrics applies the same principle to a different type of primary material: rather than textual descriptions, Lomax and his collaborators systematically code archival recordings of traditional songs. In all cases, decentralized and context-rich observations are transformed into comparable cross-cultural measures.

[3]Because the coding relies on observable features of performance rather than interpretation of meaning, the resulting measures are transparent and replicable. At the same time, the use of a common protocol across societies ensures comparability, allowing the dataset to be used for systematic cross-cultural analysis despite the heterogeneity of the underlying source material.

integration with a wide range of variables describing social organization, economic practices, and political institutions.

With 825 groups, Cantometrics achieves broad geographic coverage across all inhabited continents.[4] Figure 1 illustrates the spatial distribution of sampled groups, highlighting the wide dispersion of observations across regions. On average, each society is represented by 6 recorded songs (median: 4), and exhibits a high degree of variation across coded dimensions, with 79 distinct musical features on average (median: 78). These summary statistics underscore both the depth of the musical coding and the richness of cross-cultural variation captured in the data.

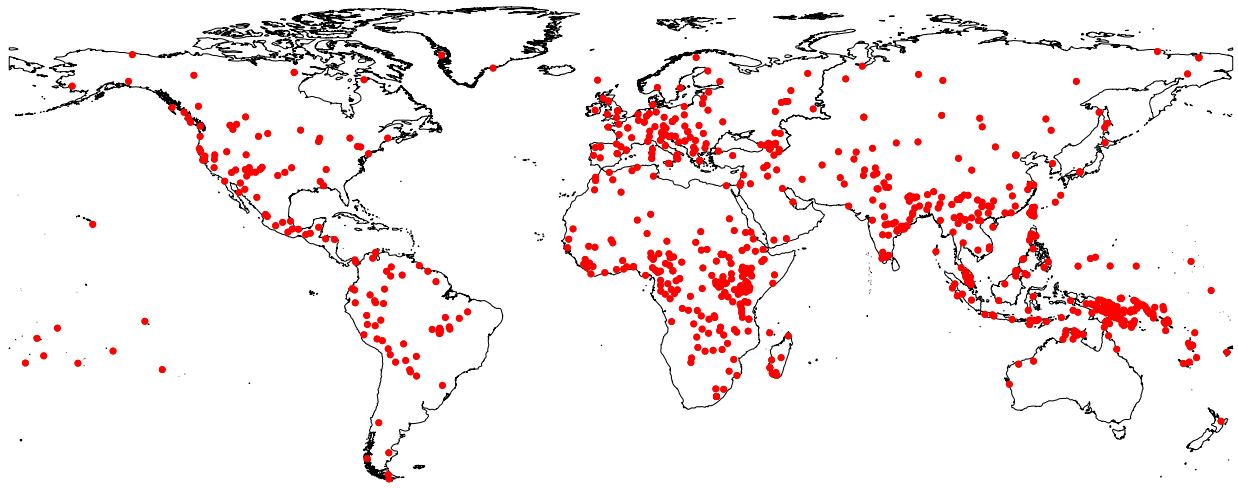


Figure 1: Global Distribution of Cantometric Groups

Notes: *The figure displays the geographic location of the 825 ethnic groups included in the Cantometrics dataset. Each point corresponds to a group for which traditional songs were recorded and coded. The map highlights the broad spatial coverage of the dataset across all continents.*

3. Music Embedded in the Process of Cultural Evolution

This section examines whether musical traditions encode information about broader dimensions of cultural variation. The central premise is that if music is embedded within a society's cultural system—reflecting dimensions such as modes of subsistence, social organization, or institutional structures—then musical traits should contain systematic information about these ethnographic characteristics, thereby validating Cantometrics as a useful data

[4]The original Cantometrics compilation contains 1,024 societies. Of these, 88 correspond to non-indigenous groups and are therefore excluded. From the remaining 936 societies, I harmonize the data into a consistent sample of 825 groups with known geographic coordinates and reliable linguistic classification. This harmonization builds directly on the important contribution of Wood et al. (2022), who digitized, systematically organized, and publicly released the original Cantometrics materials while assigning Glottolog classifications to the underlying musical traditions.

source for studying cultural evolution and its implications for contemporary social structure and economic development.

3.1. Predictive Content of Musical Traits

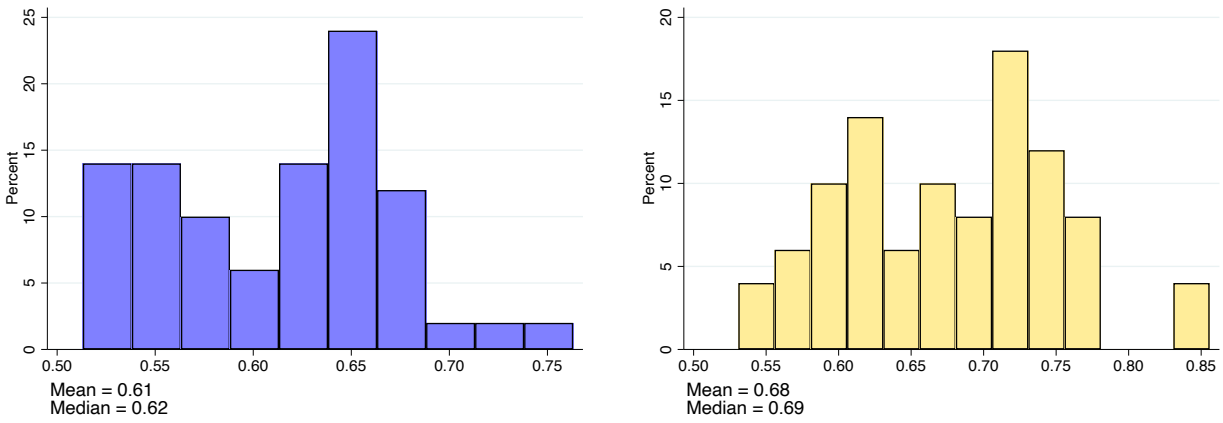
To evaluate whether musical traditions encode broader cultural information, I implement a supervised machine-learning approach that maps musical features from the Cantometrics dataset to ethnographic traits from the Ethnographic Atlas (Wainstock, 2026). Each society is represented by a high-dimensional vector of musical characteristics, capturing variation in rhythm, melody, vocal style, and performance structure. For each ethnographic outcome, I estimate models that use these musical features as predictors, with performance evaluated out-of-sample using repeated train–test splits.

Given the heterogeneity in coding across ethnographic traits, outcomes are harmonized into binary indicators to ensure comparability across domains. I focus on balanced accuracy and the area under the receiver operating characteristic curve (ROC–AUC) as summary measures of predictive performance. Balanced accuracy is particularly well suited to this setting, as it accounts for potential class imbalance across outcomes, while ROC–AUC captures the model's ability to separate societies with different underlying characteristics.

Figure 2 summarizes predictive performance across a broad set of ethnographic traits. The results indicate that musical features systematically predict ethnographic characteristics at levels well above naive benchmarks, with median performance around 0.62 for balanced accuracy and 0.69 for ROC–AUC. This pattern holds across a wide range of domains, including economic organization, social structure, and political institutions, indicating that the relationship is not confined to a narrow subset of variables.

These findings show that musical traditions are not orthogonal to other dimensions of cultural variation. Importantly, they also indicate that Cantometrics is sufficiently informative to capture this structure in the data. While musical traditions may co-evolve with other cultural characteristics, this relationship would not be empirically detectable if musical features were measured with substantial noise. The observed predictive performance therefore suggests that Cantometrics provides a meaningful representation of musical traditions, capable of revealing their connection to broader cultural dimensions. This evidence supports the interpretation of traditional songs as part of the broader process of cultural evolution, in which different dimensions of culture co-vary and are jointly shaped by shared historical and social forces. The key implication is that musical data contain meaningful signal about broader cultural characteristics, providing a novel empirical window into cultural evolution.

3.2. Musical and Ethnographic Distance



(a) *Balanced accuracy* (b) *AUC-ROC*

Figure 2: Model Performance: Ethnographic Atlas outcomes

Notes: *This figure summarizes predictive performance across Ethnographic Atlas (EA) outcomes. The panels report the distribution of (a) balanced accuracy and (b) ROC-AUC across outcomes, computed on held-out societies. Each outcome's performance is averaged over repeated random train–test splits and hyperparameters are tuned within the training sample. Balanced accuracy equals the average of sensitivity and specificity and is robust to class imbalance.*

This subsection builds on the previous analysis by examining whether similarity in musical traditions corresponds to similarity in broader ethnographic characteristics across societies. While the predictive exercise shows that musical traits contain systematic information about individual ethnographic outcomes, it does not directly speak to how societies relate to one another in the joint distribution of cultural attributes. I therefore shift to a dyadic perspective, constructing measures of distance in both musical and ethnographic space to assess whether societies that are closer in their musical traditions are also closer in broader cultural dimensions.

For musical distance, I use the pairwise distance measure constructed from the Cantometrics musical-feature vectors. Let M_i denote the vector of musical traits associated with society i, where each element indicates the presence or absence of a given musical feature. The musical distance between societies i and j is defined using the Jaccard metric:

$$d_{ij}^{M} = 1 - \frac{\sum_k \mathbf{1}\{M_{ik} = 1 \text{ and } M_{jk} = 1\}}{\sum_k \mathbf{1}\{M_{ik} = 1 \text{ or } M_{jk} = 1\}}.$$

This measure captures the extent to which two societies share musical features relative to the total set of features present in either society, and ranges from zero (identical musical profiles) to one (no shared features).

I construct an analogous distance measure using ethnographic characteristics from the Ethnographic Atlas. Let E_i denote the vector of ethnographic traits for society i.[5] Because EA variables are measured on different scales, I standardize each trait across societies and compute pairwise distance using the cosine metric:

$$d_{ij}^{EA} = 1 - \frac{\sum_{k \in K_{ij}} \widetilde{E}_{ik} \widetilde{E}_{jk}}{\sqrt{\sum_{k \in K_{ij}} \widetilde{E}_{ik}^2} \sqrt{\sum_{k \in K_{ij}} \widetilde{E}_{jk}^2}}.$$

I then relate ethnographic distance to musical distance across society pairs. The estimating equation is

$$d_{ij}^{EA} = \alpha + \beta d_{ij}^{M} + \mu_i + \mu_j + \varepsilon_{ij},$$

where μ_i and μ_j denote society fixed effects. A positive coefficient β indicates that societies that are more distant in musical space are also more distant in ethnographic space. Standard errors are clustered two-way at the level of the musical traditions i and j. In additional specifications, I include controls for pairwise characteristics, such as geodesic distance between societies.

The results show a positive association between musical distance and ethnographic distance. This pattern indicates that societies farther apart in their musical traditions are also farther apart in broader ethnographic characteristics. As in the predictive exercise, the result speaks both to the cultural content of music and to the quality of Cantometrics as a data source. Even if musical traditions co-evolved with other dimensions of culture, this relationship would not be empirically detectable if the musical-feature codings were dominated by noise or failed to capture systematic variation across societies. The observed alignment between musical and ethnographic distance therefore suggests that Cantometrics captures deeper cultural differences rather than idiosyncratic stylistic variation.

[5] For each pair (i, j), the distance is computed using only the subset of traits observed for both societies. I verify that the results are robust to imposing stricter minimum-overlap requirements on the number of shared traits (e.g., 5, 10, or 20).

Table I: Musical Distance and Ethnographic Distance

	Musical distance	
	(1)	(2)
Ethnographic distance	0.021***	0.024***
	(0.0020)	(0.0016)
Ethnicity FE		✓
Dep. var. mean	0.66	0.66
Observations	187578	187576
Adjusted R^2	0.0023	0.74

*Notes: This table reports the relationship between musical distance and ethnographic distance across pairs of societies. Musical distance is constructed from Cantometric musical-feature vectors using the Jaccard metric, while ethnographic distance is computed from standardized Ethnographic Atlas traits using cosine distance. Column (1) reports the baseline specification, while column (2) includes society fixed effects for both members of the dyad. Standard errors, reported in parentheses, are two-way clustered at the level of the musical traditions. *** Significant at the 1 percent level. ** Significant at the 5 percent level.* Significant at the 10 percent level.*

4. Musical Distance and Contemporary Social Networks

This section examines whether ethnic groups with more similar musical traditions exhibit stronger contemporary social and economic connectedness. The analysis combines measures of social ties from Facebook friendship networks and business interactions from GDELT with the GREG database of ethnic homelands. The interpretation throughout is that musical distance captures broader dimensions of long-run cultural relatedness across populations, rather than narrowly defined differences in musical style.

4.1 Social and Economic Connectedness

The Social Connectedness Index (SCI) measures the relative frequency of Facebook friendship links between geographic units (Bailey et al., 2018). Because relative differences in the index correspond to relative differences in friendship links, the SCI provides a natural measure of the intensity of bilateral social connectedness across space. A central advantage of the SCI for the purposes of this paper is its ability to capture realized patterns of interpersonal interaction at a global scale.

To link contemporary social connectedness to ancestral cultural distance, I combine the SCI with the GREG database of ethnic homelands. GREG provides spatial polygons for

ethnolinguistic homelands (Weidmann et al., 2010), allowing contemporary administrative regions to be mapped into the territories historically associated with ethnic groups. I first construct a transition matrix from SCI administrative units to GREG homelands. Each element of this matrix records the share of an administrative region that overlaps with a given ethnic homeland. I then use this matrix to aggregate SCI links between administrative-region pairs into predicted links between GREG homeland pairs.

To complement this measure of interpersonal connectedness, I also construct a measure of economic connectedness using the Global Database of Events, Language and Tone (GDELT) (Leetaru and Schrodt, 2013). GDELT records worldwide events from news reports and other online sources and identifies both the type of interaction and the geographic location of the actors involved. Following Campante and Yanagizawa-Drott (2018), I focus on events classified as material cooperation and restrict attention to events in which at least one actor is categorized as a multinational corporation (MNC) or business (BUS). These events capture observed forms of bilateral economic interaction, including partnerships, agreements, investments, and other cooperative business activities.

I map the coordinates of each actor into GREG ethnic homelands. This procedure allows the construction of a bilateral network of business interaction between ethnic groups. Specifically, for each pair of ethnic homelands, economic connectedness is defined as an indicator for whether at least one business-related event is observed during the sample period. This measure captures the existence of realized economic interaction across groups and provides a complementary outcome to the SCI. While the SCI measures interpersonal social ties, the GDELT-based measure captures observed forms of intergroup economic cooperation.

4.2 Matching Ethnic Homelands to Musical Traditions

The next step links GREG ethnic groups to Cantometric musical traditions. I follow the matching strategy commonly used in the literature on historical ethnographic data, prioritizing matches according to their degree of ethnolinguistic proximity. Exact ethnonym matches are used whenever possible. When an exact match is unavailable, groups are matched based on alternative ethnonyms, dialect affiliation, or membership in the same broader linguistic cluster.

In cases where the GREG category reflects a modern national identity rather than a premodern ethnolinguistic group, I additionally adjust the mapping using the Putterman and Weil ancestry matrix (Putterman and Weil, 2010). This procedure assigns contemporary national populations to ancestral source populations and therefore avoids treating modern national labels as culturally homogeneous historical groups.

4.3 Empirical Strategy

The empirical specification relates contemporary connectedness between ethnic homelands to musical distance:

$$Connectedness_{ijc_1c_2} = \beta MusicalDistance_{ij} + X'_{ijc_1c_2}\Gamma + \mu_{ic_1} + \mu_{jc_2} + \varepsilon_{ijc_1c_2},$$

where $Connectedness_{ijc_1c_2}$ denotes either the social connectedness or economic connectedness between ethnic homeland i in country c_1 and ethnic homeland j in country c_2. $MusicalDistance_{ij}$ is the Jaccard distance between the Cantometric musical profiles associated with ethnic homelands i and j. $X_{ijc_1c_2}$ denotes bilateral controls, such as folkloric, genetic, geodesic, linguistic, and religious distance. The terms μ_{ic_1} and μ_{jc_2} denote origin and destination ethnic-homeland fixed effects. Standard errors are two-way clustered at the musical-tradition level.

4.4 Main Findings

Table II reports the relationship between musical distance and contemporary connectedness across ethnic homeland pairs. Panel A examines interpersonal social ties using the Social Connectedness Index, while Panel B considers economic connectedness measured through bilateral business interactions from GDELT events.

The specifications progressively absorb increasingly demanding sources of heterogeneity. Columns (1) and (5) report the unconditional relationship between musical distance and contemporary connectedness. Columns (2) and (6) introduce musical-tradition fixed effects, comparing pairs within the same musical traditions. Columns (3) and (7) instead introduce ethnic-homeland fixed effects, comparing connectedness across homeland pairs while absorbing time-invariant characteristics specific to each ethnic homeland. Finally, columns (4) and (8) include country-specific ethnic-homeland fixed effects, exploiting variation across ethnic homelands nested within contemporary national boundaries.

Across all specifications, greater musical distance is associated with significantly lower contemporary connectedness. The estimates are remarkably stable across alternative fixed-effect structures, suggesting that the relationship is not driven solely by broad regional or national patterns. The results therefore indicate that the cultural signal embedded in traditional songs is strongly associated with the structure of contemporary social and economic interaction across groups.

4.5 Alternative Measures of Cultural Distance

Table II: Musical Distance and Contemporary Connectedness

	Social Connectedness				Economic Connectedness			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Musical distance	−0.74***	−3.71***	−3.69***	−3.65***	−0.015***	−0.020***	−0.020***	−0.019***
	(0.24)	(0.33)	(0.33)	(0.32)	(0.0034)	(0.0027)	(0.0027)	(0.0026)
Musical tradition FE		✓				✓		
Ethnicity FE			✓				✓	
Country-Ethnicity FE				✓				✓
Observations	1,282,900	1,282,899	1,282,899	1,282,898	1,500,829	1,500,828	1,500,827	1,500,827
Dep. var. mean	3.99	3.99	3.99	3.99	0.0048	0.0048	0.0048	0.0048
Adjusted R^2	0.0023	0.20	0.23	0.29	0.00077	0.030	0.041	0.087

*Notes: This table reports the relationship between musical distance and contemporary connectedness across ethnic homeland pairs. Columns (1)–(4) use the Social Connectedness Index, while columns (5)–(8) use a measure of bilateral business interaction constructed from GDELT events. Musical distance is measured using the Jaccard distance between Cantometric musical profiles. Specifications progressively introduce fixed effects for musical traditions, ethnic homelands, and country-specific ethnic homelands. Standard errors, reported in parentheses, are two-way clustered at the musical-tradition level. *** Significant at the 1 percent level. ** Significant at the 5 percent level.* Significant at the 10 percent level.*

Table III compares musical distance to alternative measures of cultural distance constructed from domains such as linguages (Desmet et al., 2009), folklore (D'Amato and Russo, 2026), and religion. The objective of this exercise is not to interpret these measures as competing explanations for contemporary connectedness. Rather, the analysis builds on the evidence from Sections 2 and 3 showing that traditional songs encode broader dimensions of cultural similarity across societies. Musical distance is therefore interpreted as an observable manifestation of deeper cultural differences.

The comparison instead evaluates whether the cultural signal embedded in traditional songs remains informative alongside alternative proxies for long-run cultural relatedness. Because these measures capture overlapping dimensions of cultural proximity, conditioning on additional cultural distances mechanically absorbs part of the variation contained in musical traditions. Consequently, attenuation in coefficient magnitudes or statistical significance is not, by itself, informative about the relevance of music. The relevant question is whether musical distance continues to retain explanatory content once other manifestations of cultural similarity are taken into account.

The results show that musical distance remains negatively associated with both social and economic connectedness even after accounting for alternative cultural distances. At the same time, the estimates also reveal substantial overlap between musical distance and other measures of long-run cultural similarity, particularly linguistic and folkloric distance. This pattern is consistent with the interpretation advanced throughout the paper: traditional

songs do not merely capture musical styles narrowly defined, but instead encode broader dimensions of cultural evolution across societies.

Table III: Musical Distance, Alternative Cultural Distances, and Contemporary Connectedness

	A. Social Connectedness								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Musical distance	−3.58***		−2.44***	−3.66***		−2.26***	−3.62***		−3.39***
	(0.32)		(0.23)	(0.32)		(0.26)	(0.37)		(0.37)
Linguistic distance		−2.25***	−2.16***						
		(0.13)	(0.12)						
Folkloric distance					−11.7***	−11.3***			
					(0.60)	(0.59)			
Religious distance								−0.75***	−0.72***
								(0.060)	(0.060)
Observations	1,266,972	1,266,972	1,266,972	1,268,530	1,268,530	1,268,530	694,272	694,272	694,272
Dep. var. mean	3.98	3.98	3.98	4.00	4.00	4.00	4.02	4.02	4.02
Adjusted R^2	0.29	0.36	0.36	0.29	0.39	0.40	0.29	0.30	0.31
	B. Economic Connectedness								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Musical distance	−0.018***		−0.013***	−0.020***		−0.0092***	−0.022***		−0.020***
	(0.0026)		(0.0025)	(0.0026)		(0.0023)	(0.0032)		(0.0030)
Linguistic distance		−0.010***	−0.010***						
		(0.0013)	(0.0013)						
Folkloric distance					−0.090***	−0.088***			
					(0.0099)	(0.0099)			
Religious distance								−0.0055***	−0.0053***
								(0.00080)	(0.00078)
Observations	1,483,223	1,483,223	1,483,223	1,445,963	1,445,963	1,445,963	812,977	812,977	812,977
Dep. var. mean	0.0047	0.0047	0.0047	0.0049	0.0049	0.0049	0.0066	0.0066	0.0066
Adjusted R^2	0.088	0.090	0.090	0.090	0.096	0.096	0.10	0.11	0.11

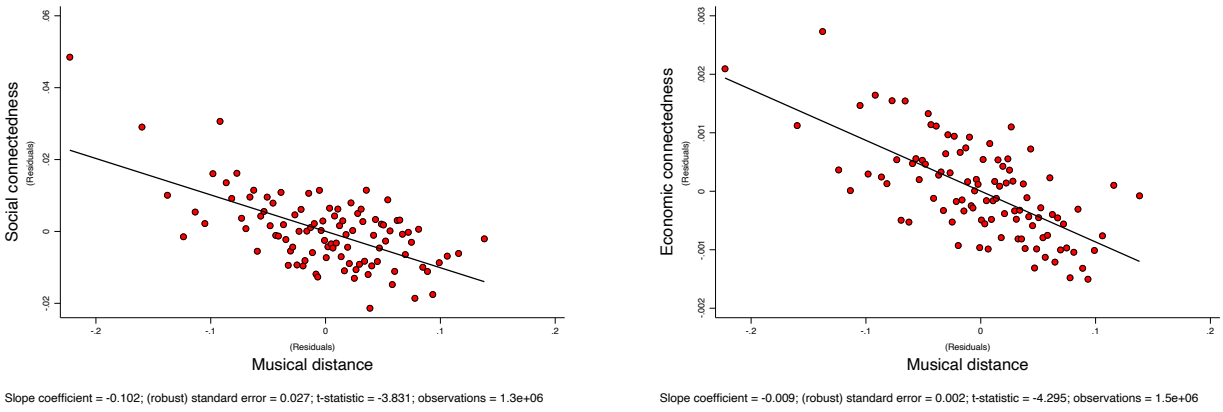
*Notes: This table compares the relationship between contemporary connectedness and alternative measures of cultural distance across ethnic homeland pairs. Panel A uses the Social Connectedness Index, while Panel B uses bilateral business interaction constructed from GDELT events. Musical distance is measured using Cantometric musical profiles, linguistic distance captures differences across language families, folkloric distance is constructed from the distribution of traditional narrative motifs, and religious distance is computed from differences in the distribution of religious denominations using data from the World Religion Database. All specifications include country-specific ethnic-homeland fixed effects. Standard errors, reported in parentheses, are two-way clustered at the musical-tradition level. *** Significant at the 1 percent level. ** Significant at the 5 percent level.* Significant at the 10 percent level.*

4.6 Within-Country-Pair Comparisons

Contemporary country borders are themselves shaped by deep cultural forces. As a result, country-pair fixed effects are not orthogonal to the mechanisms studied in this paper;

rather, they are endogenous to cultural differences across ethnic groups. Absorbing bilateral country-pair effects therefore removes part of the relationship between long-run cultural similarity and contemporary connectedness identified in the baseline estimates, and is consequently expected to attenuate the estimated relationship between musical distance and connectedness. Nevertheless, examining whether the relationship survives under such circumstances remains informative about the extent to which cultural similarities embedded in traditional songs explain finer-grained patterns of connectedness across groups.

Figure 3 presents the corresponding estimates. As expected, the coefficients are attenuated relative to the baseline specifications. Nevertheless, the negative relationship between musical distance and both social and economic connectedness remains economically and statistically significant even within country pairs. Ethnic homelands with more similar musical traditions exhibit stronger friendship links and business interactions even after absorbing all country-pair-level variation. These findings therefore illustrate the extent to which the cultural similarities embedded in traditional songs continue to explain the structure of contemporary global connectedness even under substantially more restrictive comparisons.



(a) *Social connectedness* (b) *Economic connectedness*

Figure 3: Musical Distance and Connectedness within Country Pairs

Notes: *This figure plots the residual relationship between musical distance and contemporary connectedness after absorbing country-pair fixed effects and baseline controls. Panel (a) uses the Social Connectedness Index, while Panel (b) uses bilateral business interaction constructed from GDELT events. Musical distance is measured using the Jaccard distance between Cantometric musical profiles. Each point corresponds to a bin of the residualized data, and the fitted line is estimated from the underlying bilateral observations. The negative relationship in both panels indicates that ethnic homelands with more similar musical traditions exhibit stronger contemporary social and economic connectedness even within the same pair of contemporary countries.*

5. Concluding Remarks

This paper introduces the Cantometrics catalogue into economics as a new source of cross-cultural data on traditional music. The catalogue provides systematic information on the structure and stylistic characteristics of songs across hundreds of societies worldwide, allowing the construction of comparable measures of musical similarity and distance across populations. Despite the centrality of music in human societies, traditional musical expression has remained largely absent from empirical economics. This paper demonstrates how musical traditions can be incorporated into quantitative research on culture and long-run development.

The analysis provides evidence that musical traditions encode systematic information about broader dimensions of culture. If music is embedded within a society's cultural system—reflecting dimensions such as modes of subsistence, social organization, or institutional structures—then musical traits should contain information about these broader ethnographic characteristics. Consistent with this view, the paper demonstrates that musical features have substantial predictive power with respect to ethnographic traits from the Ethnographic Atlas. Moreover, societies that are closer in musical space are also systematically closer in ethnographic space, indicating that musical traditions co-vary with broader dimensions of cultural variation across populations.

The paper further shows that societies characterized by greater musical similarity exhibit substantially stronger contemporary social connectedness and economic connectedness. This relationship emerges consistently across specifications and remains robust to the inclusion of country-pair fixed effects and alternative measures of cultural distance. Importantly, the objective of this analysis is not to interpret music as an isolated determinant of connectedness, nor to evaluate competing cultural measures against one another. Rather, the findings indicate that traditional songs contain persistent information about cultural proximity across populations, and that this information remains informative even alongside other manifestations of long-run cultural relatedness. More generally, the results suggest that expressive cultural forms may encode broader dimensions of cultural evolution that continue to shape patterns of social and economic interaction over long time horizons.

The paper also highlights the potential of expressive cultural forms as measurable objects of economic analysis. Traditional songs historically constituted important mechanisms through which societies transmitted norms, identities, collective memory, and shared modes of behavior across generations. The Cantometrics catalogue offers unusually broad geographic coverage and rich internal variation, creating opportunities for future work across several areas of economics, including cultural transmission, migration, and identity forma-

tion. More generally, the paper illustrates how systematically coded forms of expressive culture can expand the empirical toolkit available for studying the historical and cultural foundations of economic and social outcomes.

References

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018): "Social connectedness: Measurement, determinants, and effects," *Journal of Economic Perspectives*, 32, 259–280.

Berezkin, Y. (2015): "Folklore and mythology catalogue: its lay-out and potential for research," *The Retrospective Methods Network*, 58–70.

Blacking, J. (1973): *How musical is man?*, University of Washington Press.

Campante, F. and D. Yanagizawa-Drott (2018): "Long-range growth: economic development in the global network of air links," *The Quarterly Journal of Economics*, 133, 1395–1458.

D'Amato, M. and F. F. Russo (2026): "Cultural doorways in the barriers to development," *Journal of Economic Growth*, 31, 125–178.

Desmet, K., S. Weber, and I. Ortuño-Ortín (2009): "Linguistic diversity and redistribution," *Journal of the European Economic Association*, 7, 1291–1318.

Kampanelis, S., A. Elizalde, and Y. M. Ioannides (2023): *Songlines*, Queen's University Centre for Economic History.

Leetaru, K. and P. A. Schrodt (2013): "Gdelt: Global data on events, location, and tone, 1979–2012," in *ISA annual convention*, Citeseer, vol. 2, 1–49.

Lomax, A. (1962): "Song structure and social structure," *Ethnology*, 1, 425–451.

Lomax, A. (2017): *Folk song style and culture*, Routledge.

Michalopoulos, S. and M. M. Xue (2021): "Folklore," *The Quarterly Journal of Economics*, 136, 1993–2046.

Murdock, G. P. (1967): "Ethnographic atlas: a summary," *Ethnology*, 109–236.

Putterman, L. and D. N. Weil (2010): "Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 125, 1627–1682.

Wainstock, D. C. (2026): "A Culture of Human Capital Formation in Africa," *Mimeo, University of Oxford, Department of Economics*.

Weidmann, N. B., J. K. Rød, and L.-E. Cederman (2010): "Representing ethnic groups in space: A new dataset," *Journal of Peace Research*, 47, 491–499.

Wood, A. L., K. R. Kirby, C. R. Ember, S. Silbert, S. Passmore, H. Daikoku, J. McBride, F. Paulay, M. J. Flory, J. Szinger, et al. (2022): "The Global Jukebox: A public database of performing arts and culture," *PloS one*, 17.