Imagine All the People: Characterizing Social Music Sharing on Reddit

Veniamin Veselovsky, Isaac Waller, Ashton Anderson
University of Toronto
venia@cs.toronto.edu, walleris@cs.toronto.edu, ashton@cs.toronto.edu

Abstract

Music shapes our individual and collective identities, and in turn is shaped by the social and cultural contexts it occurs in. Qualitative approaches to the study of music have uncovered rich connections between music and social context but are limited in scale, while computational approaches process large amounts of musical data but lack information on the social contexts music is embedded in. In this work, we develop a set of neural embedding methods to understand the social contexts of online music sharing, and apply them to a novel dataset containing 1.3M instances of music sharing in Reddit communities. We find that the patterns of how people share music in public are related to, but often differ from, how they listen to music in private. We cluster artists into *social genres* that are based entirely on aggregate sharing patterns and reflect where artists are invoked. We also characterize the social and cultural contexts music is shared in by measuring associations with social dimensions such as age and political affiliation. Finally, we observe that a significant amount of sharing is attributable to extra-musical factors—additional meanings that people have associated with songs. We develop two methods to quantify the extra-musicality of music sharing. Our methodology is widely applicable to the study of online social contexts, and our results reveal novel cultural associations that contribute to a better understanding of the online music ecosystem.

Introduction

Music binds us together. It is universal, being present in every observed society (Mehr et al. 2019); it spans the full range of the human experience (Blacking and Nettl 1995); and it is a critical component in the formation and maintenance of social identity (Tarrant, North, and Hargreaves 2002; Gregory 1997). As the sociomusicologist Simon Frith contends, "Social groups... only get to know themselves *as groups* through cultural activity." (Frith 1996).

Several bodies of literature in the social sciences have been dedicated to understanding the social and cultural contexts of music. Ethnographic, sociological, and anthropological studies have documented a plethora of ways in which music is linked to social constructs including gender, age, education level, class, and politics. These approaches typically focus on a small number of examples, either of communities or pieces of music, in order to deeply understand the social and cultural contexts they are embedded in.

Computational approaches to the study of music, on the other hand, have largely been dedicated to processing musical data. Millions of songs can be categorized into fine-grained musical genres using automated and semi-automated techniques, signal processing methods can perform complex tasks on acoustic data, and higher-level semantic qualities of songs can be extracted from the music alone. However, the *social contexts* in which music is invoked has not been subject to the same level of analysis. Detailed audio signals corresponding to millions of pieces of music are readily available, but it has been more difficult to construct large-scale datasets corresponding to the social and cultural contexts of music.

There is thus a gap between these two literatures: qualitative investigations of music are rich but limited to relatively few artists and communities, and computational studies of music are largely lacking social and cultural context. In this work, we introduce a computational methodology for measuring the social and cultural contexts of large-scale online music sharing, and apply our methodology to a novel social music sharing dataset. We study invocations of music on Reddit, one of the world's largest online social platforms, in which people publicly discuss various topics in thousands of communities. Reddit is an ideal environment for our analysis. It is a public forum with millions of instances of people sharing music, which enables a large-scale study. Crucially, we can measure the social and cultural contexts of this sharing by applying recently-developed neural embedding methods to characterize subreddits along social dimensions such as age, gender, and political partisanship. For each musical artist that is shared on Reddit sufficiently often, we can then measure the social and cultural orientations attached to them by aggregating over the subreddits they are shared in. In this way, we develop an understanding of the social contexts of musical artists by harnessing large-scale patterns in how they are shared by listeners online.

We focus on three main areas of interest. First, we cluster artists together based on how they are shared into what we term "social genres". In contrast with conventional musical genres, these groups are *socially constructed* in that they are collections of musical artists that are similar in how they are shared online. The similarities and differences between how musical artists are invoked in public discussions are related to, but often distinct from, their musical similarities and differences. In many cases, musically dissimilar artists appear in similar social contexts and are thus grouped together by our method.

Second, we infer social and cultural contexts associated with musical artists by harnessing information about the communities in which they are shared. By building on previous work using neural community embeddings to identify cultural axes of meaning in Reddit, we quantify the cultural contexts artists are shared in. For example, we can quantify if a particular artist is disproportionately shared in left-wing communities, or in communities with a younger demographic. Furthermore, we compare our social and cultural associations for all musical artists in our dataset, providing a comprehensive, macro-scale view of music sharing on Reddit. These features can act as a new source of information for both artists and researchers to deepen their understanding of the online music ecosystem.

Finally, we find evidence that a significant amount of online music sharing is driven by extra-musical factors—many songs or pieces of music acquire additional social meaning online. One manifestation of this is when an artist becomes associated with meme culture. We develop methods to measure the extent to which the social sharing of music is extra-musical. These extra-musicality scores could be valuable to designers of recommendation systems, as current state-of-the-art recommendation algorithms are typically blind to social context. Our methods are one approach to determining whether there is additional social or cultural context driving the consumption of particular music. Taken together, our methodology and results provide a framework for studying music by understanding how it is shared in online platforms.

Related Work

Our work draws upon several lines of research that have laid the theoretical and empirical foundations of the social, cultural, and societal contexts of music sharing. We detail three particularly important existing directions here: classic work elucidating the many social contexts that music is embedded in; theories of "extra-musicality", in which music acquires additional meaning beyond its formal musical components; and empirical studies of online music sharing.

Social contexts of music. Music is a key component of social identity (Frith 1996). Vast literatures spanning musicology, sociology, and anthropology are dedicated to understanding the social contexts of music, which we can only briefly survey here. Bourdieu introduced a foundational theory relating social contexts to cultural behavior. In his framework, cultural practices including musical preferences are not only tightly linked to social contexts, but are a product of them (Bourdieu 1984). He documented many examples of how musical taste reflects and reproduces social stratification, and is strongly correlated with indicators of class such as education (Prior 2013). Bourdieu's findings, based on French society in the 1960s, were largely replicated and extended to further demographic variables, including age and gender, in an analysis of British society in the 2000s (Bennett et al. 2009). Subsequent lines of inquiry building upon this foundation have developed our understanding of how music relates to all manner of cultural concepts. We focus in particular on how music is used as a political object, how it manifests gender divides, and how it acts as a signal of affluence, but important work also addresses how music structures and affirms other sociocultural differences, *e.g.* ethnicity (Stokes 1997), national identity (Vianna 1999), and social class (Chan and Goldthorpe 2007).

Politics. Music has long been a powerful political tool and agent of change. This power has been harnessed by countries to form a sense of national identity (Colley 2005), for various propaganda purposes (Street 2003), and by revolutionary forces to rally support for their movements (Roy 2013). This function of music has been referred to as a "device of social ordering" (DeNora 2000).

Gender. The relationship between music and gender has also been deeply studied. Ethnomusicologists have shown that men and women do not "have equal access to all musical experiences and opportunities within a given society", which leads to differences in the consumption of music (Koskoff 2014). A body of fieldwork has documented how these differences play out in music production and consumption (Hamessley 2006; Magrini 2003).

Affluence. Music tastes have been shown to depend on the class of an individual, with affluence being one component of social class (Seeger 1957). The concept of "affluent music" evolves over time with changes in taste and as music that once signalled high social status becomes popularized (Rochberg 1968).

Extra-musicality. Music is laden with meaning, some of which is not clearly derived from musical form alone. A body of work is dedicated to studying these "extra-musical" associations and meanings. Extra-musical features have been shown to affect our emotional response to listening to music (Susino and Schubert 2020; Vuoskoski and Eerola 2015) and extra-musicality often differs across national boundaries (Kristen and Shevy 2013). Meyer advanced a taxonomy of three "sign qualities" along which music can communicate meaning: iconic sign qualities, indexical sign qualities, and symbolic sign qualities (Meyer 2008). Symbolic sign qualities, those which measure meaning from explicitly extra-musical associations, are of particular interest to our analysis. We propose a methodology to score songs according to how "extra-musical" their sharing patterns appear to be. Our approach of measuring extra-musicality through understanding patterns in how listeners share music responds directly to DeNora's call to analyze the "relations of production"—DeNora emphasized that it is not enough to analyze how artists and listeners interact, but also how listeners themselves interact (DeNora 1986). Our paper aims to fill this gap by proposing a methodology to understand large-scale patterns in how listeners interact by sharing music online.

Online music sharing. Finally, there are precedents to our empirical approach to understanding online music sharing. Studies of music platforms such as Spotify have investigated the dynamics of musical identity (Way et al. 2019), how diversity of consumption is affected by algorithmic recommendations (Anderson et al. 2020a), and how music preferences are associated with personality (Anderson et al. 2020b). The piracy of online music and the culture surrounding it has also attracted attention (Condry 2004). Our work also fits into a larger literature devoted to cultural transmission. In particular, the concept of horizontal cultural transmission—how cultural objects are passed between members of a similar generation—relates to our findings about age (Cavalli-Sforza and Feldman 1981). This form of horizontal cultural transmission is frequently used to study the spread of disinformation (Carrignon, Bentley, and Ruck 2019) and meme sharing (He et al. 2014).

Data

We begin with a dataset that consists of all Reddit submissions and comments from 2006 to the end of 2018, plus all comments in 2019, and aim to extract a subset of comments and submissions containing music sharing (Baumgartner et al. 2020). Naïvely searching for posts and comments containing artist or song names would be the simplest approach, but this risks including extraneous mentions of unrelated concepts that share the same terms. For example, searching for the well-known rapper Pitbull would surface many posts and comments discussing the pitbull breed of dog. For this reason, we instead search for comments and submissions that link to either a Spotify track (or album) or a YouTube video corresponding to a song. URLs are long and unique, making them accurate and robust identifiers.

While extracting posts containing Spotify links is easily achieved by searching for the `open.spotify.com` domain, the same method cannot be used for YouTube, as YouTube links point to all kinds of videos. To create a list of YouTube links that refer to a song, we first extracted all posts and submissions that contain a Spotify link. For each Spotify link, we then used the Spotify API to extract artist and track information, including genres, popularity, track name, artist name, and album name. For album shares, we collected all tracks in the album and all metadata for those tracks. In total, we collected 536,860 unique tracks from Spotify shares on Reddit. We linked each of these tracks with a YouTube video by scraping the Last.fm link associated with the Spotify track and extracting the YouTube link from the Last.fm page. Last.fm contains links to 194,067 of our tracks (36%). Some links did not align with the artist due to the scraping occasionally picking up the wrong link from the Last.fm website. For this reason, if a YouTube video was associated with more than one track, we validated the link by extracting the title of the YouTube video and verifying whether it matched the song title.

We finally combined our lists of Spotify and YouTube song URLs to select all comments and submissions on Reddit containing a link to a song or album on either service. After removing artists that were shared fewer than 20 times, as estimating artist sharing patterns with so few data points risks painting an inaccurate picture of the artist, our final dataset contains 1.3 million music shares on Reddit of 6,600 unique artists.

Characterizing Sharing with Embeddings

Community Embedding

Reddit consists of tens of thousands communities called *subreddits*, which act as forums for users to discuss various topics. Every submission and comment is associated with a single subreddit. Communities exist for a wide array of topics, ranging from sports to music, politics to mathematics, and jokes to photos. Music sharing occurs not only in music-dedicated communities but is widespread across the platform; this means that many songs are shared in many different communities. The communities in which a song is posted represents the social contexts in which it is shared.

Given the large number of Reddit communities and their unstructured nature, it is difficult to meaningfully compare sharing patterns of songs or artists directly. To understand the similarities and relationships between communities in which music is shared, we use a *community embedding*, an adaptation of word embeddings used in natural language processing. As a word embedding embeds words in a vector space such that similar words are close to each other, a community embedding embeds communities in a vector space such that communities with similar user-bases are close to each other. We use the Reddit community embedding developed by Waller and Anderson, which treats communities as words and users as contexts and applies the word2vec algorithm (Waller and Anderson 2019). Each community is represented as a 150-dimensional vector in a Euclidean space. Notably, the textual content of comments and submissions are not used in the creation of this community embedding, thus the similarity between two communities is related only to *who* comments in a community, not *what* they say. Communities with many users in common are close together in the resulting space, and communities with few users in common are further apart. This embedding achieves high accuracy on community analogy problems, meaning semantic relationships are preserved in this space. For example, simple vector arithmetic will correctly answer the analogy r/Torontobluejays $-$ r/toronto $+$ r/nba $\approx$ r/torontoraptors. A community embedding is thus a useful tool to quantify the similarity of disparate communities on the site.

Artist Embedding

We use this community embedding to create a vector for each of the artists in our dataset using the following process. Let P represent the set of all *music sharing posts* (where posts can be either comments or submissions), and let P_a denote the subset of P restricted to shares of artist a. For each $p \in P_a$, we define $\vec{p}$ to be the vector of the community in which the comment or submission was posted. An artist's vector $\vec{a}$ is then:

$$\vec{a} = \sum_{p \in P_a} \frac{\vec{p}}{|P_a|} \tag{1}$$

i.e. the arithmetic mean of the vectors of the communities where a was shared, weighted by the number of shares in each community. Importantly, an artist's position in the embedding is purely a function of *where they are shared*.

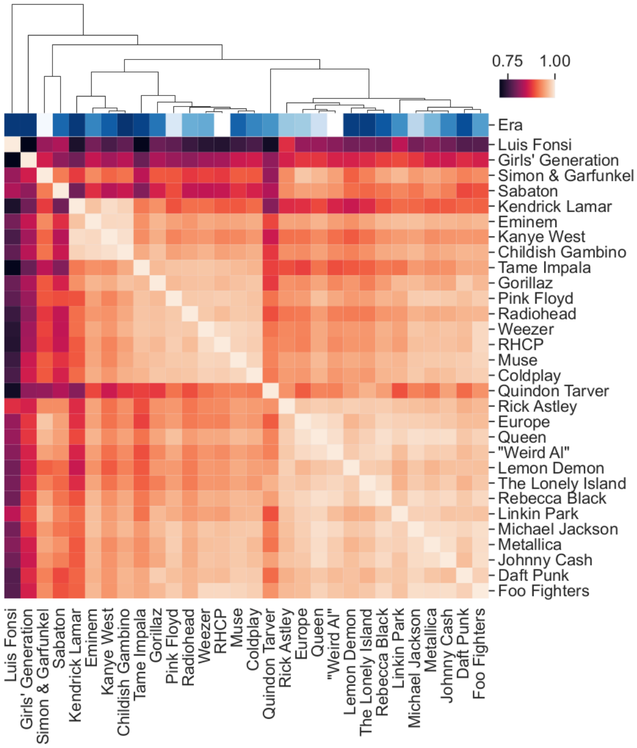

Figure 1: Heatmap of artist similarity for the top 30 most shared artists on Reddit, including a dendrogram of the clustering solution. The blue bars represent the year of release for each artist's most popular song (lightest represents 1968, darkest represents 2018).

The similarity of any two artists a_1 and a_2 can be computed using the cosine similarity $\cos(\vec{a_1}, \vec{a_2}) = \frac{\vec{a_1} \cdot \vec{a_2}}{\|\vec{a_1}\| \|\vec{a_2}\|}$, which returns a value between -1 and 1. The value is 1 if their vectors point in the same direction, 0 if they are orthogonal, and -1 if they point in opposite directions. Since artists' positions in the embedding are derived only from how they are shared, the cosine similarity between artists is a measure of similarity of sharing patterns. This similarity will be high when artists are shared in similar communities, and is therefore reflective of the social contexts in which they are invoked online.

Artist similarity. We apply our artist similarity metric to compute the similarity between all pairs of popular artists on Reddit, and visualize the similarities between the top 30 most-shared artists in Figure 1. Examining these connections reveals the structure of how popular artists are shared. Many artists are shared in clusters that reflect their traditional music genres. For instance, Kendrick Lamar, Eminem, Kanye West, and Childish Gambino all belong to the same Spotify genre, and are also mutually similar in how they are shared on Reddit.

However, other artists have virtually no Spotify genres in common, but are nevertheless similar to each other in how they are shared. Rebecca Black and the Foo Fighters are quite musically distinct, but have a very high similarity score (0.98). To understand this, we inspect their sharing patterns. For both artists, only a few of the top 20 subreddits that they are shared in are related to music. The majority of their mentions occur in non-music-related communities including r/videos, r/todayilearned, and r/funny. Rebecca Black is famous for her viral Internet hit "Friday", which is mostly invoked as a meme in response to mentions of the day of the week Friday. Foo Fighters' similar sharing is driven by the lyrics in their hit song "My Hero": "There goes my hero, watch him as he goes". This song is often invoked in response to an act perceived (either sarcastically or genuinely) as heroic. This is an example of how musically distinct artists can still have high social similarity driven by the symbolic sign qualities (*i.e.* extra-musical associations) that they have acquired.

Social Genre: *pop 1*

Spotify genres		Subreddits		Artists	
Name	Count	Name	Count	Name	Count
pop	15122	Music	3212	Taylor Swift	1733
indie pop	2802	popheads	2748	CHVRCHES	1111
dance pop	2541	AskReddit	2030	Ariana Grande	809
electropop	1043	tipofmytongue	1320	Lorde	746
rock	647	listentothis	1265	St. Vincent	725
detroit hip hop	288	indieheads	811	Grimes	661
latin	265	hiphopheads	806	Janelle Monáe	566
pop rock	207	CasualConv.	762	Kylie Minogue	565
pop rap	191	videos	480	Azealia Banks	560
modern rock	157	ifyoulikeblank	460	Kero Bonito	536

Table 1: The Spotify music genres, subreddits, and artists that occur most often within the social genre we call *pop 1*.

Social Genres

The way artists are shared is related to, but sometimes distinct from, their musical similarities and differences. To measure the macro-scale structure of music sharing on Reddit, we cluster artists based on their positions in the artist embedding. This produces groupings of artists that are similar in how they are shared, which we term *social genres*. Although conventional music genres also include social meaning in their formulation, our social genres are entirely driven by patterns of social contexts.

Creating social genres. To create social genres, we applied hierarchical clustering to all artists' vectors in the artist embedding (other clustering methods resulted in qualitatively similar groupings). As a result, each artist is placed in the social genre with the other artists whose sharing patterns are most similar to their own. We used a combination of silhouette scoring and the elbow method to determine that 50 is an appropriate number of clusters.

We named each social genre based on three properties of the artists within it: the top communities their shares appear in, their top Spotify genres, and in some cases the top artists in the genre. In Table 1, this information is shown for the social genre named *pop 1*. In this example, pop dominates the Spotify genres, r/Music and r/popheads are the main communities of transmission, and the top artists fea-

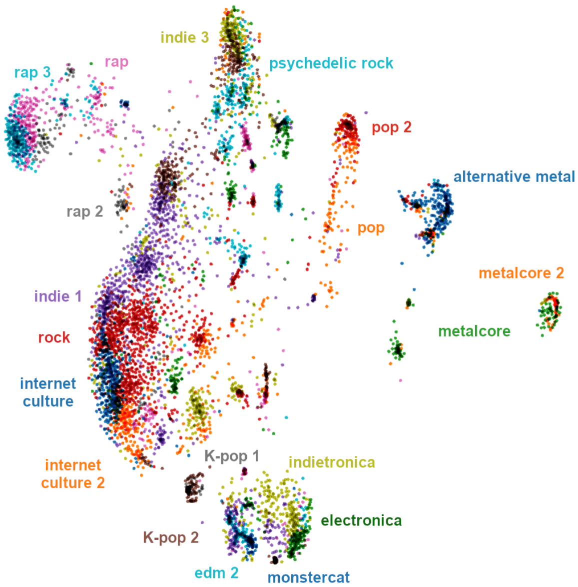

Figure 2: Projection of 6,600 artists in a 150-dimensional embedding into a 2-dimensional representation using t-SNE. Artists are represented as points and color represents the social genre an artist is in. Similarity between artists is shown through proximity.

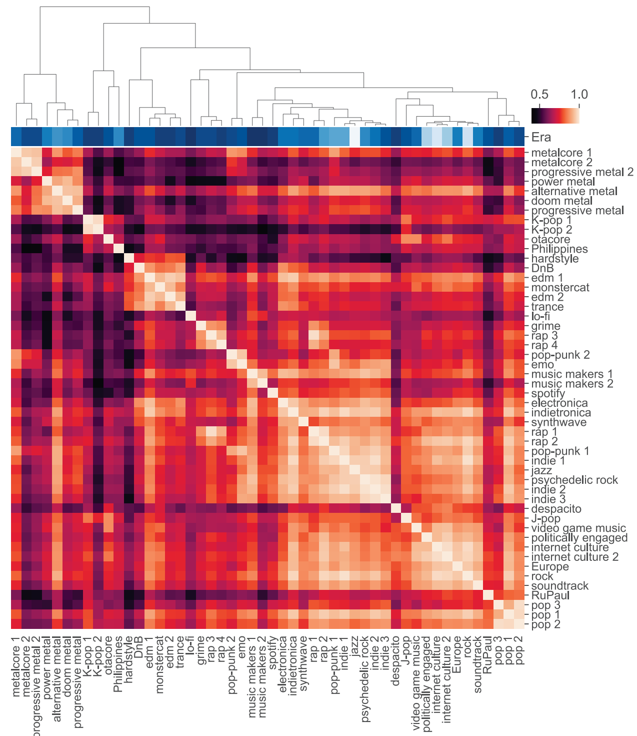

Figure 3: Heatmap of social genre similarity (as measured by cosine similarity of the genre vectors) for the top fifty social genres. The blue bars represent the average year of release of the most popular song by each artist in the genre (lightest represents 1995, darkest represents 2018).

ture pop musicians. If one Spotify genre was predominant, then that genre name was taken as the name of the social genre. Often, more than one social genre was dominated by the same top Spotify genre. These genres were named sequentially, as in *pop 1* and *pop 2*, and ordered by descending popularity (*e.g. pop 1* is the most popular pop genre). When there was a diverse array of Spotify genres present in a social genre, we conducted a more in-depth analysis by examining the communities and artists. Finally, there were a few genres that could not be understood by analyzing the artists, subreddits, and Spotify genres. These social genres capture something entirely unrelated to conventional genres. For example, the top twenty artists in *internet culture 1* include Queen, The Lonely Island, OK Go, Hans Zimmer, and Bon Jovi. An analysis of how these artists are shared reveals that they are often deployed in meme-like ways. Songs such as Rebecca Black's "Friday" or Queen's "Don't Stop Me Now" have become part of internet culture and operate mostly on a meta level.

The artist embedding is clustered into highly distinct groups, suggesting that social genres capture real differences in the social contexts that music is shared in. Figure 2 shows a 2-dimensional t-SNE projection of the artist embedding where artists are colored by social genre. As the original embedding space is 150-dimensional, this representation contains far less data than the original, but nevertheless many specific clusters are clearly identifiable. Furthermore, similar social genres are in close proximity to each other.

Analysis of social genres. Examining these social genres uncovers connections that would be difficult to observe using conventional music analyses. Social genres fall on a spectrum between two extremes: those that align with traditional musical genres, and those that are driven by additional layers of meaning. This distinction aligns with Meyer's taxonomy of iconic sign qualities and symbolic sign qualities, where social genres that align with traditional genres are more likely to be dominated by artists that share iconic sign qualities. Many social genres correspond directly with traditional music genres. Pop-punk, indie, alternative metal, metalcore, electronic dance music (EDM), lo-fi, hardstyle, and some types of pop tend to cluster together in their respective traditional genres, as most of the sharing in these categories takes place in communities dedicated to these genres. For example pop-punk has a community called r/PopPunkers, and metalcore has a community called r/Metalcore.

However, not all social genres are in clear correspondence with a traditional genre. For instance, two social genres feature creators who share their music to try to gain exposure in communities such as r/WeAreTheMusicMakers, r/ThisIsOurMusic, and r/Songwriters. Importantly, these artists are not necessarily musically similar. Instead, they are connected by sharing their music in communities dedicated to small music makers. As another example, the *RuPaul* genre is dominated by music sharing in the r/RuPaul community, which is associated with the reality TV show

"RuPaul's Drag Race", billed as the "search for America's next drag superstar".

Social genres can represent extra-musical associations as well. For example, the *despacito* social genre represents music shared for a comedic purpose, and is dominated by Luis Fonsi's hit "Despacito", which developed a meme meaning on Reddit. In contrast, the *politically engaged* social genre captures sharing from political communities on Reddit. This genre contains shares of artists who are known for political themes in their music, and are accordingly shared in political communities such as r/politics, r/LateStageCapitalism, r/conspiracy, and the now-banned r/The_Donald and r/ChapoTrapHouse. This political nature can also be observed through the artists in the social genre. The top three most shared artists are Rage Against the Machine, who express revolutionary political views; Knife Party, whose hit song "Centipede" became connected with Donald Trump; and Killer Mike, who is known for promoting social justice.

Social genres can be compared to one another to understand the overall structure of music sharing on Reddit. Figure 3 shows all pairwise similarities between the 50 social genres. Analyzing this plot reveals the emergence of several key social genre groupings. First, the metal group highlights a diverse range of metal music including metalcore, progressive metal, alternative metal, and doom metal. There is also an Asian music cluster featuring K-pop, otacore, and music from the Philippines. Electronic music comprises a distinct group, and includes various flavors of EDM such as trance and monstercat. The social genres in the electronic cluster are also highly similar to the two other electronic social genres *indietronica* and *electronica*. The large group ranging from *video game music* to *soundtrack* represents extra-musical genres that have developed new meanings on the Internet. This group features everything from "rickrolling" to *Hamilton: An American Musical*. Finally, the three pop genres form a tight cluster, and also display high similarity with the indie and rap clusters. The full dataset of social genres is available in our data repository[1].

Linguistic Analysis

A useful property of social music sharing is that it is embedded in discussion. As such, it is surrounded by language that can shed light on music sharing patterns. To associate artists with language, we embed words used on Reddit in much the same way that we embed artists. By analyzing the relationships between the positions of artists and words in the embedding, we can learn how words are associated with both artists and social genres.

Embedding words. We embed words analogously to how we embed artists. Instead of creating a new embedding, we co-embed words into the artist embedding by assigning a vector to each of the top 10,000 most frequent words. Each word's vector representation is the average of the vectors of communities it is used in, weighted by the positive pointwise mutual information (PPMI) between the word and each community. In essence, this means each word's vector representation is a weighted average over Reddit community vectors, where the weights correspond to how much above expectation the word appears in each community.

[1]https://github.com/CSSLab/reddit-music

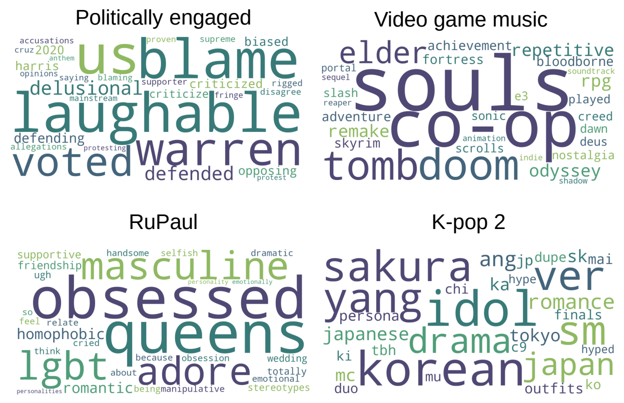

Figure 4: Word clouds of the most similar words to four social genres. Word-genre similarity is calculated by taking the cosine similarity of the word vector and the genre vector.

Comparing words to genres. Once we have embedded both words and artists into the same space, we can calculate a similarity score between words and social genres. We perform this calculation for the top 10,000 words and all 50 social genres and illustrate four of the results in Figure 4. Each word cloud features the 30 most associated words to the indicated social genre, which all have a cosine similarity of 0.7 or higher. Notice that the *politically engaged* word cloud includes language associated with politics, such as "voted", "biased", and "harris". Similarly, the *video game music* genre, which mainly includes shares of artists that have produced music used in video games, has words that are connected to video games such as "rpg", "sonic", and "skyrim". *RuPaul* includes words that are associated with drag, such as "queens", "adore", and "lgbt". Finally, the words most associated with *K-pop 2*, pop music from South Korea, are "idol", "drama", and "sakura". These examples demonstrate that this linguistic tool can enrich our understanding of social music sharing by surfacing hidden connections between music and language.

Social Contexts of Online Music Sharing

Community embeddings, like word embeddings, preserve meaningful relationships between communities in an internally consistent Euclidean space. This property implies that one can find axes in the space that represent meaningful concepts. Prior work developed a technique to extract robust, externally validated social dimensions in community embeddings such as age (older to younger) and gender (masculine to feminine) (Waller and Anderson 2020).

We apply this methodology to examine how patterns of music sharing relate to six salient aspects of identity and culture. These six dimensions in the community embedding, which are simply vectors in the space, correspond to age, gender, sociality, affluence, U.S. political partisanship, and partisan-ness (strength of politicization). Projecting the vector representations of communities, artists, and social genres onto these dimensions provides accurate measures of their social positions along these axes. For example, projecting the vector representation of the r/teenagers community onto the age dimension results in a score that indicates its user-base is much more aligned with the younger end of the spectrum than the older end. We apply these scores to identify genres and artists with notable social associations, as well as to understand the macroscale structure of musical sharing along cultural lines.

Social Dimensions

Background. We review the methodology developed by Waller and Anderson (2020) here. Each social dimension is created using two input "seed" communities, which are subtracted to obtain a vector for the axis. The seed communities are chosen such that they differ only along the target social dimension and are otherwise similar. The seeds used are: r/teenagers and r/RedditForGrownups (age), r/AskMen and r/AskWomen (gender), r/democrats and r/Conservative (partisan), r/vagabond and r/backpacking (affluence), r/nyc and r/nycmeetups (sociality). The partisan-ness axis, which represents how political a community is, is calculated by adding the vectors of the two seeds for the partisan axis (r/democrats and r/Conservative). Finally, seed augmentation and validation strategies are used to increase the robustness of the axes and validate their accuracy.

Communities, artists, and social genres are scored on these axes by projecting their vector representations onto the axis vector. This score reflects how similar it is to the seeds. If it is more similar to the first seed, it will have a negative score; if it is more similar to the second seed, it will have a positive score; and if it is equidistant between the two seeds, it will have a score of zero. This score thus represents the community's relative association with the two seeds and approximates the community's association with the desired social concept on Reddit. Note that these dimensions comprise a small subset of all possible social dimensions in the space, and therefore each of them is a simplification of the underlying cultural concept; we focus on these six dimensions for practical reasons and to illustrate a few of the broader cultural trends underlying music sharing on Reddit.

In what follows, we conduct both macro- and micro-level analyses of the social contexts of music sharing. On the macro-level, we examine how the contexts of music sharing as a whole differ from the contexts of typical Reddit content. On the micro-level, we extract insights from the projection of artists and social genres onto social axes.

Distinguishing characteristics of music sharing contexts. How do the social and cultural contexts of music sharing on Reddit differ from the contexts of typical Reddit activity more broadly? Presumably the contexts in which music is shared, and the communities for whom music sharing is disproportionately prominent, differ from Reddit at large—but how? To answer this question, in Figure 5 we measure how activity is distributed across the social dimensions separately for music (top), a time-matched sample of all Reddit comments (middle), and the difference between the two (bottom). The bottom row shows that contexts in which music is shared are systematically different from typical Reddit content. The area above the x-axis shaded in yellow is where music sharing is over-represented, and the area below the x-axis shaded in blue is where it is under-represented.

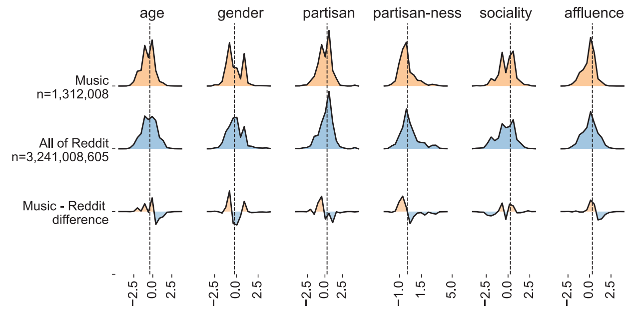

Figure 5: The distribution of activity along social axes for all instances of music sharing (top), a time-matched set of all Reddit comments (middle), and the difference between the two (bottom). Orange areas indicate where music sharing is over-represented relative to typical Reddit content.

Clear differences emerge. The sharing of music is much more likely to occur in less political contexts (low partisan-ness is over-represented). Similarly, it occurs disproportionately in younger, more left-wing, and more affluent contexts. Music sharing is also much more common towards both of the extremes on the gender axis. These results show that music sharing is organized along social and cultural lines.

Social contexts of genres. Taken together, Reddit music sharing on average thus has distinctive social signatures. Moving from the macro-level to the micro-level, what are the social and cultural contexts of how particular social genres and artists are shared? To answer this, we measure how every social genre and artist relate to the social axes by projecting their vector representations onto the dimensions discussed above. A high projection on affluence, for example, indicates an artist tends to be aligned with more "affluent" communities, and a low projection on the partisan-ness dimensions denotes they tend to be shared in less political communities.

First we examine the projection of social genres onto our social axes (see Figure 6). For each social genre on each dimension, we have shown the genre's z-score relative to all communities on that dimension (*e.g.* a score of 1.0 indicates a social genre is one standard deviation above the mean on that axis). The result is a fine-grained picture of how the contexts of sharing vary between social genres. On the age dimension, for example, rap, electronic, and K-pop are shared in much younger contexts than progressive metal, jazz, and DnB (drums and bass). There is a strong and significant Spearman correlation between the social genres' age scores and the median era of the music within them ($r = 0.55, p < 10^{-5}$)—newer music is shared in younger contexts. It is also clear that music sharing in general occurs

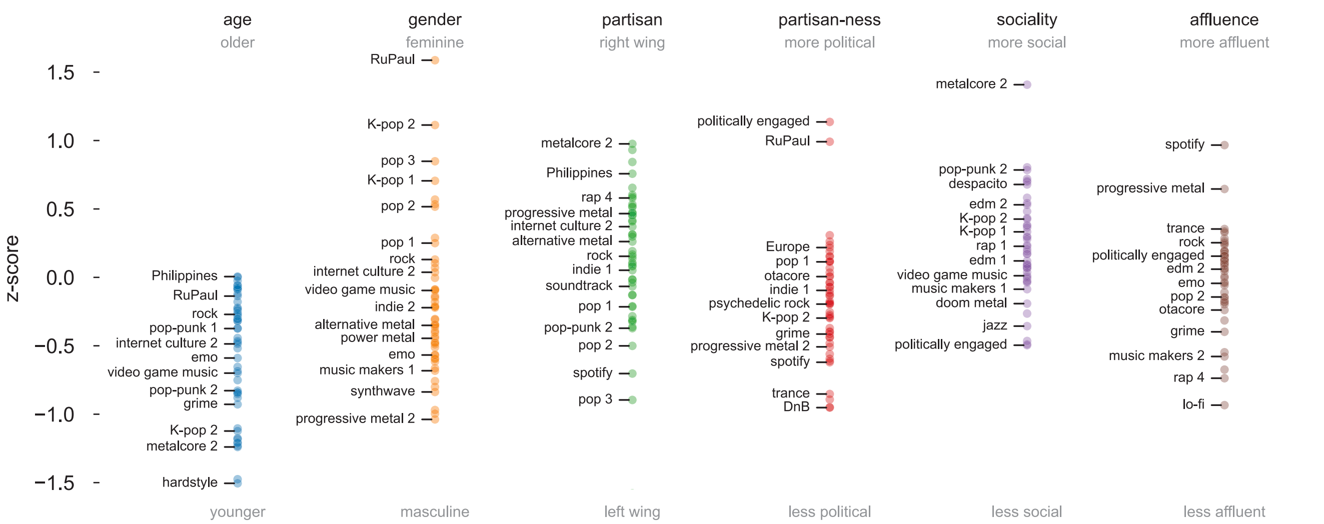

Figure 6: The social genre scores on each social dimension. Each genre score is expressed as its z-score relative to all communities on the respective dimension.

in younger contexts, since virtually every social genre has a negative score on the age axis. On the gender dimension, *RuPaul* is projected furthest in the feminine direction, followed by a preponderance of pop genres. In line with music on the right, we find metal and electronic music tend to lean more masculine. On the partisan axis, *RuPaul* is the farthest left. This aligns with the language most associated with *RuPaul*, featuring the words "lgbt", "queens", and others, which also lean left on Reddit. Other left-wing social genres include those encompassing pop, pop-punk, and indie music. The right-wing side of the spectrum is predominantly made up of metal and K-pop. Finally, the partisan-ness dimension confirms that the *politically engaged* social genre is shared in the most political contexts, followed closely by *RuPaul*.

Social contexts of artists. We also project artists onto the dimensions to examine the relationships between artists' scores on multiple axes. Figure 7 illustrates a strong correlation between affluence and age ($r = 0.70, p < 10^{-16}$). Artists that are shared in older contexts also tend to be shared in affluent contexts. Rappers such as Lil Pump, Travis Scott, and BROCKHAMPTON are over-represented in younger and less affluent contexts. Some alternative and pop artists are shared in mid-age and mid-affluence contexts, including Twenty One Pilots, Arctic Monkeys, and Ed Sheeran. Finally, Phish, Living Colour, and Frank Zappa are all older artists and are shared in older, more affluent contexts. This high-resolution social map of how individual artists are shared could be used to understand the cultural associations in which they are discussed. Beyond the strong relationship we've documented here, one can compare artists that are similar in one dimension but differ in the other. For example, the 80s punk rock band Dead Kennedys and contemporary rapper Meek Mill score similarly on the affluence dimension, but Dead Kennedys are shared in much older contexts than Meek Mill. Figure 8 shows artist scores

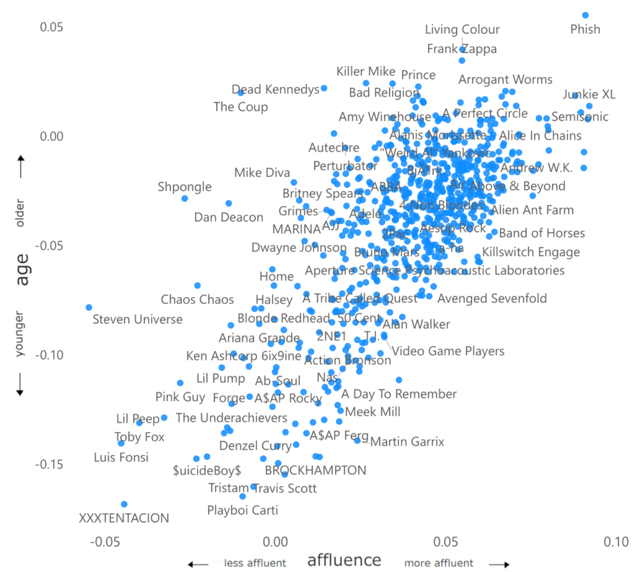

Figure 7: The joint distribution of artists along the age and affluence social axes.

on the gender and age axes, which also illuminates notable relationships. For example, the correlation between scores on these two axes is much weaker, as there are more artists with diverse combinations of age and gender scores ($r = 0.19$, $p < 10^{-6}$). Notably, however, among artists shared in younger contexts, there are far more that lean masculine. As before, controlled pairings of artists can be analyzed with this visualization. For example, A$AP Rocky and BTS draw similarly young audiences, but differ greatly in the gender associations of the contexts they are shared in.

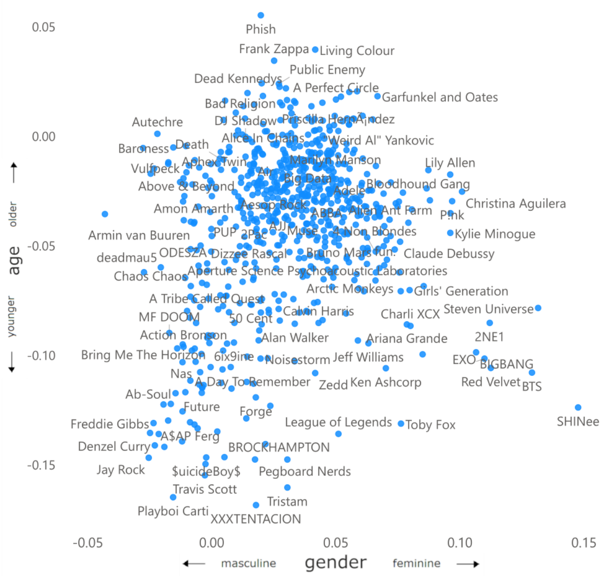

Figure 8: The joint distribution of artists along the age and gender social axes.

Quantifying Extra-musical Sharing

Frith argued that "While music may be shaped by the people who first make and use it, as experience it has a life of its own." (Frith 1996). We have seen this reflected in our analyses of social genres; how music is shared is not always purely driven by the music itself. Often, there are other meanings that become attached to certain pieces of music, and the patterns in who invokes them in public discussions differ greatly from the patterns in who listens to them in private. Following existing literature, we refer to this important behavior as *extra-musical sharing*.

But which music has developed "a life of its own" in the experience of how it is shared online? In this section, we develop two approaches to measuring extra-musical sharing. First we outline a general method based on the observation that artists who are shared in an unusually wide variety of contexts tend to have developed extra-musical meanings that drive these diverse sharing patterns. Then we propose a method to study a main driver of extra-musicality on Reddit: meme sharing. As cultural artifacts including songs are reproduced, transformed, and transmitted on the Internet, they sometimes acquire secondary meanings and are deployed in discussion as memes. We introduce an embedding-based approach to quantify meme sharing of music. ts see a large divide between listening and sharing habits, and can we create a new metric for artists that attempts to score this deviation.

Generalist-Specialist Scoring

To measure the extra-musicality of how an artist is shared, we first observe that extra-musical sharing is more likely than ordinary musical sharing to occur in a wide variety of contexts. When artists are straightforwardly shared for their music, as opposed to invoking them at another level of meaning, the sharing often happens in related communities. For example, fans of contemporary rap music share rap songs in rap-related communities. But when there are extra-musical factors that affect the sharing of an artist, the communities they are shared in can be quite different from each other, since the reason behind the sharing is decoupled from the music itself. For example, Rick Astley's "Never Gonna Give You Up" is shared in a diverse set of communities because a lot of *different* users understand the online practice of rick-rolling. Our approach in this section follows this insight: to measure the extra-musicality of how an artist is shared, we will measure the diversity of contexts they are shared in.

Measuring diversity. How can we measure diversity of sharing contexts? Although entropy is often used to quantify diversity, it does not capture similarities or differences between communities. For example, say we have two distinct artists a_1 and a_2, and both are shared in two communities. Artist a_1 is shared in r/classicalmusic and r/piano whereas a_2 is shared in r/classicalmusic and r/NFL. The entropy of sharing will be equal for these two artists, but intuitively we think of a_2 as appearing in a more diverse set of contexts than a_1. For this reason, we use the Generalist-Specialist score (GS-score), a simple, validated measure of diversity that incorporates similarities between communities (Waller and Anderson 2019). An artist's GS-score is defined as the average cosine similarity between the artist and all the communities the artist was shared in, weighted by the number of shares in each community:

$$GS(a_i) = \frac{1}{J}\sum_j w_j \frac{\vec{c_j}\cdot\vec{a_i}}{\|\vec{a_i}\|} \tag{2}$$

where $\vec{a_i}$ is the artist vector, w_j is the number of times a_i was shared in community j, and $J = \sum w_j$ is the total number of times a_i was shared. The higher the score, the more narrowly shared a_i is.

After calculating the GS-score for each artist, we restricted our attention to artists in the 70th percentile of sharing popularity or above (those who were shared at least 124 times). Since artists with more shares have a natural tendency to have a lower-GS score, we follow previous work and use an activity-adjusted GS-score to correct this bias (Anderson et al. 2020a). This activity-adjusted GS-score is calculated as the percentile GS-score within each activity level (for the 80th, 90th, and 100th popularity percentile groups). An artist with an activity-adjusted GS-score of 100 has the narrowest sharing patterns among artists of their popularity level. Finally, we normalize these scores to lie in $[0, 1]$ and reverse the order so that 0 corresponds to narrowest sharing contexts and 1 corresponds to the widest.

Results. This methodology results in a coherent picture of musical versus extra-musical sharing (see Table 2). The most widely-shared artist given their popularity is The Pointer Sisters, whose song "I'm So Excited" is frequently brought up to express excitement. Similarly, Kool & The Gang wrote the song "Celebration", which is often invoked in online discussion in a celebratory way. Some of the most extra-musical artists wrote popular songs used in films, including Dwayne Johnson's "You're Welcome", Kenny Loggins' song from Top Gun, and "One Day More" from Les

	Artist	Top Track	GS
Top ten	The Pointer Sisters	I'm So Excited	1.0
	Kool & The Gang	Celebration	1.0
	Les Misérables	One Day More!	1.0
	Bonnie Tyler	Total Eclipse Of The Heart	0.998
	Dwayne Johnson	You're Welcome	0.998
	Tina Turner	Goldeneye	0.998
	Simon & Garfunkel	The Sounds Of Silence	0.997
	Lionel Richie	All Night Long	0.995
	Noisestorm	Crab Rave	0.995
	The HU	Wolf Totem	0.995
Bottom ten	MC Rich	Down the Coast	0.005
	Lovelyz	Destiny	0.003
	Immortal Technique	Dance With the Devil	0.003
	WINNER	SENTIMENTAL	0.003
	Phinehas	White Livered	0.002
	Jay Rock	Vice City	0.002
	Wvs	Cotton Candy	0.002
	dom champ	I Feel Better	0.0
	Invent, Animate	Darkbloom	0.0
	SHINee	View	0.0

Table 2: Top and bottom artists ranked by our extra-musical sharing score.

	Artist	Top Track	Meme
Top ten	Luis Fonsi	Despacito	1.0
	Video Game Players	DuckTales	0.713
	Blonde Redhead	For the Damaged Coda	0.703
	Valve Studio Orchestra	Team Fortress 2	0.68
	Rick Astley	Never Gonna Give You Up	0.639
	Starbomb	It's Dangerous to Go Alone	0.616
	Pink Guy	Stfu	0.612
	Kirin J Callinan	Big Enough	0.59
	Noisestorm	Crab Rave	0.587
	Chaos Chaos	Do You Feel It?	0.571
Bottom ten	Björk	It's Oh So Quiet	0.05
	St. Vincent	Digital Witness	0.046
	Janelle Monáe	Make Me Feel	0.044
	Baroness	Chlorine & Wine	0.042
	Purity Ring	bodyache	0.038
	Slowdive	Alison	0.037
	Jon Hopkins	Open Eye Signal	0.029
	Deafheaven	Dream House	0.019
	Autechre	feed1	0.001
	Pavement	Cut Your Hair	0.0

Table 3: Top and bottom artists ranked by our meme score.

Misérables. Although Simon & Garfunkel are often shared musically, "The Sounds of Silence" has become an Internet meme due to the lyric "Hello darkness my old friend", hence their high extra-musicality score.

On the other end are artists who are shared in very specialized contexts. These artists are dominated by rap and international music, since discussion of these genres is often localized in a small set of communities. For example, the most specialized artist is the K-pop group SHINee, whose sharing almost entirely occurs in K-pop communities, which occupy a small niche in the community embedding. The other artists tend to be rappers whose sharing is mostly concentrated in hiphop-related communities. This aligns with the rap social genres we found above.

Meme Scoring

The last ten years saw the rise of online memes as a central component of the Internet's participatory culture. "Meme culture" extends to music; as we have seen above, some artists and songs have become associated with memes. This occurs frequently and explains a large amount of music sharing on Reddit, and so here we aim to quantify the "meme score" of an artist—how meme-driven their sharing is. The prototypical example of this is Rick Astley's "Never Gonna Give You Up", in which people "rickroll" others into clicking on an interesting-seeming link that actually directs to Astley's music video. As a result, the vast majority of Rick Astley sharing is driven by the meme association this song has developed. Here we present a novel method of quantifying how meme-oriented an artist's sharing is.

Quantifying meme sharing. Our approach is to construct a "meme dimension" in the embedding space that corresponds to meme sharing of music, similar to the social dimensions discussed above. In order to measure this, we first need to find communities that are extremely associated or disassociated with the meme sharing of music. We analyzed the sharing of "Never Gonna Give You Up" and calculated which communities disproportionately shared this song. Specifically, we calculated the z-score between the song and all subreddits with at least 100 music shares:

$$z_{a,s} = \frac{p_{a,s} - \mu_a}{\sigma_a} \tag{3}$$

where $z_{a,s}$ represents the z-score between artist a and subreddit s, $p_{a,s}$ is the proportion of music shares that artist a represents in subreddit s, and μ_a and σ_a are the mean and standard deviation of the proportion of music shares a represents on average. For example, on average we expect "Never Gonna Give You Up" to represent 9.45% of music shares on a subreddit, with a standard deviation of 0.104. In the r/futurebeats community, however, this song makes up only 0.038% of its shares, resulting in a low z-score of -0.905. But in the r/dankmemes community, Astley's song is responsible for a staggering 79.6% of the music shares, resulting in a high z-score of 6.73.

We used these z-scores to define the sets of communities that are extremely associated or disassociated with meme sharing of music, from which we can construct a meme vector. We created a pro-meme vector, denoted v_m^+, by averaging all communities with a z-score greater than 1.7, and created an anti-meme vector, denoted v_m^-, by averaging all communities with a z-score less than -0.3. We constructed a meme dimension by normalizing these two vectors and subtracting them ($v_m = \hat{v}_m^+ - \hat{v}_m^-$), which we finally normalized to produce $\hat{v_m}$. By projecting all artists onto this meme dimension $\hat{v_m}$ and normalizing, we can calculate their meme-sharing scores.

Results. Table 3 shows the top and bottom ten artists by normalized meme score, restricted to artists above the 90th percentile of popularity for ease of interpretability. All top ten artists are associated with well-established memes, and we discuss the top five in detail here. The top artist is Luis Fonsi, whose hit song "Despacito" became a meme on Reddit, in which users reply to sad news by saying "So sad, Alexa play despacito", triggering a Reddit bot to automatically respond with a link to the song. The vast majority of this sharing took place in r/Darkjokes and r/memes. The song "DuckTales" by Video Game Players has also become a meme online, especially within the video game playing community.

The context for the shares ranges widely, but in general this artist is shared in gaming communities including r/gaming, r/GamePhysics, r/Overwatch, and r/Warframe. Blonde Redhead's song "For the Damaged Code" was sampled in a song that appeared in a popular comedy TV show, and has become frequently used in videos posted in the community r/WatchPeopleDieInside. Similar to DuckTales, Valve Studio Orchestra's "Team Fortress 2" was the theme song for the popular video game Team Fortress 2. Finally, Rick Astley has already been discussed.

The artists with low meme scores tend to be award-winning and critically-acclaimed musicians. Pavement and Slowdive produce indie and experimental music, St. Vincent and Björk are singer-songwriters, and Jon Hopkins, Autechre and Purity Ring are electronic bands.

Discussion

In this work, we presented a set of methodologies for understanding the macro-scale structure of online music sharing and its relation to social and cultural dimensions, and we applied our methodologies to conduct an analysis of music sharing on Reddit. The research presented aims to fill a gap between traditional qualitative approaches, which are limited in the number of analyses they can conduct, and computational approaches, which often lack crucial social and cultural context. Our computational approach to the study of music sharing complements these existing methods.

We first contribute a novel dataset of millions of instances of music sharing on Reddit. This dataset is of interest for several reasons. First, music sharing on Reddit is an explicitly social activity, as opposed to consuming music on platforms such as Spotify or YouTube, which is often a solitary activity. In contrast with audio signals and musical data, our inherently social, large-scale dataset opens the door for a computational complement to sociological and ethnographic approaches to studying music sharing. Second, since Reddit is a diverse platform with thousands of communities, we can connect instances of music sharing with the broader social and cultural contexts that exist on Reddit. Third, Reddit is virtually entirely devoid of algorithmic curation. As such, the patterns we observe are more likely the result of genuine user choices and are not shaped by algorithmic forces. Finally, our focus on music sharing enables us to heed DeNora's call to study music via understanding how their listeners interact.

We clustered artists into socially constructed groups with similar sharing patterns. These *social genres* tend to fall into two categories—artists that are shared alike due to a similarity in their music forms (what Meyer calls iconic sign qualities), and artists that are connected by some external social aspect (symbolic sign qualities). The importance of these extra-musical features echoes findings in the sociological literature (Kristen and Shevy 2013; DeNora 1986). Our computational approach allows us to cluster all 6,600 artists into 50 coherent social genres, which could provide a basis for further sociological investigation.

We applied community embeddings to understand the social contexts of music. The social dimension scores we obtained for each musician provide both a high-resolution understanding of the social contexts each artist is embedded in, as well as a big-picture view of how social differences are expressed via music on Reddit. Using these scores, future researchers can quickly access a collection of, for example, musical artists that are associated with the right-wing, or high affluence. Moreover, this technique provides a computational approach of quantifying social differences in taste, much like Bourdieu's original work. Our research can also help the music curation industry by proposing a new set of social features to understand music.

Finally, we developed methods to measure two forms of extra-musicality—how broadly a musician is shared and the degree to which a musician is associated with meme culture. The first metric acts as a quantification of the idea that some music "reaches across boundaries and bridges social relationships" (Roy 2013). We directly measure the breadth of contexts that a particular artist is shared in. Additionally, our meme scoring method can enable new analyses on cultural transmission (Cavalli-Sforza and Feldman 1981). It can be used to study how cultural objects become associated with memes over time, as well as differentiate sharing that is meme-driven compared to more conventional music sharing. Furthermore, our quantitative scores of extra-musicality could be used to distinguish musical-form-driven from culture-driven preferences. Modern recommendation systems are blind to the underlying social and cultural contexts of the large-scale data they harness to algorithmically recommend music to users. By conflating extra-musical associations with musical ones, their recommendations could be socially inept or even culturally harmful. Algorithm designers could build on our methods to construct algorithmic systems that are more attuned to the social and cultural contexts that music is embedded in.

Our research is limited by several factors. While our focus on Reddit allows us to conduct the first large-scale computational study of social music sharing, it is important to acknowledge that Reddit comprises a specific slice of online activity that is not representative of the broader music sharing ecosystem. In particular, Reddit has a young and Anglocentric user base. We also restricted our attention to links that are verifiably connected with a specific artist. However, this ignores less structured invocations of music, such as free-text mentions.

Our work introduces a computational methodology to study music by understanding how it is shared online. As such, we aim to derive insight not from the musical form a piece takes, but from the "life of its own" it experiences out in the world. We hope that others build on our work to deepen our understanding of both music and the online communities that share it. Sociologists and musicologists could use our methods to generate hypotheses that warrant further investigation. Also, while it wasn't our main focus, there is much to be learned about online groups through the music they share. Finally, our approach of studying music by understanding its contexts is not limited to music. We believe the study of many types of online content and cultural forms would benefit from our computational, context-driven framework.

Acknowledgments

We thank Georgina Born for helpful discussions, and NSERC and CFI for their support.

References

Anderson, A.; Maystre, L.; Anderson, I.; Mehrotra, R.; and Lalmas, M. 2020a. Algorithmic effects on the diversity of consumption on spotify. In *The Web Conference 2020*.

Anderson, I.; Gil, S.; Gibson, C.; Wolf, S.; Shapiro, W.; Semerci, O.; and Greenberg, D. M. 2020b. "Just the way you are": Linking music listening on spotify and personality. *Social Psychological and Personality Science*.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The pushshift Reddit Dataset. In *International AAAI Conference on Web and Social Media*.

Bennett, T.; Savage, M.; Silva, E. B.; Warde, A.; Gayo-Cal, M.; and Wright, D. 2009. *Culture, class, distinction*. Routledge.

Blacking, J., and Nettl, B. 1995. *Music, culture, and experience: Selected papers of John Blacking*. University of Chicago Press.

Bourdieu, P. 1984. *Distinction: A social critique of the judgement of taste*. Harvard University Press.

Carrignon, S.; Bentley, R. A.; and Ruck, D. 2019. Modelling rapid online cultural transmission. *Nature Humanities and Social Sciences Communications*.

Cavalli-Sforza, L. L., and Feldman, M. W. 1981. *Cultural transmission and evolution: A quantitative approach*. Princeton University Press.

Chan, T. W., and Goldthorpe, J. H. 2007. Social stratification and cultural consumption: Music in england. *European Sociological Review*.

Colley, L. 2005. *Britons: forging the nation, 1707–1837*. Yale University Press.

Condry, I. 2004. Cultures of music piracy: An ethnographic comparison of the us and japan. *International Journal of Cultural Studies*.

DeNora, T. 1986. How is extra-musical meaning possible? music as a place and space for "work". *Sociological theory*.

DeNora, T. 2000. *Music in everyday life*. Cambridge University Press.

Frith, S. 1996. Music and identity. *Questions of Cultural Identity*.

Gregory, A. H. 1997. *The roles of music in society: The ethnomusicological perspective.* Oxford University Press.

Hamessley, L. 2006. Women's voices across musical worlds. *Women and Music: A Journal of Gender and Culture*.

He, S.; Zheng, X.; Bao, X.; Ma, H.; Zeng, D.; Xu, B.; Tian, G.; and Hao, H. 2014. Ranking online memes in emergency events based on transfer entropy. In *2014 IEEE Joint Intelligence and Security Informatics Conference*.

Koskoff, E. 2014. *A feminist ethnomusicology: writings on music and gender*. University of Illinois Press.

Kristen, S., and Shevy, M. 2013. A comparison of german and american listeners' extra musical associations with popular music genres. *Psychology of Music*.

Magrini, T. 2003. *Music and gender: Perspectives from the Mediterranean*. University of Chicago Press.

Mehr, S. A.; Singh, M.; Knox, D.; Ketter, D. M.; Pickens-Jones, D.; Atwood, S.; Lucas, C.; Jacoby, N.; Egner, A. A.; Hopkins, E. J.; et al. 2019. Universality and diversity in human song. *Science*.

Meyer, L. B. 2008. *Emotion and meaning in music*. University of Chicago Press.

Prior, N. 2013. Bourdieu and the sociology of music consumption: A critical assessment of recent developments. *Sociology Compass*.

Rochberg, G. 1968. Contemporary music in an affluent society. *Music Educators Journal*.

Roy, W. G. 2013. *Reds, whites, and blues: Social movements, folk music, and race in the United States*. Princeton University Press.

Seeger, C. 1957. Music and class structure in the united states. *American Quarterly*.

Stokes, M. 1997. *Ethnicity, identity and music*. Berg Oxford.

Street, J. 2003. 'fight the power': The politics of music and the music of politics. *Government and Opposition*.

Susino, M., and Schubert, E. 2020. Musical emotions in the absence of music: A cross-cultural investigation of emotion communication in music by extra-musical cues. *PloS one*.

Tarrant, M.; North, A. C.; and Hargreaves, D. J. 2002. Youth identity and music. *Musical identities*.

Vianna, H. 1999. *The mystery of samba: popular music & national identity in Brazil*. Univ. of North Carolina Press.

Vuoskoski, J. K., and Eerola, T. 2015. Extramusical information contributes to emotions induced by music. *Psychology of Music*.

Waller, I., and Anderson, A. 2019. Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. In *The World Wide Web Conference*.

Waller, I., and Anderson, A. 2020. Community embeddings reveal large-scale cultural organization of online platforms.

Way, S. F.; Gil, S.; Anderson, I.; and Clauset, A. 2019. Environmental changes and the dynamics of musical identity. In *International AAAI Conference on Web and Social Media*.