Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)

Gender Differences in the Global Music Industry: Evidence from MusicBrainz and The Echo Nest

Yixue Wang
Northwestern University
2240 Campus Drive
Evanston, Illinois 60208
yixue.wang@u.northwestern.edu

Emőke-Ágnes Horvát
Northwestern University
2240 Campus Drive
Evanston, Illinois 60208
a-horvat@northwestern.edu

Abstract

With digital music consumption being at an all-time high, online music encyclopedia like `MusicBrainz` and music intelligence platforms like `The Echo Nest` are becoming increasingly important in identifying, organizing, and recommending music for listeners around the globe. As a by-product, such sites collect comprehensive information about a vast amount of artists, their recorded songs, institutional support, and the collaborations between them. Using a unique mash-up of crowdsourced, curated, and algorithmically augmented data, this paper unpacks an unsolved problem that is key to promoting artistic innovation, i.e., how gender penetrates into artistic context leading to the globally perceived gender gap in the music industry. Specifically, we investigate gender-related differences in the sonic features of artists' work, artists' tagging by listeners, their record label affiliations, and collaboration networks. We find statistically significant disparities along all these dimensions. Moreover, the differences allow models to reliably identify the gender of songs' creators and help elucidate the role of cultural and structural factors in sustaining inequality. Our findings contribute to a better understanding of gender differences in music production and inspire strategies that could improve the recognition of female artists and advance gender equity in artistic leadership and innovation.

Introduction

Recently, there has been an increasing level of awareness about the differential treatment of artists based on their gender in a variety of creative sectors (Smith, Choueiti, and Pieper 2017; 2018). The music industry is at the forefront of several mediatized debates related to systemic gender biases (Newman 2018), yet scholarly works that investigate the issue at scale are largely missing. Primarily due to difficulties in extracting and quantifying musical features in a non-automated way, previous studies focused on general compositional structures. For instance, recent work found no differences between male and female artists in terms of their high-level compositional quality (Sergeant and Himonides 2016). Yet, gender inequalities in the music context seem to be the status quo, mirroring disparities seen in other creative fields and beyond. While these gaps are well-documented in areas ranging from the workspace and wages (Blau and Kahn 2000; Kuhn and Villeval 2013) to educational opportunities (Hausmann et al. 2009), entrepreneurship and capital markets (Brooks et al. 2014; Kanze et al. 2017; Horvát and Papamarkou 2017), as well as leadership (Ragins 1998; Burke and Collins 2001), it is largely underexplored in the case of the global music scene.

Data about world-wide music production is becoming increasingly available in conjunction with the heavy use of music streaming services like Spotify, Tidal, Apple Music, Amazon and Google Play. These data enable a comprehensive study of music, where the complementary input of men and women is not only desired, but essential. By building on the interdependent voices, sensibilities, and associated artistic worlds of the two groups, music is considered a highly gendered form of expression (Treitler 2011), i.e., "fraught with gender-related anxieties" and "strongly informed by erotic imagery" (McClary 1991). Gender effects and musical expression are thus exceptionally hard to untangle, making the problem more severe here than in other creative fields or mundane settings that have been studied before (Lutter 2015; Altenburger et al. 2017; Wachs et al. 2017; Altenburger and Ugander 2018). Firmly ingrained human biases are at the root of emerging gender inequalities, stereotyping, and discrimination in the music industry (Boimabeau 2009). Recent research indicates that statistical tools and appropriate algorithmic implementations can have the potential to detect, raise awareness of, and eventually help overcome some of the harmful individual biases (Kleinberg et al. 2017; 2018). Approaching the gender problem in music from the algorithmic de-biasing perspective would not only represent progress in a difficult area, but it would do so in a context that is recognized to be an important pioneer and trendsetter for the gig economy (Baym 2018). Results obtained in the case of the music industry can thus be expected to translate to novel directions in today's changing labor markets.

In this paper, we thus tackle empirically the open question about the existence of a specifically female style in music. Our analysis has an unprecedented scale and operates along a varied set of dimensions. Specifically, we ask: *Is male* *and female sound distinguishable? Do listeners use different categories to describe male and female artists' songs? Is this work produced under the flagship of different institutions and via different collaboration patterns?* The answers to these questions led to a modelling endeavor, which shows that gender differences are pronounced enough to enable the reliable identification of the gender of a song's creator. These investigations were based on a global dataset that collects and collates crowdsourced, edited, curated, as well as algorithmically filtered and augmented information on various aspects of gender differences in music. In the U.S. alone, 251.9 billion people have used online streaming platforms like Spotify and YouTube as their primary source of music consumption in 2016 (Crawford 2016). Many of them regularly contribute to online music encyclopedia, discuss preferences and discoveries on social media, and feed data into recommendation engines. The identification of the most relevant sources, the matching and cleaning of their content has resulted in the comprehensive dataset that we used here.

Our systematic assessment of gender disparities in music production contributes a framework that evaluates differences in the created musical content alongside dissimilarities of cultural and structural origins. Specifically, the framework compares data summarizing songs' underlying sonic features, tagging by listeners, institutional support, and collaboration networks around artists of the two genders. It leads to models that allow comparing the effect of the considered factors on inequality. Eventually, a better understanding of gender differences points to a "female way" of producing music that could inform attempts at better nurturing women on the global music scene.

Data

The dataset that we use describes the sonic features of 232,798 songs combined with detailed artist metadata for 8,247 solo artists. This sample represents commercially recorded popular music between 1960 and 2000 signed by solo artists with accurate gender assignment. The first component of the dataset compiles song-level information from `The Echo Nest`[1], a music intelligence platform now owned by Spotify. The key innovative part of this data is the fine-grained *sonic feature* set that the platform deduced using state-of-the-art Music Information Retrieval (MIR) (Bertin-Mahieux et al. 2011). MIR extracts numeric information from digital audio files, generating quantifiable descriptions of songs' underlying musical attributes (Friberg et al. 2014). The key idea is that individuals' musical preferences are linked to a series of features that structure the musical space, such as speed, repetition, and loudness (Greenberg et al. 2016). The sounds perceived by listeners are mapped via MIR to sonic features that reliably distill the complexity of music and make statistical comparisons feasible. Our dataset comprises the following important MIR-based sonic features: standard musical attributes *tempo*, *mode*, *key*, and *time signature*; the simple quantification of the *duration* of songs; as well as a series of measures that represent particular aural or emotive dimensions of music: *valence*, *loudness*, *danceability*, *acousticness*, *energy*, *liveness*, and *speechiness*. The sonic features are thus either binary (e.g., major/minor mode), discrete (e.g., distinct values of estimated overall song key from C through B), or continuous (e.g., tempo quantified as an average number of beats per minute). Table 1 summarizes the sonic features. A slightly altered set of sonic features from `The Echo Nest` have been used recently to study optimal differentiation in music (Askin and Mauskapf 2017).

Besides song and album listings, `The Echo Nest` contains detailed artist-level information as well. Most importantly, a large set of crowdsourced *tags*, cross-validated between several listeners, combined with social media mentions, and finally thresholded based on the reliability of association. To limit our dataset to popular music, we drop artists who have been tagged with versions of the labels classical, opera and soundtrack. The 571 remaining distinct attributions position artists on a fine-grained tag landscape that ranges from jangle pop and bluegrass to singer songwriter and Greek music. The mapping between artists and tags is not one-to-one. Instead, artists typically have 2.08 ($\pm$1.36) associations.

Additionally, we collected from `MusicBrainz`[2] information about artists' affiliation with record labels and collaborations within the industry. First, the extra information we obtained from this leading crowdsourced platform for music metadata enables us to track the institutional affiliation of artists. There are 6,077 different *record labels* that are spread around the world. With semi-manual cleaning we eliminate typos and account for different languages and local subsidiaries of the same franchise. This way we collate 4,873 verified record labels with 11 being the median number of songs produced by a label. Second, we build a *collaboration network* that describes the shared recordings of solo artists. We do not take into account band memberships, as in those cases the dynamics of production is considerably different and a clear attribution of creative input is extremely difficult. Note that when doing the matching between the two data platforms, on the one hand, we could not find collaborators for every artist with available sonic and tag information. While part of this could be due to missing data, a great portion of artists never actually collaborate throughout their careers. On the other hand, we also found several artists who never released a song on their own, but only collaborated on others' projects, which is also typical for the industry. We included these collaborators in the network, which finally consisted of 8,757 artists and 16,577 connections between them. Although the network has 358 components, the giant component contains 89.7% of the artists indicating that it might have the core-periphery structure observed in other artistic and scientific collaboration networks (Csermely et al. 2013; Yang and Leskovec 2014).

Unlike most large-scale studies of gender differences, which often infer gender assignment through an analysis of individuals' names and/or profile photos (Karimi et al. 2016; Wachs et al. 2017), here we used binary gender codes that were explicitly attributed by listeners and checked by

[1]http://the.echonest.com/

[2]http://musicbrainz.org/

Sonic feature	Description	Mean ♂	Mean ♀	p
Tempo	Average tempo in beats per minute	117.07	116.33	< 0.001
Mode	Minor or major key	0.70	0.69	0.41
Key	Estimated overall key from C through B	3.95	3.93	< 0.001
Time signature	Estimated overall time signature in beats per bar	3.85	3.84	< 0.001
Duration	Length in seconds	245.93	229.26	< 0.001
Valence	Musical positiveness	0.54	0.50	< 0.001
Loudness	Perception of sound pressure determined by the mastering technology	-12.34	-11.68	< 0.001
Danceability	How suitable a song is for dancing based on tempo, regularity of beat, and beat strength	0.56	0.55	< 0.001
Acousticness	Likelihood of recording solely by acoustic means (as opposed to electronic means)	0.49	0.53	< 0.001
Energy	Perception of intensity based on speed, loudness, and noisiness	0.49	0.45	< 0.001
Liveness	Likelihood of recording in the presence of a live audience (as opposed to studio production)	0.22	0.20	< 0.001
Speechiness	Likelihood of existing spoken words (vocals are not considered)	0.09	0.06	< 0.001

Table 1: Summary of the used sonic features. Shown are averages for male and female artists' songs as well as results of two-tailed Kolmogorov-Smirnov tests that compare the two distributions.

`MusicBrainz` editors. We manually confirmed the correctness of gender assignment for a random selection of 100 artists. The compiled data comprises 6,164 male and 2,083 female solo artists who released 177,856 and 54,942 songs, respectively. Figure 1 compares the prevalence of female versus male solo artists over the forty years covered by our dataset. In agreement with research on creative occupations (Grow and Deng 2014; Koppman 2014), we find that men are consistently overrepresented in artistic fields. Although the percentage of females increased from roughly 20% to 25% between 1960 and 2000, throughout the entire time period, men have released more songs than women, even after controlling for the imbalance in representation.

Data coverage and limitations

Our dataset has three important limitations. First, despite having a technically sound gender-ascription methodology, we acknowledge the major simplification we make by coding gender as a dichotomous variable. This coding is aligned with the conventional binary gender model that combines physical and biological markers with cultural narratives of purpose and social norms of interaction into the groups of males and females (West and Zimmerman 1987; Lamont and Molnár 2002; Keener 2015). Contemporary thinking challenges the wide-spread binary gender concept and advocates for more fluid categories (Richards et al. 2016). Data on these categories is slowly becoming available, but as of now it does not exist at the scale and for the population studied here.

Second, much like other large-scale datasets about cultural production (Hochman and Manovich 2013; Spitz and Horvát 2014), our sample is skewed toward Western popular music. Figure 2 illustrates nonetheless a broad global coverage and enables a cross-country comparison of the percentage of female solo artists. Among countries represented in our dataset by at least 10 solo artists, Egypt has 12 male and no single female musician, followed by Algeria and Puerto Rico with 7.14% and 8.33% female artists. Latvia, Israel, Portugal and Philippines have the highest female representation (from 60% to 50%).

For artists to be included in our data, they need to have been affiliated with at least one established record label in their region. Musicians who have never recorded with a label are missing from our analysis. Our sonic feature and artist-level data represents a good coverage of all recorded music available online and there is no reason to believe that there is a systemic gender bias in this sampling. However, a third limitation of our data is that it is hard to assess the comprehensiveness of its crowdsourced components. The gender labeling, listener tags, record labels as well as collaborations could be biased. The nature and extent of this bias has not been explored yet. Literature documents lower female contribution rates on peer production platforms, online repositores and freelance communities (Robles et al. 2014; Vasilescu et al. 2015; Hargittai and Shaw 2015; Wachs et al. 2017). This lesser contribution is reflected in an inferior coverage of content about women in these examples. While we cannot exclude the possibility of a similar gender bias in our crowdsourced data about music production, comparisons with small-scale data available for specific local music industries reflects reassuring agreements. For instance, a recent report based on the top 600 songs from 2012 to 2017, curated from Billboard charts, found that 22.4% of them were women (Smith, Choueiti, and Pieper 2018), in agreement with our finding about the overall participation rate of women. Thus, although we certainly err on the side of exact numbers, we expect the trends to be representative.

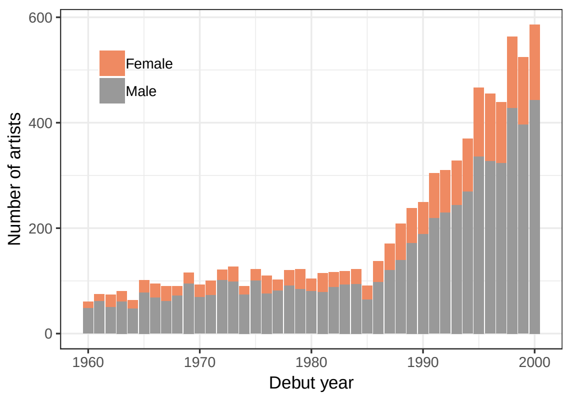

Figure 1: The distribution of years in which solo artists debuted shows the overrepresentation of men in the music industry.

Methods

To systematically explore gender differences in music, we propose a framework that statistically compares male and female artists along a set of dimensions that have been discussed in previous literature in areas ranging from musicology to psychology and management.

Sound-level analysis The underrepresentation of women in the arts has raised the question whether creative output was gender-specific or not (Citron 1993). One line of work has found no characteristic male-or-female-sound (McClary 1991), albeit it was mainly based on a couple of high-level musical properties (Sergeant and Himonides 2016). However, research has also hypothesized that a broader range of elements which together constituted music needed to be considered to make an accurate determination (Halstead 1991). In line with this proposition, we use developments in MIR to perform a large-scale comparison between men and women at the elemental level of several sonic features (see Table 1). These features are closest to the essence of music as a form of artistic expression and, arguably, they are responsible for the most fundamental connection between listeners' preferences and individual songs. Thus, we first perform a *sound-level analysis* that centers on the question whether there are any differences in the sonic features of songs released by male and female artists.

Tag-level analysis As opposed to debating about the existence of a specifically female musical style, researchers largely agree on the role of listeners in ascribing masculinity and femininity to musical artists' work (Bem 1987; Sergeant and Himonides 2014). Even experiments on expert listeners that have shown no competence in identifying the gender of classical music composers, indicated strong gender stereotyping: subjects of both genders made significantly more male-composer attributions than female (Sergeant and Himonides 2016). This suggests that a considerable part of the gendered impressions experienced by listeners are imposed subjectively as a result of their own, previously established gender conceptions. We thus perform a *tag-level analysis* that asks whether listeners tend to attribute different tags to male and female artists. To find tags that are statistically overrepresentative of men or women, we compute log-odds-ratios with an informative Dirichlet prior, which control for the variance in tag frequency based on tag usage across both genders (Monroe, Colaresi, and Quinn 2008).

Distribution-level analysis Gender stereotypes are not only descriptive of typical male or female behaviors, but they often become prescriptive (O'Neill and Boultona 1996). Given how pervasive subjective prescription is among general audiences and experts alike, it can be expected that institutions involved in music distribution actively promote socially agreed gendered behavioral styles, e.g., through the choice of artists they sign. This leads inevitably to a record label culture that disadvantages most non-conforming expressions (Negus 1999). To address this problem, we include a *distribution-level analysis* that is concerned with differences in the institutional context men and women are embedded in as measured through their affiliation with record labels. The importance of record labels cannot be understated: established institutions play a key role in marketing songs and are often the main facilitators of the wide distribution of music. Lacking access to major record labels in an artist's area is thus likely to prevent them from reaching their target audiences. Similarly as with tags, we assess the significance of associations between record labels and the two genders using the log-odds-ratio with an informative Dirichlet prior.

Network-level analysis Aside of institutional affiliations, informal social structures around artists, i.e., the composition of their collaboration networks, might affect men and women differently and could highlight some of the structural barriers that female artists face throughout the advancement of their careers. Social capital has been shown to impact career success in various occupations (Seibert, Kraimer, and Liden 2001). Research about creative success on Broadway and in jazz further emphasized this link (Uzzi and Spiro 2005; Vedres 2017), and a recent study identified collaboration patterns that penalize film actresses (Lutter 2015). Our *network-level analysis* asks whether male and female artists have collaboration networks of different structure. The idea is simple: aside of the fact that well-positioned collaborators might link lesser-known musicians with the appropriate institutions, collaboration networks also have a key role in assuring increased exposure to others' ideas and resources, and thus they might result in more creative recombination (Madlock-Brown and Eichmann 2016). The Smurfette principle states that women might be less likely to take advantage of social capital as they tend to be positioned in the periphery of collaboration networks with a core composed of men (Pollitt 1991; Wagner et al. 2015). To test whether this is the case in global music production, we define collaboration as the co-recording of songs and examine the structure of such collab-

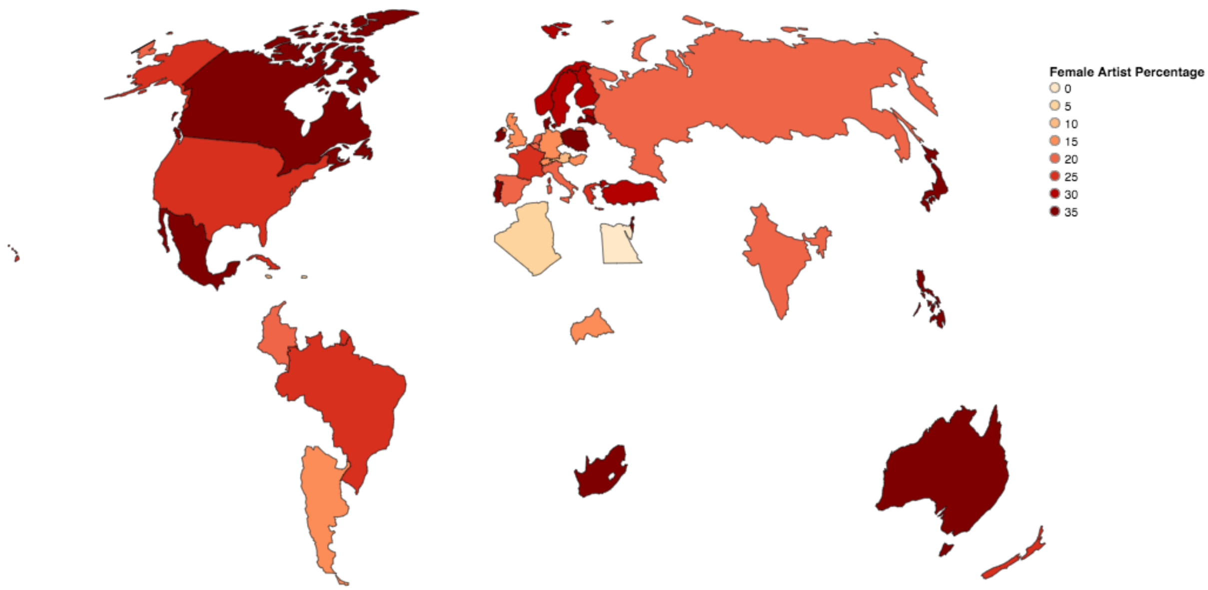

Figure 2: The map shows the percentage of female solo artists in our sample for countries with at least 10 individual artists.

orations using various notions of network centrality (Newman 2010):

a.) Who has more collaborators or, in network terms, who has a higher degree?

b.) Whose collaborators are more highly connected themselves? We measure this using the eigenvector centrality, which is a variant of Google's PageRank and weights collaborators according to their importance.

c.) How "close" is an artist to everyone else in the collaboration network though chains of collaborators? We quantify this using closeness centrality, i.e., the mean shortest-path distance from an artist to all other artists in the collaboration network.

d.) Is the artist positioned in the core or at the periphery of the network? Coreness is computed through a repeated pruning based on degree such that in the k-core of the network, every artist has at least k collaborators.

After using these modalities to quantify structurally important positions in the network, we investigate the propensity to form collaborations within one's own gender and across genders. The well-known social mechanism of homophily dictates that similarity based on demographics like race, age, or gender breeds connection (McPherson, Smith-Lovin, and Cook 2001). We compute a homophily index that captures the number of female collaborators relative to the total number of collaborators (Currarini, Jackson, and Pin 2009). In the case of music, however, we expect that due to the complementarity of male and female voices and sensibilities, solo artists would often record with the opposite gender. Recent research has shown that even in networks with weak gender-homophily, there could be a gender preference at two hops from the focal person. This tendency is called monophily and it entails a systematic preference for women *or* men (i.e., one of the groups). Regardless of the focal person's gender, monophily affects the choice of friends-of-friends, instead of friends like homophily does (Altenburger and Ugander 2018). Hence, as Altenburger and Ugander phrase it, it has implications for "the company you are kept in" instead of "the company you keep". The presence of monophily implies that individual agency has limited effect on overall network structure and might explain why collaboration patterns that arise due to gender-preferences are so hard to change. These gender-driven tendencies in tie formation represent the last aspect of gender differences in music that we consider in this paper.

Gender classification Sonic feature, listener tag, institutional distribution, and collaboration network dimensions incorporate elements that are crucial to music production, while they also cover creative units ranging from individuals to teams and institutions. To increase the value of our comparisons, we trained *classifiers that predict whether a song was created by a man (1) or a woman (0)*. We gradually expanded our classifiers by adding different dimensions, i.e., sonic features, dummy variables for tags and record labels, and finally network structural features, and monitored the increase in the accuracy of the classification measured by the F-score. We ran logistic regressions and random forests using the `scikit-learn`[3] Python package. For the logistic regression models, we standardized the data and used lasso regularization strength of 10. In the random forest models, we built 300 trees. In both cases, we adjusted the weights to account for class imbalance. To eliminate the effects of

[3]http://scikit-learn.org/

potential bias towards productive artists, we performed all our experiments at the artist level by randomly selecting one song from each artist in our data. At every step of including a new group of features (i.e., sonic, tag, record label and network features), we performed 20 rounds of experiments and report F-scores based on averaging 20 predicted scores from both logistic regression and random forest models. The test sample consisted of a holdout subset comprising 2% of the songs chosen randomly from the entire dataset.

Results

This section presents our empirical results on gender differences in sound (i.e, musical and sonic features), listener attributions (i.e., tags), institutional distribution (i.e., record labels), and collaboration patterns. Then, it reports the results of our binary classification, compares the role of the considered dimensions, and evaluates feature importance.

Statistical comparisons

We compare songs released by men and women based on available sonic features and show the results in Table 1. Except for acousticness, female artists have lower average sound-level scores than male musicians. When comparing statistically the distributions for men and women, we find that aside of mode, all sonic features are significantly different for the two groups ($p < 0.001$, two-tailed Kolmogorov-Smirnov test). Mode is the highest level feature of songs, and hence it is not surprising that there are no significant differences between men and women in that respect.

Exploring the tags attributed by listeners, we find that nearly all of the 571 different fine-grained genres have been assigned at least once to male artists. However, listeners associated only less than half of the genres with female artists. Accounting for the over-representation of men, tempers this imbalance and shows that average male artists have 2.12 tags, while average female artists receive 1.96 tag attributions. Certain tags are statistically more frequently associated with one gender than the other. Based on a log-odds-ratio with an informative Dirichlet prior (Monroe, Colaresi, and Quinn 2008), rock, electronic, rap, techno, and reggae are linked with men, while pop, vocal, R&B, vocal jazz, and soul are attributed to women. Even though tags should reflect the actual sound of artists' music, we find a disparity in the tag distributions for the two genders as most people interpret music through pre-built gender stereotypes (James 2017). This gender-specific tag association could contribute to the difficulties of female musicians in overcoming barriers to enter and be recognized in new corners of the genre space.

Record label affiliations show similar patterns as tag attributions. Specifically, male artists work overall with nearly three times more record labels than female artists: 4,256 out of 4,873 companies are signing men and only 1,563 have a history of producing women's music. Using the same log-odds-ratio with Dirichlet prior (Monroe, Colaresi, and Quinn 2008), on the one hand, we identify Fania, ECM Records, Prestige Records, TriStar Music, and TrojanRecords to be significantly associated with male artists. On the other hand, Chansophone, Pony Canyon, Philips, Speedstar, and Chesky Records are the most salient labels for female musicians. From the major labels, MCA Records and Universal Music Group rank high on working with women. Just like in other industries, adequate female representation at an institution is linked with female leadership. As anecdotal confirmation, we find Michele Anthony, one of the most powerful women in the music industry, to be the Executive VP at Universal Music Group.

To uncover the differences in the way male and female musicians collaborate, first we examine their networks with respect to various notions of centrality (specifically, degree, eigenvector, closeness, and coreness). Being a highly connected artist in the center of the network secures a more advantageous position. Figure 3 shows the differences between men and women in terms of their degree centrality and coreness. The distributions reveal that, in general, the tails are longer for male artists. On average, female artists have 3.2 collaborators, male artists 3.9. The average coreness of women is 1.8, the average for men is 2.1. This indicates that female artists tend to have fewer collaborators and are more frequently on the periphery of the collaboration network. The distance between the distributions is statistically significant ($p < 0.005$, two-tailed Kolmogorov-Smirnov test). We find the same trends for eigenvector centrality and closeness with both distributions being statistically significantly different ($p < 0.001$, two-tailed Kolmogorov-Smirnov test). Finally, the average eigenvector centrality of male artists is two orders of a magnitude higher than the eigenvector centrality of female artists. Altogether, all four centrality measures point to significant differences in network position and indicate that men are better-connected and tend to be at the core of the collaboration network.

Within this network analytic framework, we also evaluate gender-preferences in forming collaborations. Based on the homophily index, there is no statistically significant difference between men and women's tendency to collaborate with women. However, as the distribution of the fraction of female collaborators in Figure 3 shows, there is a strong variance in this fraction for both genders. When we compute the overdispersion quantified as excess variance of gender-preferences in comparison with a homophily-only model (Altenburger and Ugander 2018), we find overdispersions of $\phi_F = 0.11$ and $\phi_M = 0.14$ that capture how much female (F) and male (M) artists vary in allocating their in-group versus out-group collaborations. To evaluate the usefulness of homophily and monophily in identifying the gender of song's creators, we include them in the models that we devise next in form of the fraction and number of female collaborators.

Classification models

Table 2 shows the performance of our models trained to predict whether a song was created by a male or female artist. As we add more and more dimensions, the number of artists in our training data (N_{train}) decreases. We find that relying simply on sonic features (the *Sound model*), we obtain an F-score of 0.62 with logistic regression and 0.87 with random forest. With various specifications of the

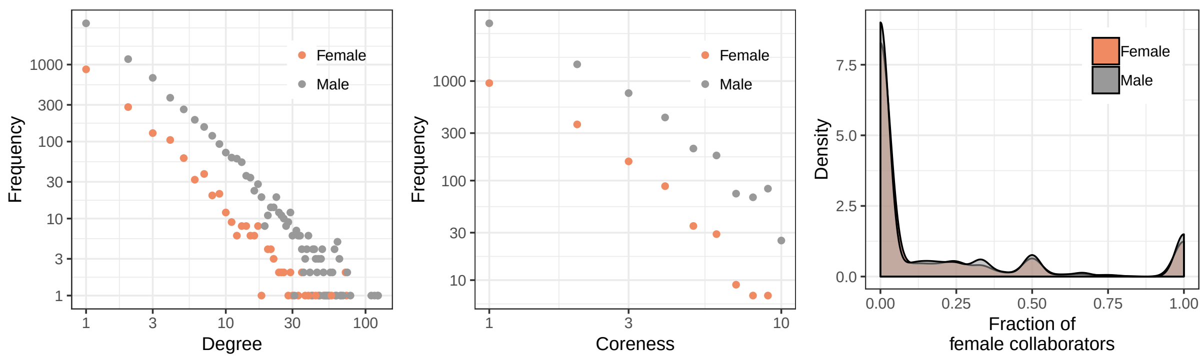

Figure 3: Differences in collaboration patterns. The log-log distributions of degree and coreness for the two genders show that men tend to have more collaborators and are in higher k-cores. Gender preferences in tie formation: Empirical distributions of the number of female collaborators for men and women show a tendency for monophily.

Model	Features	$N_{training}$	F_{LR}	F_{RF}
○ *Sound*	Sonic features	8,233	0.62	0.87
+ *Tags*	Listener attributions	8,010	0.75	0.88
+ *Distribution*	Record labels	8,010	0.94	0.91
+ *Network*	Collaborations	4,017	0.95	0.92

Table 2: Improvement of classification performance as dimensions are added to the model. Average F-scores with logistic regression (F_{LR}) and random forest (F_{RF}) models are comparable. Feature availability determines the size of the training sample (N_{train}). Results indicate that all considered dimensions incorporate aspects of gender differences.

sound model, we obtain scores that are considerably superior to random guessing. This indicates that algorithms trained with musical content features can successfully retrieve gender information. Adding listener-provided tag information to the model increases the performance to 0.75 with logistic regression and 0.88 using random forest (the *Sound+Tag model*). This increase in performance resonates with our expectation based on previous literature that artist gender identification improves upon adding listener tags (see section *Methods*). The sonic features and tags used here are a reflection of how music sounds to and is perceived by listeners. The F-scores suggest thus that factors related to sound and perception alone lead to uncovering a considerable part of the differences that emerge between the two genders' songs. Focusing on the contexts of and inputs into production, we further improve the predictive power of the model. When we add details about record label affiliations to account for characteristic distribution modalities, we increase the F-scores to 0.94 and 0.91, respectively (the *Sound+Tag+Distribution model*). Taking into account also the features describing collaboration patterns, the complete model (*Sound+Tag+Distribution+Network*) achieves an F-score of 0.95 and 0.92, respectively. The increase in F-scores to above 0.90 is enabled by the differences between male and female musicians in obtaining formal and informal support.

Overall, the performance of the logistic regression and random forest models is comparable. This observation holds also when we use other measures of evaluation like the AUC, accuracy, precision, and recall. We notice that the random forests built on our dataset are better at classifying male artists than logistic regressions: the recall of random forest is around 0.99 at all four levels, while the recall of logistic regressions goes from 0.49 in the case of the *Sound model* to 0.91 in the case of the full model. Conversely, logistic regression models capture female artists better: the precision of random forests varies between 0.78 and 0.85, remaining below the precision of logistic regressions (0.84 to 0.99).

To further unpack the role of these dimensions in differentiating the songs of men and women, we inspect the significance of used features. Table 3 highlights selected features along with their estimated coefficients and standard deviation in the logistic regression model that includes all four dimensions (adjusted R^2=0.84). Danceability and valence are statistically significant: higher danceability indicates female artists, while higher valence indicates male artists. Aside of tempo, all other sonic features are statistically significant in the complete model, albeit with small coefficients. Regarding tags and record labels, the associations of pop, R&B, and vocal jazz as well as Chansophone, Pony Canyon, and Philips with female solo artists follow the hypothesized patterns and are highly significant. Similarly, as anticipated, rock, rap, and techno as well as the labels Fania, ECM and Prestige Records are significant for men. With respect to the network features, eigenvector centrality and the fraction of female collaborators (i.e., the homophily index) are more useful predictors than degree and coreness, both being significantly associated with male artists.

Feature	Coefficient	SD
Sound features		
Danceability	-0.03	<0.01
Valence	0.04	<0.01
Tag features		
Pop	-0.24	<0.01
R&B	-0.12	<0.01
Vocal jazz	-0.32	0.02
Rock	0.23	<0.01
Rap	0.15	<0.01
Techno	0.06	0.02
Distribution features		
Chansophone	-0.70	0.02
Pony Canyon	-0.39	0.03
Philips	-0.20	<0.01
Fania	0.78	0.02
ECM Recods	0.15	<0.01
Prestige Records	0.03	<0.01
Network features		
Degree	-0.01	<0.01
Coreness	0.01	<0.01
Eigenvector	4.86	0.12
Fraction ♀ collaborators	0.10	<0.01

Table 3: Selected features of the complete regression model (adjusted R^2=0.84). All shown features are significant (p <0.01).

Discussion

A lack of understanding gender differences in the music industry might contribute to the low recognition of female artists and could result in gender-related disparities in artistic leadership and innovation. To address this problem, we evaluated gender differences in popular music along a set of dimensions that have been linked by previous literature to gender inequalities, stereotyping, and discrimination. We found that male and female artists' songs sound measurably different; artists tend to be associated by listeners with distinct genres and roles based on their gender; they are typically affiliated with different record labels; and male artists have more and better-connected collaborators in addition to being positioned in the core of the collaboration network. Our models, trained with features corresponding to these dimensions, suggest that we isolated key factors that distinguish male and female musicians' work.

This paper yields a number of interesting results. First, we found that there is an identifiable "female sound" in popular music. This result informs an ongoing debate in the literature (McClary 1991; Citron 1993; Halstead 1991; Sergeant and Himonides 2016). Our finding was enabled by a novel use of measures and global data. Specifically, we used quantitative sonic features to explore gender differences based on data that span 40 years of global popular music. Second, listener tags contributed considerably to the correct attribution of artist gender. This is closely aligned with the gender stereotyping discussed in previous research (Bem 1993; Sergeant and Himonides 2016). Third, the extent to which men and women differ in their affiliation with record labels is striking. Record labels accept male artists more easily than female musicians and the gender glass-ceiling has been apparent in artists' recognition: women accounted for less than 25% of the #1 hits between 1940 and 1990 (Dowd, Liddle, and Blylery 2005). Since awards and connections with record labels eventually reinforce the underrepresentation of female artists in the industry, an account based on a wide selection of record labels was important in quantifying the effects of institutional gatekeeping. Fourth, men are associated with a more central position in the collaboration network, which is likely to assure increased exposure to others' ideas and could result in more creative recombination (Seibert, Kraimer, and Liden 2001; Madlock-Brown and Eichmann 2016).

Our analysis here focused on identifying gender differences and evaluating the effect of these differences by testing the ability to distinguish male and female artists' songs based on features with no explicit gender connection. Note that we were not attempting to predict gender as a demographic indicator, which is an "immutable characteristic." Instead, we predicted a specific characteristic of a song, i.e., the gender of the creative entity behind it. Furthermore, this paper deliberately does not attempt to establish links between gender and commercial success or creative outcome. The reason for this is that artistic performance is quantified and evaluated in several, considerably distinct ways, which might favor one or the other gender. Since a thorough investigation into the different evaluation metrics in music is still missing, we refrain from making judgments about cultural value.

Although the dataset we analyzed is more comprehensive than was previously available, our findings are still constrained by it and the biases in our sample are not easily eliminated. As detailed in the *Data* section, *i.)* we only have information about binary genders, *ii.)* the crowdsourced parts of our data are likely to suffer from gender biases seen in other online repositories and open source platforms, and *iii.)* the data are skewed towards Western music. The extent to which these biases influence empirical studies about gender differences is a relevant avenue of investigation. We also hope that further analysis into the combination of cultural and structural factors that reinforce and sustain gender inequalities in music will enable us to make suggestions that address difficult-to-reverse gaps more efficiently than existing attempts.

Acknowledgments

We thank Michael Mauskapf for invaluable discussions and the anonymous reviewers for their helpful feedback. This work was supported in part by the U.S. National Science Foundation (IIS-1755873).

References

Altenburger, K. M., and Ugander, J. 2018. Monophily in social networks introduces similarity among friends-of-friends. *Nature Human Behaviour* 2:284–290.

Altenburger, K. M.; De, R.; Frazier, K.; Avteniev, N.; and Hamilton, J. 2017. Are there gender differences in professional self-promotion? An empirical case study of LinkedIn profiles among recent MBA graduates. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, 460–463.

Askin, N., and Mauskapf, M. 2017. What makes popular culture popular? Product features and optimal differentiation in music. *American Sociological Review* 82(5):910–944.

Baym, N. 2018. *Playing to the Crowd. Musicians, Audiences, and the Intimate Work of Connection*. New York University Press.

Bem, S. L. 1987. *Masculinity and Femininity exist only in the mind of the perceiver*. Oxford University Press. 290–304.

Bem, S. L. 1993. *The Lenses of Gender: Transforming the Debate on Sexual Inequality*. Yale University Press.

Bertin-Mahieux, T.; Ellis, D. P.; Whitman, B.; and Lamere, P. 2011. The million song dataset. In *In Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*.

Blau, F., and Kahn, L. 2000. Gender differences in pay. NBER Working Papers 7732, National Bureau of Economic Research, Inc.

Boimabeau, E. 2009. Decisions 2.0: The power of collective intelligence. *MIT Sloan Management Review* 50(2):45–52.

Brooks, A. W.; Huang, L.; Kearney, S. W.; and Murray, F. E. 2014. Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences* 111(12):4427–4431.

Burke, S., and Collins, K. M. 2001. Gender differences in leadership styles and management skills. *Women in Management Review* 16(5):244–257.

Citron, M. J. 1993. *Gender and the Musical Canon*. Cambridge University Press.

Crawford, E. 2016. Nielsen music year end report U.S. Technical report, Nielsen Music.

Csermely, P.; London, A.; Wu, L.-Y.; and Uzzi, B. 2013. Structure and dynamics of core/periphery networks. *Journal of Complex Networks* 1(2):93–123.

Currarini, S.; Jackson, M. O.; and Pin, P. 2009. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica* 77(4):1003–1045.

Dowd, T.; Liddle, K.; and Blylery, M. 2005. Charting gender: The success of female acts in the u.s. mainstream recording market, 1940-1990. *Research in the Sociology of Organizations* 23.

Friberg, A.; Schoonderwaldt, E.; Hedblad, A.; Fabiani, M.; and Elowsson, A. 2014. Using listener-based perceptual features as intermediate representations in music information retrieval. *The Journal of the Acoustical Society of America* 60:1951–1963.

Greenberg, D. M.; Kosinski, M.; Stillwell, D. J.; Monteiro, B. L.; Levitin, D. J.; and Rentfrow, P. J. 2016. The song is you: Preferences for musical attribute dimensions reflect personality. *Social Psychological and Personality Science* 7(6):597–605.

Grow, J. M., and Deng, T. 2014. Sex segregation in advertising creative departments across the globe. *Advertising & Society Review* 14(4).

Halstead, J. 1991. *The Woman Composer: Creativity and the Gendered Politics of Musical Composition*. Routhledge.

Hargittai, E., and Shaw, A. 2015. Mind the skills gap: the role of internet know-how and gender in differentiated contributions to Wikipedia. *Information Communication and Society* 18(4):424–442.

Hausmann, R.; Tyson, L.; Zahidi, S.; and Economic Forum, W. 2009. Global gender gap report.

Hochman, N., and Manovich, L. 2013. Zooming into an Instagram city: Reading the local through social media. *First Monday* 18(7).

Horvát, E.-A., and Papamarkou, T. 2017. Gender differences in equity crowdfunding. In *Proceedings of the Fifth Conference on Human Computation and Crowdsourcing*, 51–60.

James, R. 2017. Is the post- in post-identity the post- in post-genre? *Popular Music* 36/1:21–32.

Kanze, D.; Huang, L.; A. Conley, M.; and Higgins, E. 2017. We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal* 61:amj.2016.1215.

Karimi, F.; Wagner, C.; Lemmerich, F.; Jadidi, M.; and Strohmaier, M. 2016. Inferring gender from names on the Web: A comparative evaluation of gender detection methods. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, 53–54.

Keener, E. 2015. The complexity of gender: It is all that and more... In sum, it is complicated. *Sex Roles* 73(11):481–489.

Kleinberg, J.; Lakkaraju, H.; Leskovec, J.; Ludwig, J.; and Mullainathan, S. 2017. Human decisions and machine predictions. Working Paper 23180, National Bureau of Economic Research.

Kleinberg, J.; Ludwig, J.; Mullainathan, S.; and Rambachan, A. 2018. Algorithmic fairness. *AEA Papers and Proceedings* 108:22–27.

Koppman, S. 2014. Making art work: Creative assessment as boundary work. *Poetics* 46:1–21.

Kuhn, P. J., and Villeval, M.-C. 2013. Are women more attracted to cooperation than men? Working Paper 19277, National Bureau of Economic Research.

Lamont, M., and Molnár, V. 2002. The study of boundaries in the social sciences. *Annual Review of Sociology* 28(1):167–195.

Lutter, M. 2015. Do women suffer from network closure? the moderating effect of social capital on gender inequality in a project-based labor market, 1929 to 2010. *American Sociological Review* 80(2):329–358.

Madlock-Brown, C., and Eichmann, D. 2016. The scientometrics of successful women in science. In *2016 IEEE/ACM*

International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 654–660.

McClary, S. 1991. *Feminine Endings: Music, Gender and Sexuality*. Univ Of Minnesota Press.

McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1):415–444.

Monroe, B. L.; Colaresi, M. P.; and Quinn, K. M. 2008. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16(04):372–403.

Negus, K. 1999. *Music Genres and Corporate Cultures*. Routhledge.

Newman, M. 2010. *Networks: An Introduction*. Oxford University Press, 1st edition.

Newman, M. 2018. Where are all the female music producers? *Billboard*.

O'Neill, S. A., and Boultona, M. J. 1996. Boys' and girls' preferences for musical instruments: A function of gender? *Psychology of Music* 24(2):171–183.

Pollitt, K. 1991. Hers; the Smurfette principle. *The New York Times*.

Ragins, B. R. 1998. Gender gap in the executive suite: CEOs and female executives report on breaking the glass ceiling. *The Academy of Management Executive (1993-2005)* 12(1):28–42.

Richards, C.; Bouman, W. P.; Seal, L.; Barker, M. J.; Nieder, T. O.; and T'Sjoen, G. 2016. Non-binary or genderqueer genders. *International Review of Psychiatry* 28(1):95–102.

Robles, G.; Arjona Reina, L.; Serebrenik, A.; Vasilescu, B.; and González-Barahona, J. M. 2014. Floss 2013: A survey dataset about free software contributors: Challenges for curating, sharing, and combining. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, MSR 2014, 396–399.

Seibert, S. E.; Kraimer, M. L.; and Liden, R. C. 2001. A social capital theory of career success. *Academy of Management Journal* 44:219–237.

Sergeant, D. C., and Himonides, E. 2014. Gender and the performance of music. *Frontiers in Psychology* 5:276.

Sergeant, D. C., and Himonides, E. 2016. Gender and music composition: A study of music, and the gendering of meanings. *Frontiers in Psychology* 7:411.

Smith, S. L.; Choueiti, M.; and Pieper, K. 2017. Inequality in 900 popular films: Examining portrayals of gender, race/ethnicity, lgbt, and disability from 2007-2016. Technical report, Media, Diversity, and Social Change Initiative.

Smith, S. L.; Choueiti, M.; and Pieper, K. 2018. Inclusion in the recording studio? Gender and race/ethnicity of artists, songwriters and producers across 600 popular songs from 2012-2017. Technical report, Annenberg Inclusion Initiative.

Spitz, A., and Horvát, E.-Á. 2014. Measuring long-term impact based on network centrality: Unraveling cinematic citations. *PLOS ONE* 9(10):e108857.

Treitler, L. 2011. *Reflections on Musical Meaning and Its Representations*. Indiana University Press.

Uzzi, B., and Spiro, J. 2005. Collaboration and creativity: The small world problem. *American Journal of Sociology* 111(2):447–504.

Vasilescu, B.; Posnett, D.; Ray, B.; van den Brand, M. G.; Serebrenik, A.; Devanbu, P.; and Filkov, V. 2015. Gender and tenure diversity in github teams. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 3789–3798.

Vedres, B. 2017. Forbidden triads and creative success in jazz: the Miles Davis factor. *Applied Network Science* 2(1):31.

Wachs, J.; Hannák, A.; Vörös, A.; and Daróczy, B. 2017. Why do men get more attention? exploring factors behind success in an online design community. In *Proceedings of the Eleventh International Conference on Web and Social Media*, 299–308.

Wagner, C.; García, D.; Jadidi, M.; and Strohmaier, M. 2015. It's a man's wikipedia? Assessing gender inequality in an online encyclopedia. In *Proceedings of the Ninth International Conference on Web and Social Media*, 454–463.

West, C., and Zimmerman, D. H. 1987. Doing gender. *Gender & Society* 1(2):125–151.

Yang, J., and Leskovec, J. 2014. Overlapping communities explain core-periphery organization of networks. *Proceedings of the IEEE* 101(12):1892–1902.