QUANTIFYING MUSIC TRENDS AND FACTS USING EDITORIAL METADATA FROM THE DISCOGS DATABASE

Dmitry Bogdanov, Xavier Serra

Music Technology Group, Universitat Pompeu Fabra

`dmitry.bogdanov@upf.edu, xavier.serra@upf.edu`

ABSTRACT

While a vast amount of editorial metadata is being actively gathered and used by music collectors and enthusiasts, it is often neglected by music information retrieval and musicology researchers. In this paper we propose to explore Discogs, one of the largest databases of such data available in the public domain. Our main goal is to show how large-scale analysis of its editorial metadata can raise questions and serve as a tool for musicological research on a number of example studies. The metadata that we use describes music releases, such as albums or EPs. It includes information about artists, tracks and their durations, genre and style, format (such as vinyl, CD, or digital files), year and country of each release. Using this data we study correlations between different genre and style labels, assess their specificity and analyze typical track durations. We estimate trends in prevalence of different genres, styles, and formats across different time periods. In our analysis of styles we use electronic music as an example. Our contribution also includes the tools we developed for our analysis and the generated datasets that can be re-used by MIR researchers and musicologists.

1. INTRODUCTION

In this paper we propose to explore the editorial metadata available in the Discogs [1] database and show how its analysis can be used as a potential tool to support musicological studies and research in music information retrieval (MIR). Discogs is one of the largest online databases of editorial metadata [2] used by music collectors and enthusiasts. It hosts all metadata under Public Domain license and provides complete monthly data dumps [3] which makes it very easy to access and re-use the data.

---

[1] `https://discogs.com`
[2] Discogs mission statement is "to build the biggest and most comprehensive music database and marketplace".
[3] `https://data.discogs.com/`

Discogs metadata contains information about music releases (such as albums or EPs) including artists name, track list including track durations, genre and style, format (e.g., vinyl or CD), year and country of release. It also contains information about artist roles and relations as well as recording companies and labels. The quality of the data in Discogs is considered to be high among music collectors because of its strict guidelines, moderation system and a large community of involved enthusiasts. The database contains contributions by more than 347,000 people. It contains 8.4 million releases by 5 million artists covering a wide range of genres and styles (although the database was initially focused on Electronic music).

Remarkably, it is the largest open database containing explicit crowd-sourced genre annotations. Discogs implements a two-level genre hierarchy including top-level broad *genres* and more specific sub-genres called *styles*. [4] This taxonomy is determined by the community moderators and there are guidelines for annotation. Genre/style labels are non-exclusive, meaning that a release can be annotated by multiple genres and styles. Releases almost always contain both genre- and style-level annotations as both are required by the submission system of the database.

To our surprise, little is known about Discogs metadata among the MIR community and there is a lack of musicological studies using this data. We identified two MIR studies, using this data in the past. One study used it for music recommendation by building and comparing artist tag-cloud profiles including genres, styles, record labels, years, and countries associated with each artist [3]. Another study analyzes graphs of artist collaborations in order to identify music clusters that can then be associated with genres [5].

There has been done some MIR research in the context of mining genre annotations using other music databases, including AllMusic [14], Wikipedia [16] and Last.fm [15], however, all these studies are limited in their scope and their use-cases of the data. Furthermore, the employed data sources are either ill-structured (Wikipedia, Last.fm), miss explicit genre information (Last.fm), or contain proprietary data (AllMusic) which cannot be reused on the large scale. In order to address these issues we are currently creating a new genre metadata corpus for MIR tasks, including Discogs metadata [10].

---

[4] `https://reference.discogslabs.com/wiki/Style-Guide`

In this paper we propose new ways of how editorial metadata can be used in MIR, specifically in the context of studies directed to assist musicological research. To this end, we present a number of example studies using the data available in Discogs in which we identify and quantify some music trends and facts, some of which were previously documented by musicologists and music journalists. In these examples we consider overall trends across broad genres together with a more specific analysis for the case of electronic music which commonly lacks attention of the research community.[5] Our analysis is exploratory and is not aimed at concrete musicological conclusions. Instead we present how data can be used to identify interesting facts and raise questions for further research. Importantly, our study is focused exclusively on music recorded and published on physical or digital media in form of collectable artifacts. We share the tools we developed for the analysis, preprocessed datasets and the complete results for future use by the MIR community and musicologists.[5]

Figure 1: Genre coverage (%).

Figure 2: Number of releases per country by year.

2. DISCOGS DATASET

We downloaded the data dump of music releases dated April 2017. Each release contains the information about a specific album, single, EP, or compilation.[6] For our exploratory study we are interested in artist name, tracklist including track durations, genre and style, format, year and country of each release.

As we wanted to perform per-year analysis of the data, we decided to discard all releases from the ongoing year 2017 to avoid any bias due to incompleteness of data.[7] We extracted the desired data from the dump. The resulting dataset includes 7,954,870 releases by 1,290,943 artists with the total of 67,895,500 released tracks. All releases are annotated by 11 genres and 442 styles.[8] Approximately half of the releases (52%) are annotated with track durations, the 11% of releases are annotated as compilations, and around 1% are marked as mixed.

We estimated the overall genre coverage in terms of percentage of releases, and tracks and artists associated with those releases, out of their total number in the dataset (Figure 1). The rationale behind counting track numbers is that many releases in the databases are vinyl EPs and singles with a smaller number of tracks than albums or compilations. The number of artists provides an alternative useful estimate, as artists may vary in their "release productivity". Our inspection revealed the predominance of the Rock, Electronic, and Pop genre categories in the database,
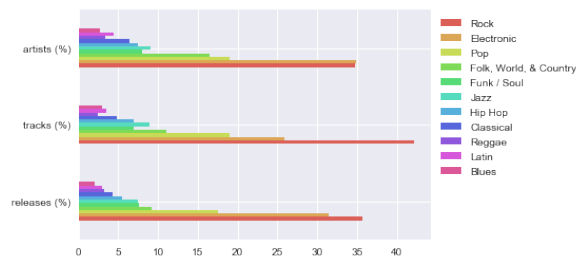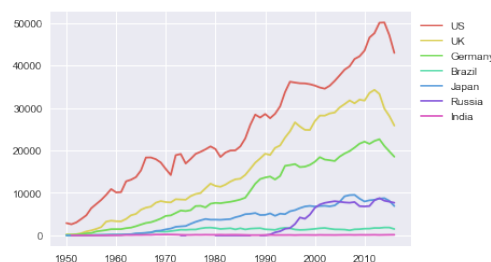
with the largest styles being Pop Rock and House. Still, even the least represented genres (such as Blues) have almost 200,000 releases. Around 90% and 74% of styles have more than 1000 and 10,000 releases, respectively.

Inspecting country distribution for releases, we can see an overall predominance of music from western countries in the dataset. Top countries included US, UK, Germany, France, Italy, Japan and Netherlands. Figure 2 presents an example of total number of releases published in various countries by year. Our observations lead us to an open question of whether the disbalance in the distribution represents the actual evolution of the recording industry in each country, or that the Discogs database has insufficient coverage of music from some countries/cultures. While both reasons are plausible, we suppose that our data is still valid for research focused on western countries and music.

3. EXAMPLE STUDIES

3.1 Average Track Duration

In our first example study we analyzed the distribution of the duration of tracks[9] across different genres and styles and its evolution in time. To this end, we discarded all mixed compilations that contain only fragments of tracks instead of full recordings. As there are music releases annotated by multiple genres, we computed tracks duration statistics twice: first, including all releases annotated by a particular genre, and second, excluding releases that were also annotated by other genres.

Figure 3 presents box plots with the obtained results. In both cases, Classical, Electronic, and Jazz music has the largest median durations and the largest variability, ac-

---

[5] We refer the reader to the additional materials online including code and the detailed analysis results: `https://github.com/dbogdanov/ismir2017-discogs`

[6] See an example of a release page: `https://www.discogs.com/LFO-Frequencies/release/3649`

[7] We can also suspect that the database is still missing some releases for 2016 as there was possibly not enough time to gather contributions from the Discogs community. Still, we decided to keep those releases.

[8] For simplicity, we ignored *Brass & Military*, *Children's*, *Non-Music*, and *Stage & Screen* genre categories present in the Discogs taxonomy due to being less represented and/or not being strictly related to music. See the complete genre taxonomy at `https://github.com/dbogdanov/ismir2017-discogs/tree/master/taxonomy`.

---

[9] Note that we are considering duration of tracks on a recording medium, not of the original music pieces. The former can be seen as a proxy for the latter, at least for some of the music styles.

(a) all releases

(b) releases annotated by a single genre
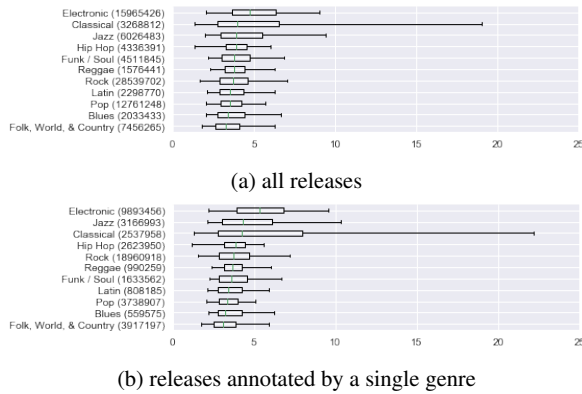
Figure 3: Boxplot of track durations (mins.) for each genre. Number of computed tracks is given in brackets. Whiskers are set at 5% and 95%.

(a) Pop

(b) Electronic

(c) Hip Hop

(d) Classical

Figure 4: Evolution of duration of Rock, Electronic, Hip Hop, and Classical tracks (mins.) by year.

cording to the interquartile range. We observed similar results when including and excluding releases with multiple genre annotations. Still there are some variations in median values, quartiles and whiskers positions, most notably for Classical, Electronic and Jazz. This suggests that music annotated by these genres on the crossover with other genres tends to be shorter. The duration of some music tracks by these genres reached over 8 minutes in contrast to other genres that were far shorter. [10]

Similarly, we analyzed all styles present in the Discogs taxonomy. [11] Overall comparison suggests that the median duration for the majority of styles is below 10 minutes, the shortest starting at 2 to 2.5 minutes (e.g., Grindcore, Crust, Surf, Doo Wop, Beat and Rockabilly). We can observe that some styles can be associated with higher variability in the duration of tracks, while others have durations condensed around a common value. For example, in the case of Electronic music, we can identify a cluster of styles with a large variability (including Harsh Noise Wall, Drone, Noise, Musique-Concrete, Berlin-School, and Dark Ambient) with the duration of tracks surpassing 15 minutes. All these styles share a unconventional, or experimental, approach to sound and music composition. In contrast, the Electronic styles with the lowest variability (such as Eurodance, Jumpstyle, or Grime, among many others) are commonly related to dance/club music.

As a next step, we checked whether there is a global trend in change of average track duration across years. We computed per-genre distributions of track durations across years, some of them shown in Figure 4. [12] We observe existence of a time period with a clear tendency of increase in almost all genre categories: Blues (early 60s to early 70s), Folk, World & Country (mid-60s to 00s), Funk/Soul (early 60s to late 70s), Jazz (late 60s to late 70s), Latin (mid-60s to mid-00s), Pop (mid-60s to mid-80s), Rock (dramatic increase in mid-60s to early 70s) and Electronic (early 70s to 2010 with a consequent decrease). It appears that all
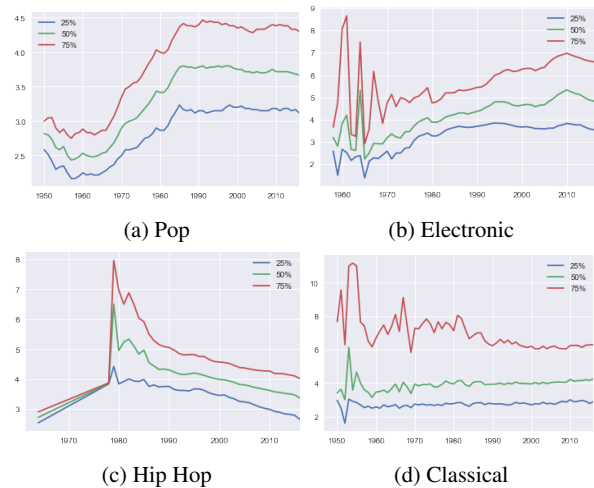
these genres have reached a plateau in median track duration after a period of stable increase. We can also see an increment in variability of durations with time.

The increase in duration may be associated with the change of record formats, which we propose to assess in future studies. Interestingly no such tendency was found for Classical music. In contrast, while median duration remains constant, we can observe a decrease in variation with the longer half of the tracks getting shorter. Furthermore, Hip Hop tracks are steadily decreasing in duration since the origin of the genre in the late 70s. All these findings can raise questions for further musicological research.
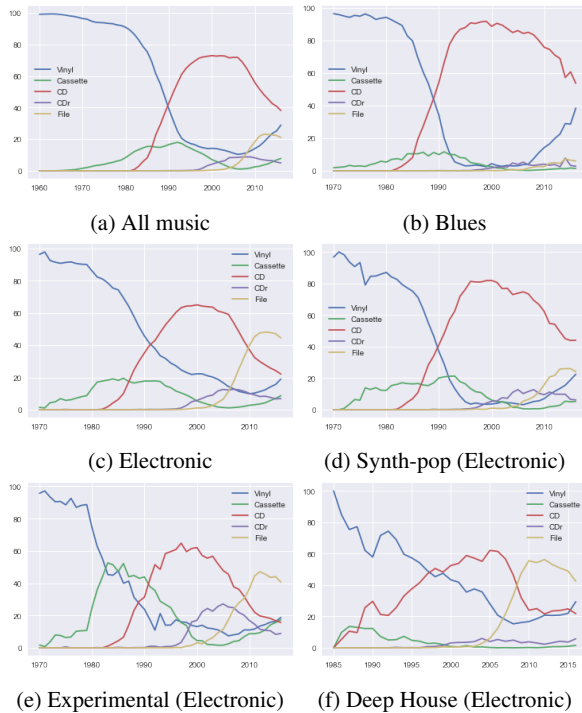
3.2 Release Formats

There is a large number of release formats registered in the Discogs database. [13] In our next example study we quantify the evolution of the most common formats of the past half-century: vinyl, cassettes, CD/CDr, and files. [14] We compare the amount of music released on each of these formats from the 1960s to nowadays. To this end, we count the overall number of releases and tracks recorded on a particular medium. We can then compute the percentage of music released on each format each year. Track and release percentages can vary significantly due to the capacity of each medium: typically CD releases contain a larger number of tracks then vinyl. Figures 5a, 5b and 5c present track statistics for all music in the dataset, and for Blues and Electronic genres in particular. [15]

Overall, our analysis corroborates existing RIAA sales reports [12]. We can evidence the commonly known rapid growth of CD format from the mid-80s to the late 90s, followed by a plateau period till the mid-00s. The following decline of CD can be clearly associated with the growth of digital file formats. Remarkably, vinyl, following its de-

---

[10] Of course, this only suggests a general trend excluding outliers.

[11] See results for all styles grouped per genre in additional materials.

[12] In the case of Electronic and Classical, large value jumps in early years may be associated to small amount of tracks annotated by duration.

[13] https://www.discogs.com/help/formatslist

[14] Note that we cannot be confident in estimations for the file format as digital releases may be under-represented in the Discogs database.

[15] See additional materials for complete results.

(a) All music

(b) Blues

(c) Electronic

(d) Synth-pop (Electronic)

(e) Experimental (Electronic)

(f) Deep House (Electronic)

Figure 5: Percentage of released tracks per format by year (%).

cline since the mid-80s, is now growing since the early 10s, which corroborates recent observations of the new "vinyl boom" [2].

Cassette releases appear to be always below vinyl or CD releases through the history of the format, descending to its supposed death in the mid-00s. However, since then, we can observe the second growth of cassettes, which confirms observations of the growing "cassette culture" by some music journalists and musicologists [8,7,6]. Interestingly, there is a considerable amount of music released on a CDr format, which appeared in the late 90s and achieved its maximum at the time of the death of cassettes. We can suppose that the observed CDr and cassette trends are linked to the DIY culture of independent music distribution [18].

Analyzing particular genres, we identified Blues, Rock, Reggae, and Funk/Soul to have the highest percentage of tracks released on vinyl in the recent years, surpassing 30%. Remarkably, these genres can be nowadays considered somewhat "old-school", and therefore of a potentially higher interest among vinyl music collectors. We hypothesize that many of new vinyl releases are reissues.

Finally, we ran per-style analysis of data, on the example of Electronic music, in order to identify peculiarities of music distribution within certain styles. Figures 5d, 5e and 5f demonstrate differences in formats on the example of Synth-pop, Ambient and Deep House. From our results we can evidence a transition to the predominance of CDs for all considered styles with the turning points [16] start-

---

[16] The year of an equal number of tracks released on both formats.

ing since 1988. The styles that moved to CD first were Ambient, Synth-pop and Experimental, with their turning points in transition between 1988 and 1991. We speculate that early transition to CD was at least partially motivated by the demand of home consumers, meanwhile other styles, supported by DJs, had a technical demand for vinyl. Such styles had their transition point to CD later between 1993 (e.g. House, Techno) to 1996-97 (Drum n Bass, Deep House). Remarkably, we identified the existence of styles with the absolute predominance of digital formats (CD, CDr and file). In the case of Glitch, this fact may be linked to musical characteristics of the style, which make releases on vinyl/cassettes aesthetically or technically unfeasible.

In 2016 the digital file format is leading in almost all styles. The turning point towards its predominance appeared between 2007 and 2011. Still, we can observe a trend in growth of vinyl in the recent 6 years. Moreover we can also see the growth of cassettes since 2005, with the most significant example being Experimental electronic music (reaching almost 20% of tracks being released on cassettes in 2016). Interestingly, almost 30% of released Ambient tracks in the early-to-late 80s, and similarly over 40% of Experimental from the early 80s to the early 90s, were released on cassettes (which again supports the existence of the DIY cassette scene in experimental music of the 80s) [17].

3.3 Genre and Style Trends

In this section we present another use-case example: an analysis of genre and style trends across time periods. We consider overall trends in genres and exemplify style analysis on Electronic music. Again, we quantified the amount of music in terms of number of releases, tracks and artists associated with those releases. Their absolute values (the amount of music in year N by genre G) and proportions (the percentage of music from year N by genre G) allow us to suggest possible trends. Figure 6 presents results of the analysis. [17] Below we summarize our observations on the tracks level.

Rock appears to be the major genre since the late 60s covering more then 40% of released tracks since the late 70s and reaching 50% nowadays. It is currently followed by Electronic music that reached its peak at 38% in 2011. Pop music is in a steady decline since its short dominance in the mid-60s, falling below Electronic since the early 90s. We can also see the decline of Jazz after its huge 50% peak in the mid-50s, being a predominant genre at that time, the rise and fall of Funk/Soul, and the growth of Hip Hop being the 4th leading genre in the 2000s. A valley in the Rock plot in the mid-70s corresponds to the peak of Funk/Soul (including "the disco boom" [9]). [18]

We then repeated the analysis for styles of Electronic music. Again, we used absolute and relative release/track/artist counts. The relative values represent the

---

[17] Note that the percentage values do not sum up to 100 because releases can be annotated by multiple genres.

[18] This is especially well seen on the release and artist-level plots, with the number of Funk/Soul artists being temporarily higher than for Rock.

(a) Tracks (all genres)   (b) Releases (all genres)   (c) Tracks (Electronic)
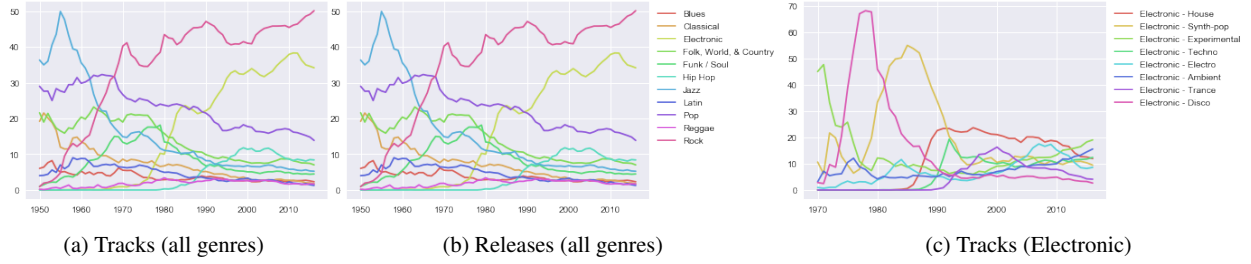
Figure 6: Percentage of released music per genre and style by year (%).

percentage of music within a style out of the total amount of Electronic music released each year. Out of 110 Electronic styles, we identified the most important ones in terms of their overall presence: we computed their relative share in all Electronic music across years from 1970 to 2016. We then summed these values for each style and identified the styles with the highest values. Figure 6c presents track-level results for those styles. [19]

According to our data, we can see how electronic Disco peaked in the late 70s followed by the Synth-pop peak in the mid-80s, both being the predominant styles of their time. The decline of Synth-pop in the late 80s/early 90s met the peak of House (1990) and the first peak of Techno (1992). Later, growing styles included Trance (peaked in 2000), and Electro (having its second peak in the late 00s, the first one in 1984). We can also guess the period of birth of each genre using our data (e.g., house in the mid-80s, techno in the late 80s, and trance in the early 90s). All of these observations seem to be well-aligned with the existing literature on the history of electronic music [11].

Interestingly, after the year 2010 we observe lower percentage values for all styles. This suggests the diversification of electronic music: more styles are taking a share in the amount of music released each year. Finally, it is worth noting that similar analyses can be run on per-country basis. This can be useful for identifying potential regional trends or analyzing the following of a particular style in various countries. [20]
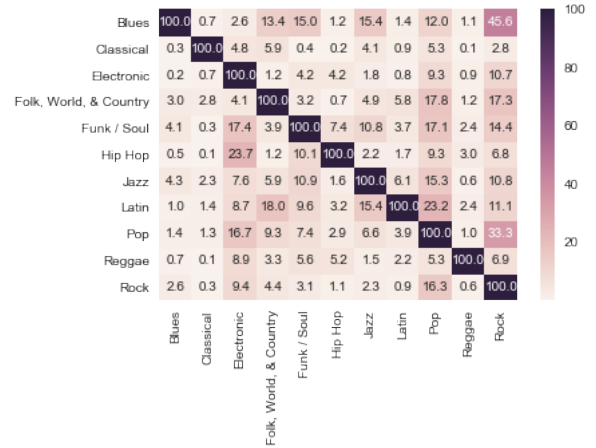
3.4 Genre and Style Co-occurrences

In this section, we consider another use-case for editorial metadata with genre/style annotations and study co-occurrences between different genre and style labels. We also attempt to assess the specificity of labels: while top-level genre categories are very broad, styles may vary a lot in their specificity and coverage.

Given that releases can be annotated by multiple genre/style labels within the Discogs taxonomy, we computed a genre (and style) co-occurrences matrix in order to identify possible relations. For each pair of genres (or styles) X and Y we counted the number times both appear on the same release across all releases in the dataset. The

---
[19] See additional materials for full results.
[20] For example, we could observe that Hardcore, Breakbeat and Drum n Bass styles, well-represented in UK in the early-to-mid 90s, were never prevailing in Germany.
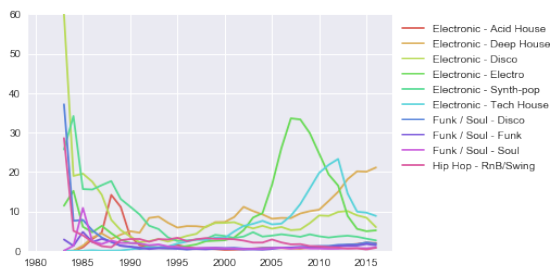

Figure 7: Genre co-occurrences (%).
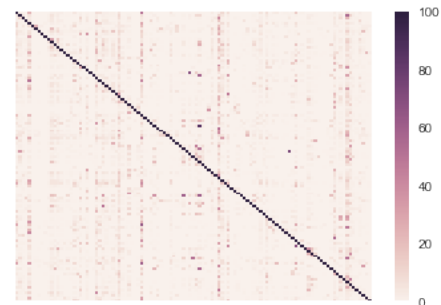
Figure 8: Electronic styles co-occurrences (%).

resulting matrix is asymmetric and its values represent the percentage of music by genre (style) X (on the x-axis) also being annotated by genre (style) Y (on the y-axis).

Figure 7 presents the resulting genre co-occurrences matrix computed using all our data. We can conclude from it that Classical, Electronic, and Reggae seem to be the genres that are well isolated from others, that is, the music under these genre labels is self-contained and all co-occurrences with other genres are relatively small (below 11%). On the contrary, the most interconnected genres are Blues (46% of it is also Rock), Pop (33% is Rock) and Hip-Hop (24% is Electronic). We can also observe how Pop and Rock, and Electronic appear to be commonly co-occurring genres for many genres, probably, due to being the most popular ones.

Figure 9: Percentage of House releases also annotated by other styles by year (%).

We proceeded with style co-occurrences in a similar manner. The resulting matrix is huge and here we analyze a portion, again, for the case of Electronic music.[21] To give an idea, Figure 8 provides an overview of this matrix containing 110 Electronic styles. As we can see, there are a number of traceable vertical lines corresponding to particular styles that often co-occur with many other styles. Those include Techno, House, Experimental, Synth-pop and Ambient — all being among the most frequent styles within Electronic music in our dataset. Less predominant, but still traceable vertical lines include Electro and Downtempo. We can suppose that these style labels are wide in coverage or generic enough to embrace some other styles. In contrast, we can also see styles that typically do not co-occur with others (e.g., Speed Garage or New Beat), which might indicate their high specificity or "nicheness".

Summing all row values for a column we can get a "genericness-vs-nicheness" score. According to it, some examples of niche styles are Beatdown, Neo Trance, Skweee, New Disco, and Italo House, while the most generic styles include Downtempo, Synth-pop, Electro, Ambient, Techno, Experimental, and House. On the other side, co-occurrence values above 50% might indicate a subgenre-to-genre relation and give us a degree of potential "sub-styleness", or "derivativeness". For example, 86% of Gabber is also annotated as Hardcore, 75% of Hardbeat as New Beat, and 68% of Power Electronics as Noise. After identifying all such examples in the matrix, we were able to corroborate this hypothesis.

Interestingly, it is also possible to compute co-occurrences for particular time epochs. Figure 9 illustrates the idea of evolution of style co-occurrences in time on the example of House. We can see how Disco, Synth-pop, Funk and Soul potentially had an influence on the style at the time of its origin (indeed these styles are often cited as such [11]), followed by a peak of Acid-House and then, later, Electro and Tech-House.

The co-occurrence matrices demonstrate the intrinsic variability in genre annotations and we believe that such data can be very useful for the MIR community in the context of evaluating music genre classifiers and for other tasks. Indeed, some studies on audio-based genre classification (such as [13, 1, 4]) reveal similar patterns in misclassifications, and they can be supported by our data.

---

[21] See additional materials for full results.

4. DISCUSSION

Naturally, the presented data analysis is limited by the coverage of the Discogs database, with a possible bias towards Western music and collectable music items, and other sociocultural factors. Digital releases are possibly underrepresented since the new online distribution models allow artists to instantly share their work and the concept of "release" might be changing. We are far more confident in the data for the former time period of predominance of physical releases. Assessment of coverage of editorial music databases is an open question for future research.

Our analysis is essentially grounded on the statistics of music production, not consumption. No analysis of music trends in terms of popularity among listeners is addressed. Instead, we deliberately focused on another aspect: what music artists tend to produce, including in the long tail. A release of 100 copies is treated equally to a release of thousands in our analysis.

We are also aware of the problem of release-level genre/style annotations: labels do not necessarily apply to all tracks. Still, we suppose a certain congruency between tracks on a release. Interestingly, our data reveals that even releases with a single track have multiple labels (1.2 genres and styles on average). This suggests that a genre annotation problem is inherently multi-label. Finally, in our analysis we are limited to the Discogs' genre taxonomy. Their broad genre categories might not be appropriate for some research tasks, but we can be much more confident in style annotations, at least for some genres.

Many music releases actually correspond to the same conceptual items (e.g., album CD version, CD version in another country, vinyl version, reissue). Discogs provides information about their groupings, and it should be considered depending on a task at hand. For simplicity, in our example studies we treat such releases as if they were independent. Such releases are often released in various formats, countries, years, and can have different track lengths, bringing additional information to our analysis.

5. CONCLUSIONS

In this paper we propose to take a closer look on the editorial metadata in the Discogs database. We believe that analysis of this data can be a valuable tool for researchers. It can help to identify and analyze various musical phenomena and raise different musicological questions. Importantly, Discogs is one of the largest sources for such data in the public domain which allows to address potential research questions on a very large scale. We demonstrated the use of this data in a number of example studies in which we attempted to quantify a number of music trends and facts, some previously documented by musicologists and music journalists. Our examples are far from being complete and of course there are more potential questions to be raised and addressed using this data. We share the analysis tools we developed, our preprocessed datasets and the complete results for our example studies for further re-use by MIR and musicology researchers.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] B. K. Baniya and J. Lee. Importance of audio feature reduction in automatic music genre classification. *Multimedia Tools and Applications*, 75(6):3013–3026, 2016.

[2] D. Bartmanski and I. Woodward. The vinyl: The analogue medium in the age of digital reproduction. *Journal of Consumer Culture*, 15(1):3–27, 2015.

[3] D. Bogdanov and P. Herrera. Taking Advantage of Editorial Metadata to Recommend Music. In *International Symposium on Computer Music Modeling and Retrieval (CMMR'12)*, 2012.

[4] D. Bogdanov, A. Porter, P. Herrera, and X. Serra. Cross-collection evaluation for music classification tasks. In *International Society for Music Information Retrieval Conference (ISMIR'16)*, 2016.

[5] J. Burke, R. Rygaard, and Z. Yellin-Flaherty. Clam!: Inferring genres in the Discogs collaboration network. 2014.

[6] J. Demers. Cassette tape revival as creative anachronism. *Twentieth-Century Music*, 14(1):109–117, 2017.

[7] C. Eley. Technostalgia and the resurgence of cassette culture. *The Politics of Post-9/11 Music: Sound, Trauma, and the Music Industry in the Time of Terror*, pages 43–55, 2011.

[8] M. Hogan. This is not a mixtape. `http://pitchfork.com/features/article/7764-this-is-not-a-mixtape/`, 2010. Accessed on 13.07.2017.

[9] T. Lawrence. *Love saves the day: A history of American dance music culture, 1970–1979*. Duke University Press, 2004.

[10] A. Porter, D. Bogdanov, and X. Serra. Mining metadata from the web for AcousticBrainz. In *International workshop on Digital Libraries for Musicology (DLfM'16)*, pages 53–56. ACM, 2016.

[11] S. Reynolds. *Energy flash: A journey through rave music and dance culture*. Soft Skull Press, 2012.

[12] RIAA. U.S. sales database. `https://www.riaa.com/u-s-sales-database`, 2017. Accessed on 13.07.2017.

[13] N. Scaringella, G. Zoia, and D. Mlynek. Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine*, 23(2):133–141, 2006.

[14] A. Schindler, R. Mayer, and A. Rauber. Facilitating comprehensive benchmarking experiments on the million song dataset. In *International Society for Music Information Retrieval Conference (ISMIR'12)*, 2012.

[15] H. Schreiber. Improving genre annotations for the Million Song Dataset. In *International Society for Music Information Retrieval Conference (ISMIR'15)*, 2015.

[16] H. Schreiber. Genre ontology learning: Comparing curated with crowd-sourced ontologies. In *International Society for Music Information Retrieval Conference (ISMIR'16)*, 2016.

[17] J. Scott. The Noise-Arch archive. `https://archive.org/details/noise-arch`, 2015. Accessed on 13.07.2017.

[18] R. Strachan. Do-It-Yourself: Industry, ideology, aesthetics and micro independent record labels in the UK. *Unpublished masters thesis, University of Liverpool, Liverpool, England*, 2003.