# arakelyan\_mining\_2018

Live performances are a crucial part of the life of a music artist.

11

discovered a connection between live events and increased digital listenership [37] (which is the second highest source of income for a band after live performances).

11

playing at top venues goes hand-in-hand with getting more digital listeners, which in turn may increase their likelihood of being signed with major music labels

11

we aim to determine whether it is possible to model and predict these career trajectories

11

we pose the question: is it possible to find open data to understand and forecast careers and success in the music industry?

11

Songkick and Discogs have sprung up to create centralized repositories of music events and music artists

11

this goldmine also allows researchers to model the music industry dynamics

11

In this paper, we are interested in the problem of characterizing and understanding the career trajectories of the artists across different genres

11

Is the choice of venues where an artist performs correlated with the eventual success of that artist (for a given definition of success)?

11

Can we predict which venues an artist/band will perform based on the history of his/her/their past performances?

11

How do we measure the relative importance of performances in specific venues and their impact on career trajectories, and how do we jointly characterize influential artists and venues?

11

We construct and present a new dataset by collecting all of the artists and concerts from the Songkick platform, and supplement this dataset with information from Discogs

12

We define a measure of success based on whether an artist has signed a contract with one of the major music record labels,

12

We demonstrate the viability of forecasting future performances of artists, and therefore their success, based on the history of past performances

12

digital music has been linked to fall in record sales, airplay and charts no longer adequately measure popularity, given numerous streaming services and listenership outside of them

12

some researchers look at the popularity of music artists on digital delivery platforms like Last.fm,

12

Success in post-digital music world can still be adequately represented by contracts with major labels.

12

Music record labels are still important players in the industry

12

even though theoretically digital technologies allow artists to perform production, promotion and sales on their own, practically this doesn't happen very often

12

Songkick3 is a concert-discovery platform that aims to link fans to artists' events.

12

The "gigography" of an artist is the term that Songkick uses to refer to all of that artist's events.

12

However, we did not consider headliners and support artists separately in the analysis presented further.

13

For each concert in the gigography, we extracted the following information: ID, date, city, country, state (if applicable), latitude and longitude of the venue, venue ID and venue name, name of the event and its popularity score

13

Collected data was organized into separate artist, event, and venue data frames

13

Each artist is indexed by its Songkick and Discogs IDs. Venues and events are indexed by their Songkick IDs

13

There are also several lists of cross-references: mapping venues to the events that happened there, and events to the venues where they took place.

13

Due to the fact that the goal of Songkick is connecting fans to their favorite artists through concerts, the platform puts less relevance on events that occurred prior to their inception

13

Songkick was founded in 2007 and there is a noticeable increase in the number of artists that have their earliest concerts recorded on Songkick in 2007 or later

13

Finally, we wanted to make sure that we have enough data about successful artists in the early stage of their career.

13

Figure 1 depicts distributions of the number of concerts per artist and number of concerts per venue

13

Both distributions are very broad and heavy tailed, with few active artists and venues hosting many events, and a very large set of artists and venues associated with very few events.

13

In such an affiliation matrix an artist is represented as a bag-of-words vector over the venues where the artist has performed. The entries in the matrix are the numbers of times the artists performed at the venue. We used those vectors as features for the prediction and forecasting tasks.

15

The dataset can be used for a variety of tasks which we exemplified by performing success forecasting and event prediction.

18

Similarly, we observed that by utilizing the underlying structure of this data, one can also predict whether an artist will have a concert in a particular venue. The performance of simple baseline models that we carried out in all three tasks indicates that much better results can be achieved with more carefully designed methods.

18

we illustrated how artist or venue influence can be measured based on analyzing a time-varying bipartite artist-venue graph.

18

For instance, the results presented here were averaged across different genres. It is plausible, however, that analysis will yield (subtle) differences when conditioned on the genre

18

we note that despite its demonstrated usefulness, the dataset presented here is not perfect and is likely to have some intrinsic biases,

18

# vanbuskirk\_songkick\_2009

they have painstakingly assembled concert information from thousands of sources to create a 1- million-song database of shows that's sort of like "Wikipedia meets Twitter" for live music.

CEO Ian Hogarth. "What we've done is crawl and cull this information from thousands of sites all across the web, and stitched it together enough to make a giant database of live musical history."

To get the ball rolling on this partially crowdsourced resource, Songkick staff added more than a million shows to the database, with new show information coming from agreements with 29 ticketing agencies

Now for the money question \x84 as in, how does Songkick intend to make any? Hogarth told us Songkick receives fees from ticketing agencies when someone buys a ticket through the site

see when your favorite bands toured most actively, using a timeline interface that represents live gigs with shaded areas

# cho\_live\_2012

We examine how the concert distribution has evolved and how broadband penetration is associated with concert attraction of particular regions

4494

we find that concert locations have been dispersed over time, but it varies across the level of popularity.

4494

The traditional profit-generating structure in many industries has been altered due to the advancement in information technology

4494

Music artists typically generate income in three ways: royalty from recorded music sales, music publishing fees for songwriters, and revenue sharing from live performances.

4494

the sharp drop in recorded music sales came hand in hand with the rise of the live performance market as a new source of income for music artists

4494

the concert industry is seen as a monopolistic competitive market that controls the process of creation of songs, ticket sales, and live tours.

4494

concert is a non-digitized experience good whose value can be fully recognized only when it is consumed.

4494

While recorded music can be substituted for digital format, the experience in the concert venue cannot be replaced by watching video clips of previously performed concerts

4494

the decisions of planning and executing concerts have

4494

become important for bands and concert promoters

4495

the concert industry is a relatively unexplored research field despite its rising importance

4495

The existing studies lack a precise estimation of the effect of the Internet on the music industry, because they only consider recorded music sales in their studies rather than having a bigger picture of the entire music market

4495

which type of artists has been more influenced between Superstars (famous and popular artists) and Underdogs

4495

to what degree does the Internet affect concert distribution

4495

This paper first examines the dynamics of concerts by focusing on the geographical distribution of concert locations, and then estimates the Internet's effect on concert distribution.

4495

For the study, the concert history data collected from two unique sources is matched with US Core-Based Statistical Area4 (CBSA).

4495

While popular and well-known bands choose to perform in small areas rather than concentrating more on big markets over the period, small bands have chosen big cities for their concerts.

4495

The empirical result indicates that higher broadband adoption has created more concerts after controlling other factors over the period

4495

the results show that the Internet affects the Underdogs' concert attraction more significantly.

4495

researchers conducted empirical works to examine the relationship between recorded music sales and concert revenues in greater details

4495

This paper is the first study to link the Internet to concerts in the US since Internet adoption has prevailed in the 2000s.

4495

The data on concert history come from Songkick.com6 and Pollstar magazine7. These two sources provide exhaustive history data of all concerts in the period from 2001 to 2010 within the US. The data include information on the main performer (or headliner8), location (city and state), and date.

4496

The concert data are mainly based on Songkick, and they were crosschecked with data from Pollstar Boxoffice.

4496

A drawback of the data is the difficulty in recognizing actual concert sales and attendance of each concert

4496

concert location distribution follows the Pareto distribution that states that a topranked 20% of CBSAs covers about 80% of concerts.

4497

The decision to select a concert location is an activity in which profit maximization and social influences play a key role

4501

The location choice is the outcome of a decision process that combines prediction of potential audience with their direct or indirect experience

4501

First, more small towns have attracted concerts over the period; simultaneously, there an incremental number of concerts are performed in big markets

4501

Internet adoption seems to play a pivotal role in delivering valuable and abundant up-to-date information about concerts to potential audience

4501

Our study is not without limitations. Firstly, we only consider the supply-side of the concert industry due to the limited data availability in terms of actual sales and the number of audience in each concert

4501

Some impressive dynamics might have occurred in the demand side, such as the impact of the increasing ticket price

4501

# gagen\_hybrids\_2019

Musical genre is multidimensional; sometimes defined by musical form or instrumentation, genre can also be derived historically, geographically or sociologically

39

Categorisation does appear to be a fun- 39

39

damental component of human cognition

40

"we cannot escape from our typological, or taxonomic, inclinations, because we have a need to orient ourselves in a world which inundates our mind in a puzzling manner with a huge amount of... impressions which can only be grasped by subsuming them under abstract, and sometimes very artificial, categories

40

genre therefore "suffers from an intrinsic ambiguity, deeply rooted in our dualist[ic] view of the world."

40

Genre categories "are convenient labels which allow critics, journalists and academics to organise their surveys, reflections and reviews

44

"draw together artists who, whatever their superficial stylistic similarities, have no concrete connection: no social ties or interdependence.

44

warns against genre classifications which lack "sociological reality

44

in an era of data curated by anybody, it is also possible for an artist to be tagged with a particular genre yet not be aware that the genre exists

44

The cultural and sociological aspects of genre are clearly important factors in understanding the mechanisms by which it operates

44

The relationship between genres and demographic categories is fundamental to commercial music marketing and "continued a process of organizing music in terms of categories of difference associated with demographic divisions

46

In popular music, the primary concern of is commerce.

46

The music industry uses demographic assumptions to mobilise genre as a means to market to different audiences, as opposed to musicians who may use genre as a means of communication when engaging with other musicians

46

genre "to mediate between producers and consumers." Genre is, therefore, not only multidimensional, but is multi-purpose.

46

"[p]opular music is a powerful cultural and economic force in modern capitalist societies."

46

genres are used by record companies as a way of integrating a conception of music (what does it sound like?) with a notion of the market (who will buy it?). Musician and audience are considered simultaneously, as a way of "defining music in its market" and "the market in its music."

47

"[t]hroughout the last half of the twentieth century, the music industry saw a proliferation of music genres, as markets became more fragmented"

47

Billboard used only three popular music genre categories in 1961, five by 1974, nine in 1982 and, by 1991, were listing 13 musical genre categories. By 2010 the number of genre categories had risen to 27

47

interventions by the music industry as it participated in the fragmentation of audiences.

47

the proliferation of genres that have occurred since the advent of the Internet

47

Genre classification is considered fundamental within theMusic Information Retrieval (MIR) community because, as well as being a difficult task, music classification and recommendation have direct commercial applications

48

Genre classification by analy- 48

48

around 40 music analysts (some working in Japan, Russia and other countries), nominate genres

48

They then make their case by citing important bands and media mentions, and a small group of editors makes the final decision

48

Data curation is a factor in the way that such systems operate. The Echo Nest, the music intelligence service owned by Spotify, allows staff and clients to "seed" new genres into the system

48

"[t]he three main ways more genres get added are: some human identifies a missing term or an unlabelled cluster

48

We"re always searching and ranking songs and artists for discovery purposes, and sometimes an emerging artist is the tip of an emerging genre. Our automated genre-miner surfaces a whole cluster of data-related music for which we don"t yet have a genre label... the code finds potential clusters, but a human... evaluates them to see if they make subjective listening sense

48

human editors looking at this data and making "subjective" decisions - not cultural gatekeepers, but cultural creators, acting from a corporate perspective.

48

sis of the audio signal has been attempted on numerous occasions "using signal processing and machine learning schemes.

49

Inconsistency among existing genre taxonomies is another part of the classification problem

49

"[c]lassifications often oscillate between... different interpretations... This semantic confusion leads to many redundancies in the taxonomy, and it is obviously a poor description scheme for automatic systems.

49

"this confusion, however, has apparently no impact on the efficiency of the taxonomy for human users. It is indeed easy to navigate in these taxonomies, and switching semantics at each taxonomic level is natural for most users"

49

The classification of music by genre is difficult due to the multidimensional nature of the problem

49

genre decisions made by people are, as often as not, based upon individual experience and cultural background

49

"genre definition and attribution is generally considered to be subjective... the establishment of any ground truth will be the study of responses to music."

49

the automatic classification of music into genre categories is a non-trivial problem

49

"[t]he ambiguity inherent to every definition of Musical Genre, together with the high dynamics that undermines its persistency over time, characterizes the complexity of the automatic genre categorization task

49

musical genres are as often subjective as not and will, therefore, defeat automation as often as not

49

claim that musical genre is essentially a cultural construct

50

We have considered genre as an historical object, as part of a system, as a social and cultural artefact, as a tool of commerce, and in the context of music information retrieval and data science

50

The system analyses a large corpora of pieces (nearly 40 million2) using digital signal processing (Jehan, 2005), data mining, and machine learning methods

72

Audio analysis and metadata gathered from multiple sources across the Internet (including user-edited sources, such as Wikipedia, Musicbrainz and Discogs) are combined and synthesised into musical "knowledge".

72

take any audio file as input, and generate various features, both low-level ("such as the time of when every beat starts") and high-level ("such as the overall "danceability\x8e").

72

The Echo Nest1 (a music intelligence service acquired by Spotify in 2014) acts, in this research, as an arbiter of genre

72

Studies evaluating these processing components are generally favourable when comparing them to other systems and methodologies, particularly in the realm of genre classification

72

"understand new music terms as quickly as they are uttered."

72

Genres can, it appears, appear and disappear from The Echo Nest. The addition of genres appeared to slow markedly throughout 2015.

73

bringing the total to more than double the number included in our main analyses.

75

MusicBrainz7 is a collectively constructed music metadata system or, as their homepage states, "an open music encyclopedia that collects music metadata and makes it available to the public.

76

it now contains information about millions of artists, tracks, recordings, producers, and performances

76

Wikidata is operated by the Wikimedia Foundation, and is a "free and open knowledge base that can be read and edited by both humans and machines

88

structured data of its Wikimedia sister projects

88

The ontology is based upon the structure of Wikipedia articles, the core concept being the "item".

88

An item has a label, a description, and any number of aliases, with statements (consisting of a property and a value) which characterise the item

88

Examination of the information held within Wikidata was initiated largely because the system is distinctly different to the commercially oriented Echo Nest

88

it was deemed important to also look directly at the data contained within the largest online, global knowledge-base

88

Wikidata, like our other sources, has inherent biases

88

Using the Wikidata Query Service, we search for instances of "music genre" items, and any relevant links to other items in the system

88

he most significant of these properties by far was "subclassOf".

89

To facilitate comparison (as shown in Section 8.3.1), manual aligning of the data is required.

89

The Wikidata information is of a different level of resolution to that from our other sources.

89

We learn only about the classification of genre, super- and sub-genre using this method; it serves only as a basis for comparison

89

We are also lacking genre-date and country information.

89

What is interesting, however, is the total number of genres. At 1262, the Wikidata genre-total is quite close to the unprocessed Musicbrainz (1294) and EchoNest (1379) totals

89

The Wikidata network is different to those generated from other data

170

There are a comparable number of genres but a fraction of the connections between them, implying that editors are more concerned with the items themselves than the relationships between them

170

The Echo Nest 170

170

minimal network, for example, has nearly 23 times as many edges as Wikidata

171

# jemielniak\_common\_2014

However, it may lead to systemic biases, based on a majority of culture, language, or gender, and consequently deter people who are underrepresented in the community from expressing their opinion

95

the consensus and the approach to the problem, relying on the Western English custom and ignoring other cultures" sensitivities, is quite likely related to the history of Wikipedia, which developed at first more dynamically in the United States and Europe

95

Issues of low symbolic importance to the dominant population of Wikimedians are significant to those who fall victim to the systemic bias

95

editors with more mainstream views tend to win argument

95

Young, male, Western, well-educated, and relatively affluent editors have more time, technical skills, commitment, and confidence to participate in Wikipedia

95

The homogeneous composition of editors affects the articles12 and, quite obviously, undermines diversity even more

95

# schindler\_capturing\_2014

Music genre classification is one of the most prominent tasks in the domain of Music Information Retrieval

229

This paper provides a performance evaluation of the Echonest audio descriptors.

230

In this paper, we presented a comparison of Echonest features - as provided by the Million Song Dataset - with feature sets from conventionally available feature extractors.

240

Experiments show that Temporal Echonest Features - a combination of MFCC and Chroma features combined with loudness information as well as the distribution of segment lengths - complimented by all calculated statistical moments - outperforms almost all datasets and classifiers - even conventional feature sets, with a prediction rate of up to 89 %.

240

# prey\_knowing\_2018

data never merely reflects reality; it always constructs that which it measures at the same time

21

Spotify requires such data: [\x83] to provide, personalise, and improve your experience with the Service and products, services, and advertising (including for third party products and services) made available on or outside the Service

22

like other online platforms, Spotify is increasingly focusing on context in an attempt to serve better recommendations.

25

Many studies have demonstrated that listeners gravitate to music that matches their current context

25

In order to recommend music that matches these contexts, streaming platforms need to collect and aggregate data points on everything from a listener"s location

25

What has been called "the contextual turn" (Pagano et al. 2016, p. 1) in recommender systems can be described as a move away from the "Immutable Preference paradigm" (ImP). ImP assumed that the user was a fixed individual, whose "goals, needs, and tastes do not develop" and in turn, "that the set of items to be recommended remains relatively static"

26

personalization in recommender systems involves recommendation for specific individuals." Instead, a context-based recommender system, "personalizes to users" context states" (ibid.) rather than to individual users.

26

"people have more in common with other people in the same situation, or with the same goals, than they do with past versions of themselves"

26

Spotify argued that context is the new genre

26

mood categorization represents a much more natural way of thinking about music; one that is more representative of how people speak about music

27

people speak very naturally about music

27

The goal, as he describes it, is to create a music player that knows precisely what music to play for any listener given their current context. The challenge, however, is to glean enough information from the listener, without them even needing to actively tell the service what they want to hear

27

future of listener understanding and segmentation will get deeper into how, when and where people actually interact with music

28

# Wang

Genres are used to organize music, but themselves reflect complex historical processes

3

Their boundaries are fuzzy, and new genres and names appear and disappear from usage over time

3

They can for example be geographically or culturally defined, based on technical requirements, or marketing considerations.

3

Artists and tracks do not necessarily "belong to", "produce" or "use" one genre alone

3

This means that there is an inherent challenge in assigning an artist to a genre

3

While more fine-grained labels are available (see for example everynoise.com for a collection of 4,852 genre labels), and while artists can make tracks that fall in multiple genres,

3

this dataset assigns one main top-level genre to each artist

3

High-level genres include for example hip hop/rap, Latin, pop, metal and R&B

3

the most frequently occurring genres in descending order are pop, hip hop, dance/electronic, rock, indie, Latin.

3

Excluded from analysis were genres not well-represented in this top 50,000, as well as a genres such as spoken-word, soundtrack, comedy and kids music.

3

We observe large differences in Wikipedia representation of artists of different genres

9

coverage of hip hop, Latin, and dance/electronic artists is particularly lacking, while rock artists appear to have the best coverage.

9

Latin artists" under-representation on English Wikipedia is not explainable by differences in language editions alone

9

nity editing work has not kept up with entertainment culture

9

hip hop is of great cultural importance and can empower underrepresented communities

9

while hip hop is one of the most popular music genres across consumers, it is often perceived as associated with "dispossessed" Black youth (Bennett 2008) while "good rock music\x8e is associated with white males by consumers (

9

very few artists in our dataset focus on metal or country, but those metal and country artists have highWikipedia coverage

9

our dataset is restricted to the most popular 50,000 artists, and there is a long tail of artists that we do not examine

9

With regard to genre labels, not all artists clearly fall within genre borders, nor are these borders set in stone

10

many artists creatively span multiple genres and subgenres

10

There are many genres and subgenres we do not examine.

10

Another limitation to consider is that the dataset is influenced by who stream from online streaming platforms

10

those who stream music versus listen via other formats such as CDs tend to be younger and have access to the internet

10

# Wardhana

Songkick provides information of music concerts around the world.

1

the systems associated with music recommendations mainly based on two approaches, collaborative filtering [5] and combination of clustering and community-based.

1

The above facts led us to implement FOAF and Music ontology on music concerts recommendation system, integrated with Facebook, Songkick, Eventbrite and DBPedia to facilitate the users in accessing music concerts information.

1