

# Social Media Technology

Group 1

Raphael, Lorenz, Juri, Michael

last.fm

GENIUS

# last.fm

- *“the social music revolution”*
- *“Last.fm is a music service that learns what you love”*

# last.fm as a website

- streaming service provider (until 2014)
- collecting user's listening behaviour
- *scrobbling* data
- *now* → links to other providers  
or enabled devices



# datasets

- last.fm dataset (2013-2014) <http://www.cp.jku.at/datasets/LFM-1b/>
  - LFM-1b Dataset *Markus Schedl*; “The LFM-1b Dataset for Music Retrieval and Recommendation”
  - ~ 43 GiB
  - 1.088.161.692 unique listening events
  - 60.675 artists
  - 120.322 users
  - 534.841 tracks
- last.fm userdata (2019)
  - last.fm api + crawling/scraping the user pages
  - 68.696 users
  - ~17.000 artists

# last.fm userdata (2019)

- API
- crawling/scraping
  - crawl/scrape only data of interest
  - SQLite database for organizing
- processing
  - reducing the dataset
    - users with missing values (e.g. playcounts)
    - duplicates
    - cleanup e.g. unreadable values

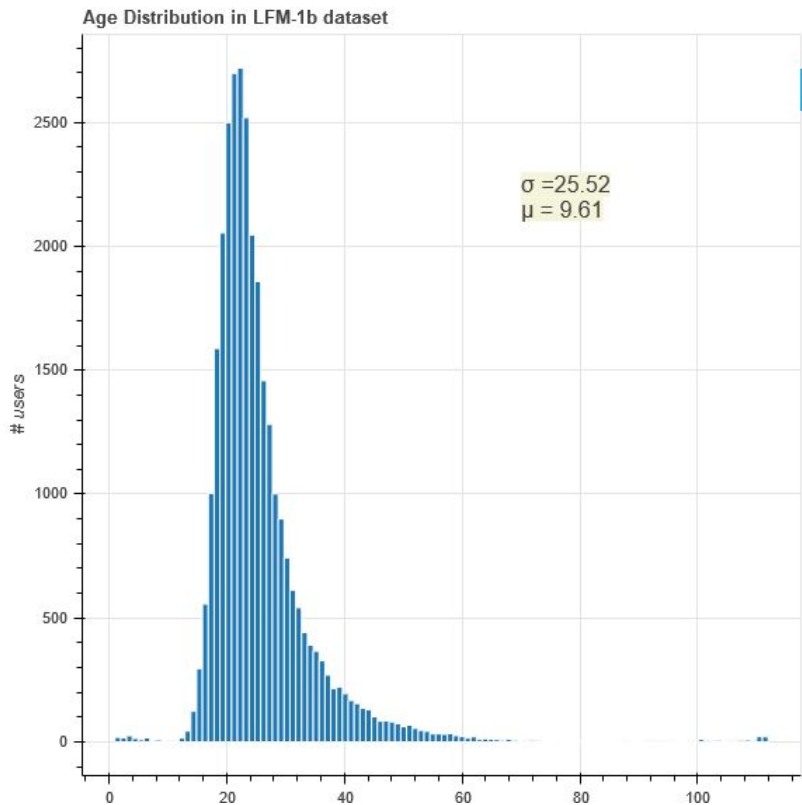
*UserData:*

username  
country  
registered  
playcount  
top\_artists  
loved\_tracks  
top\_tags  
avg\_scrobbles\_day  
week\_scrobbles  
year\_scrobbles  
year\_listening\_clock  
time\_listened  
tracks\_listened  
artists\_listened

# dataset - some key figures

key figure	2014	2019
number of users	120.322	68.696
top user <i>by scrobbles</i>	u_id: 4980702 1.000.150	KFCDELIVERY 19.438.313
scrobbles	1.954.279.102	4.700.880.842
avg. per user <i>respectively</i>	16.249	69.127
global playtime	NDA	3231 years

# Demographics lfm-1b (2014)

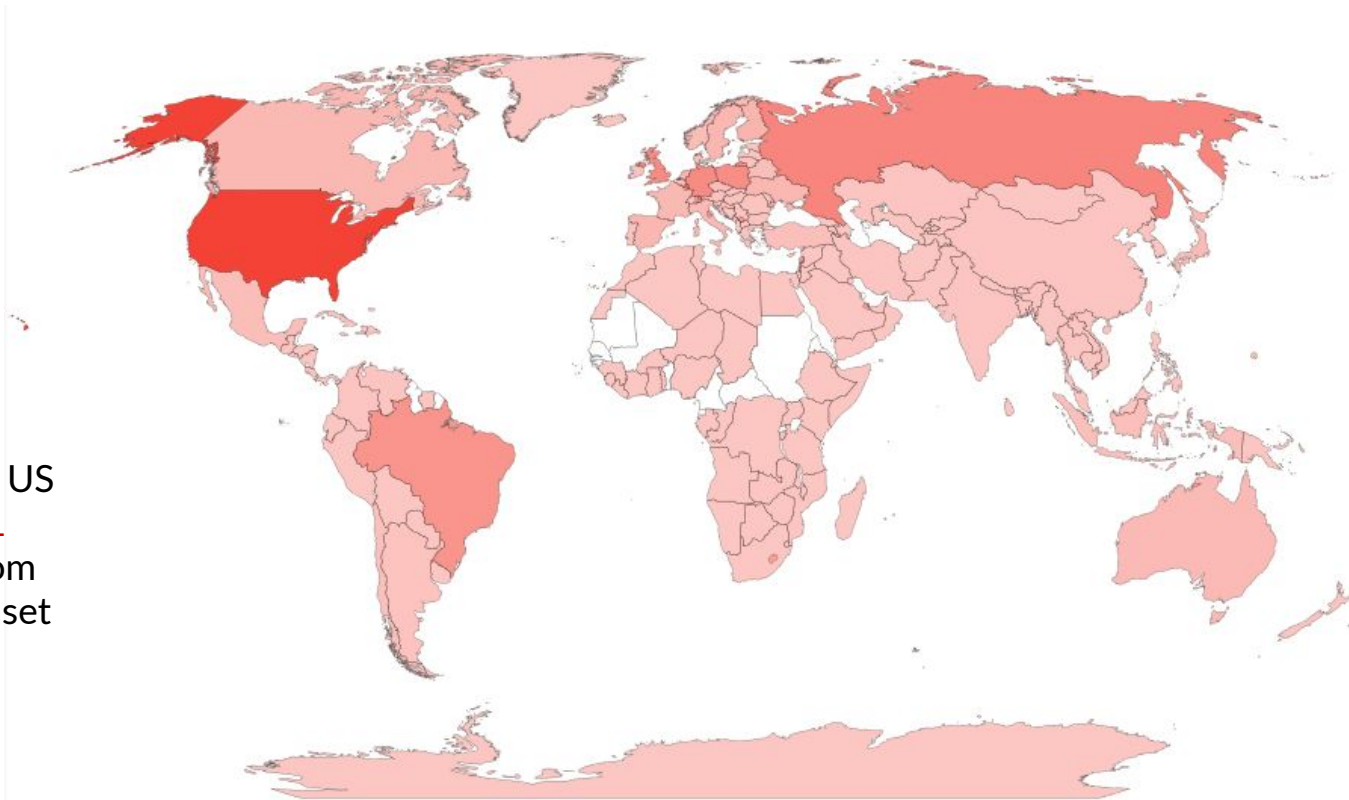


Gender	# users	percentage
male	39.830	71.63%
female	15.773	28.37%
NDA	64.307	53.63%



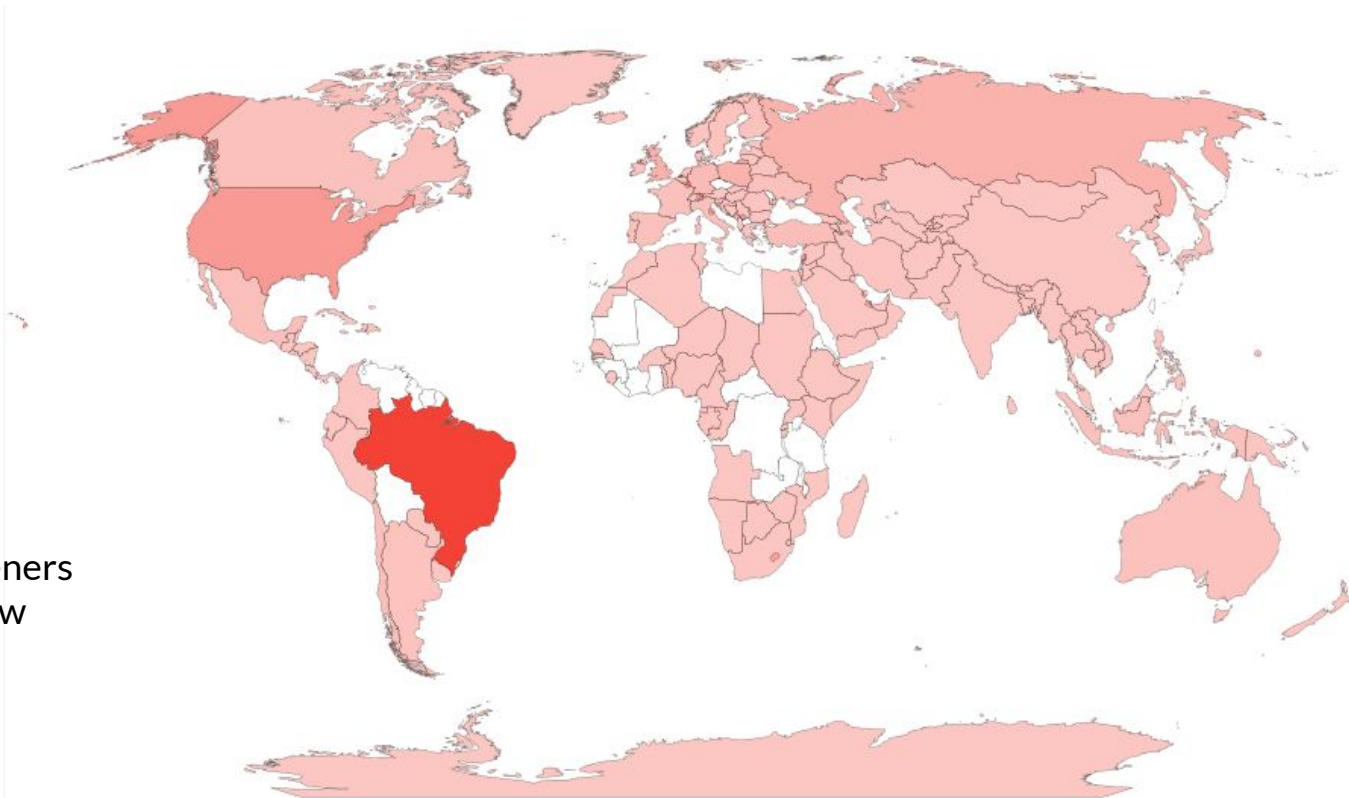
# 2014 (~52.000 listeners)

top country US  
with **10.251**  
listeners from  
the old dataset



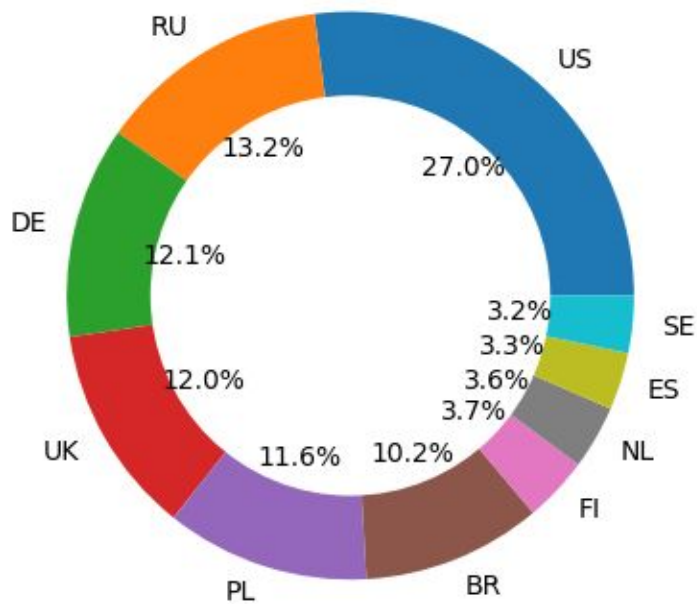
# 2019 (~52.000 listeners)

top country  
Brazil with  
**14.903** listeners  
from the new  
dataset

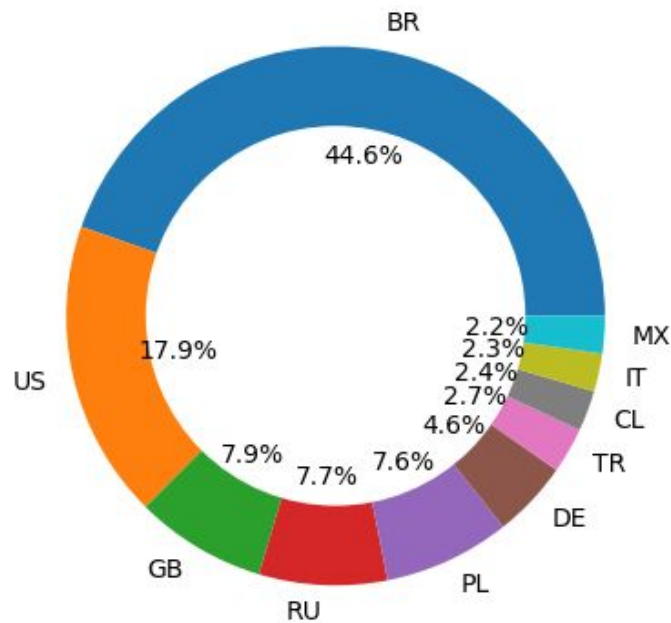


# Top Countries

data from 2014



data from 2019



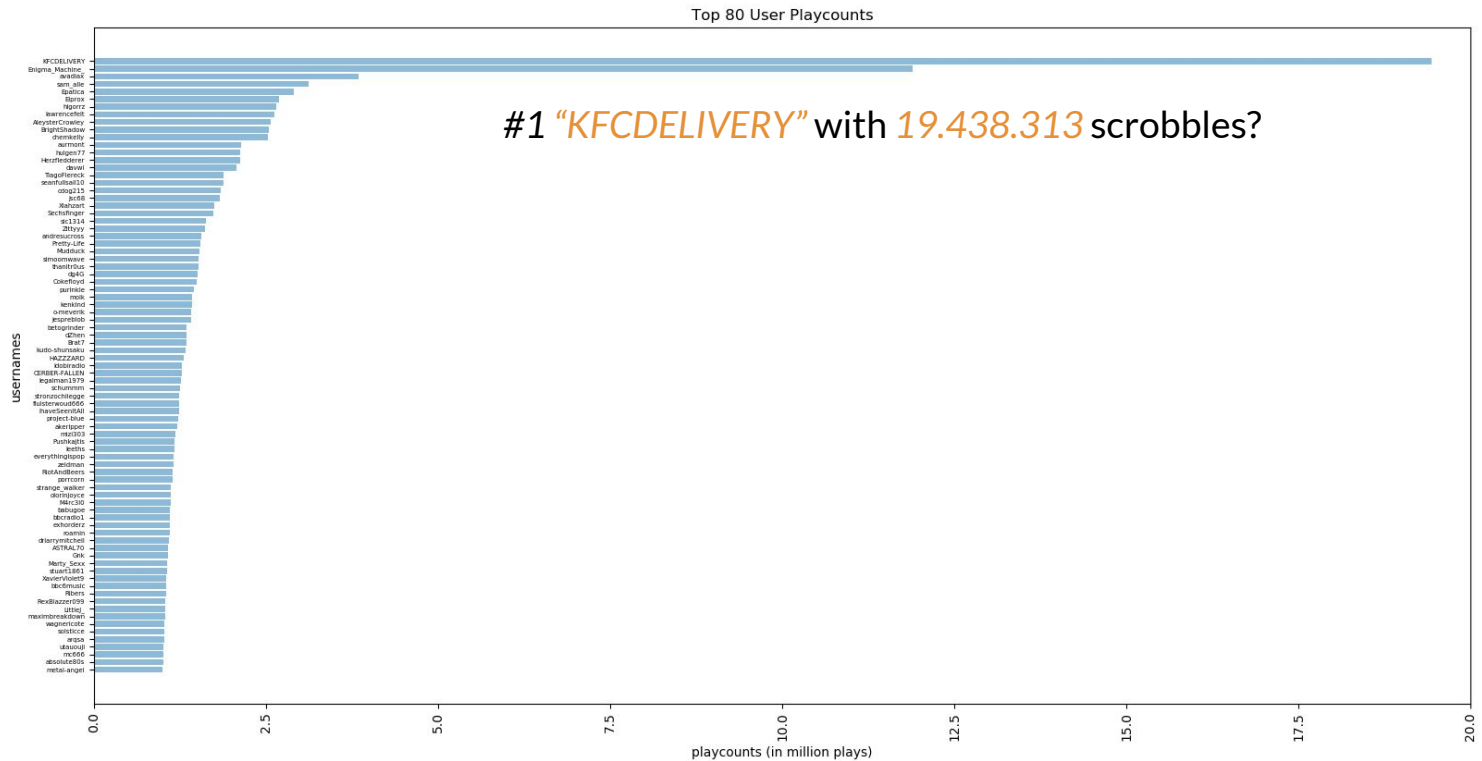
# Average scrobbles per day and country



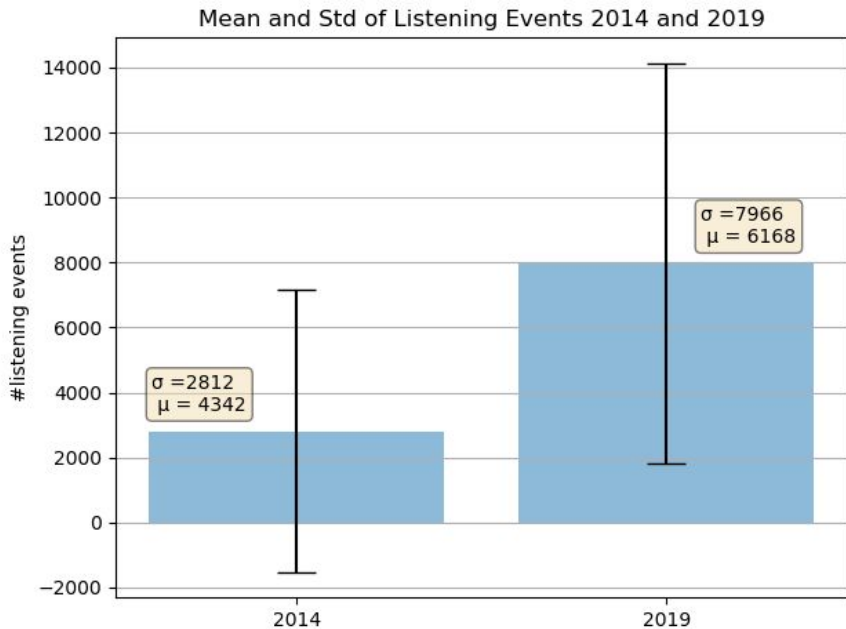
37.194 users with average scrobbles data per day in 2018

threshold: > 1000 users  
(with scrobble data)

# User ranking by scrobbles

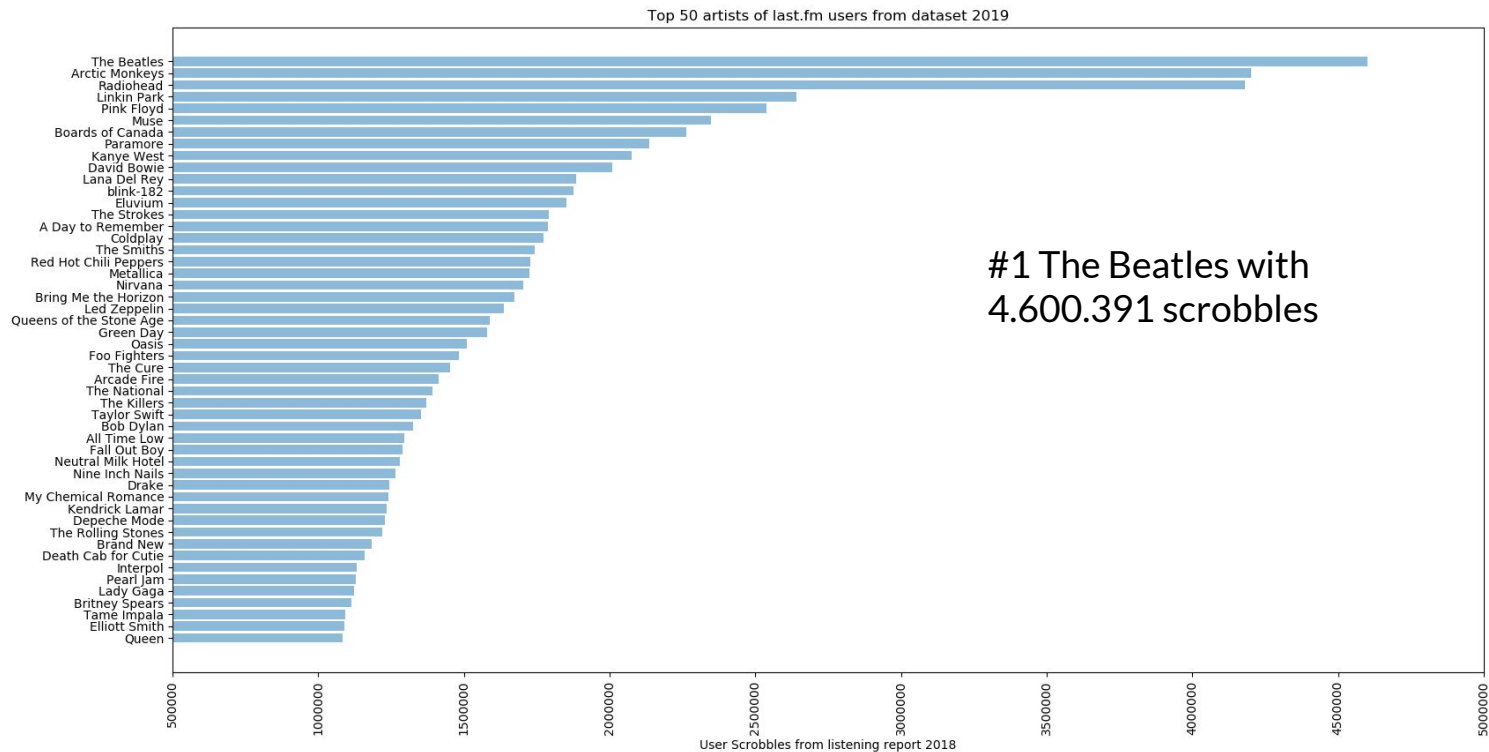


# Scrobble comparison



subset of users with listening event data  $x$  ( $0 < x \leq 20.000$ )

# Top Artists from Listening Reports



# Genre profiles and similarity of countries

- *“What genres are preferred in different countries?”*
- *“Which countries share similar genre preferences?”*
- programming language R
  - dplyr & tidyverse
  - jsonlite & httr
  - fmsb
  - corrplot
  - gridExtra & grid



# Used dataset

<b>artist</b>	<b>listening event</b>	<b>track</b>	<b>user</b>
<ul style="list-style-type: none"><li>- artist_id</li><li>- name</li></ul>	<ul style="list-style-type: none"><li>- user_id</li><li>- artist_id</li><li>- album_id</li><li>- track_id</li><li>- time</li></ul>	<ul style="list-style-type: none"><li>- track_id</li><li>- name</li><li>- artist_id</li></ul>	<ul style="list-style-type: none"><li>- user_id</li><li>- country</li><li>- age</li><li>- gender</li><li>- playcount</li><li>- time</li></ul>

# Used dataset

artist	listening event	track	user
<ul style="list-style-type: none"><li>- artist_id</li><li>- name</li></ul>	<ul style="list-style-type: none"><li>- user_id</li><li>- artist_id</li><li>- album_id</li><li>- track_id</li><li>- time</li></ul>	<ul style="list-style-type: none"><li>- track_id</li><li>- name</li><li>- artist_id</li></ul>	<ul style="list-style-type: none"><li>- user_id</li><li>- country</li><li>- age</li><li>- gender</li><li>- playcount</li><li>- time</li></ul>

# Data (pre-) processing LFM-1b

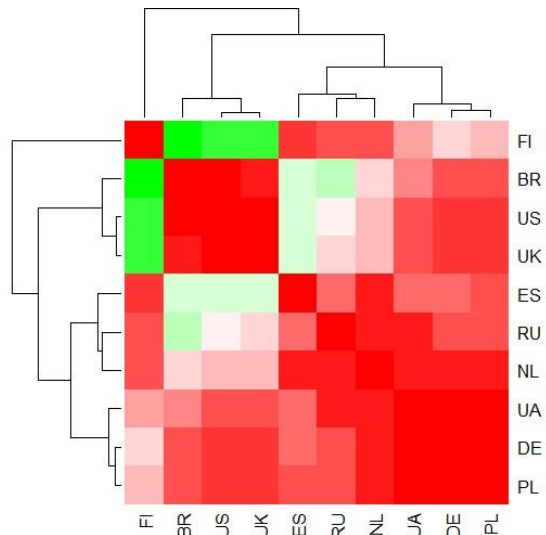
- 10.000.000 listening events (RStudio could not cope with more)
  - user\_id, artist\_id, track\_id
- 10 most frequent countries in dataset
  - BR, DE, ES, FI, NL, PL, RU, UA, UK, US
  - => reduce dataset (duplicates and entries of other countries)
- 6 most important music genres (<https://www.allmusic.com/genres>)
  - Rock, Alternative, Electronic, Pop, Folk, Metal
- API call for getting genre of track (artist + track)
  - <http://ws.audioscrobbler.com/2.0/> + track.getInfo
  - JSON-result
- takes about 200 hours

# Data visualization

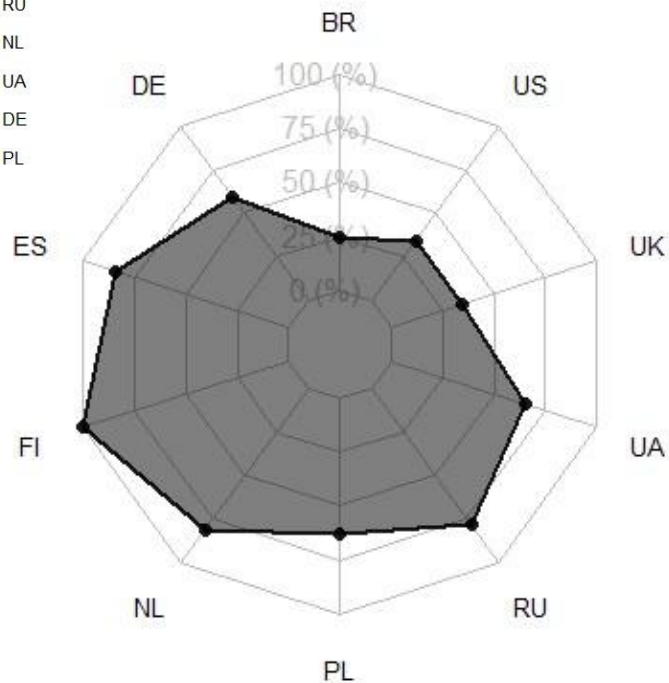
- radar charts
- table
- correlation matrix
- heat map

# Results

### Pop

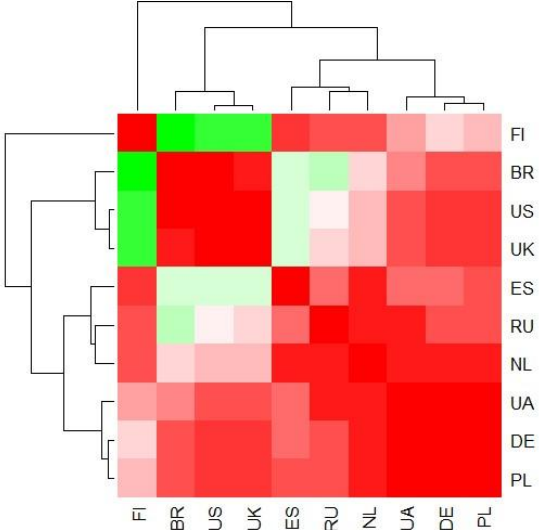


### Metal

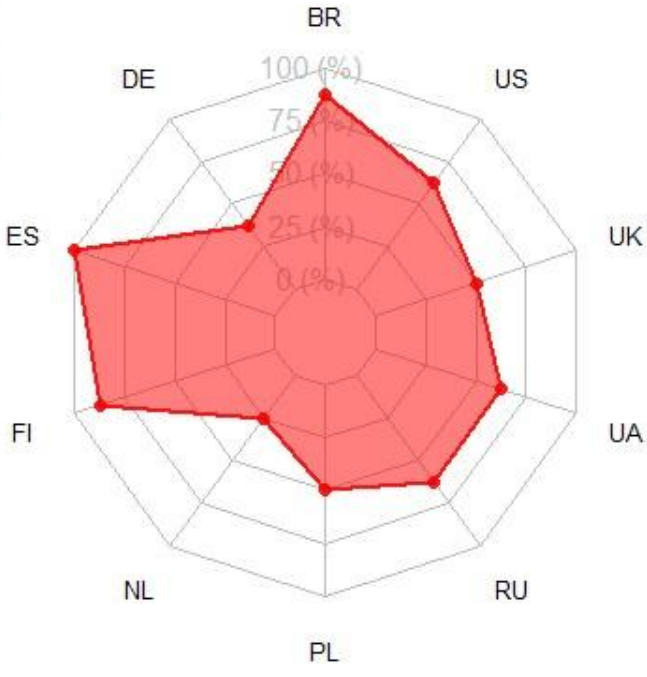


# Results

### Rock

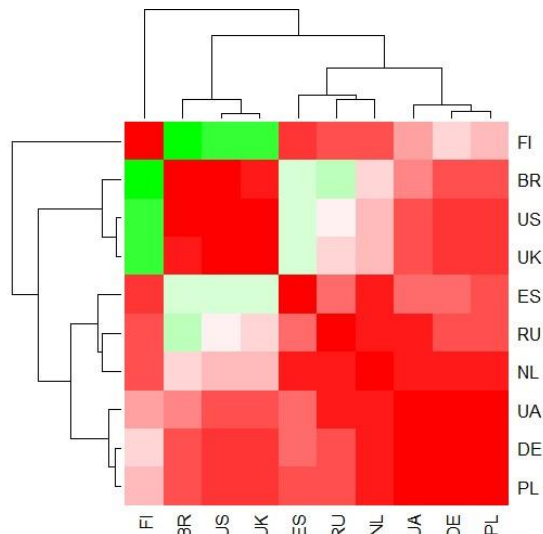


### Alternative

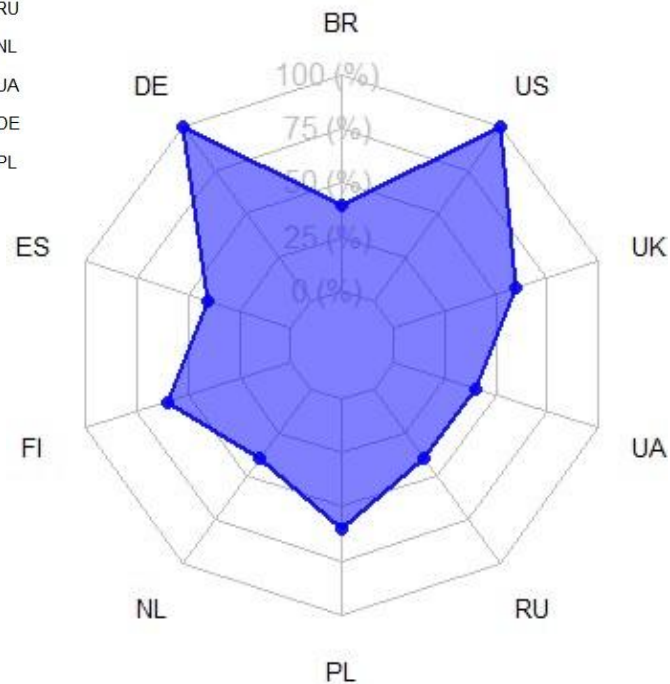


# Results

### Electronic

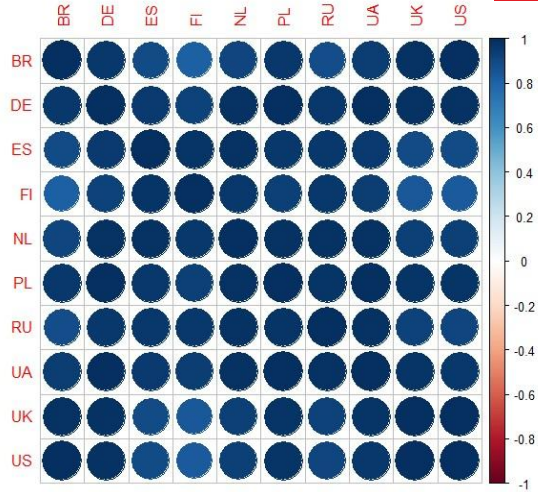
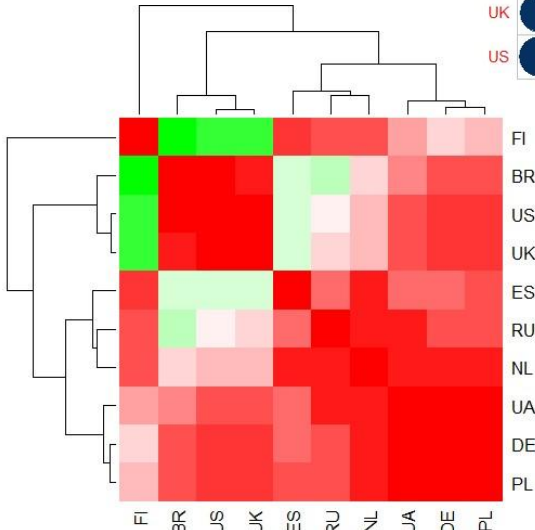


### Folk



# Results

	Rock	Alternative	Electronic	Pop	Folk	Metal
<i>BR</i>	0.66	0.07	0.1	0.07	0.02	0.08
<i>DE</i>	0.55	0.03	0.12	0.06	0.05	0.19
<i>ES</i>	0.5	0.08	0.06	0.08	0.02	0.27
<i>FI</i>	0.46	0.07	0.09	0.03	0.03	0.32
<i>NL</i>	0.55	0.02	0.09	0.06	0.02	0.26
<i>PL</i>	0.56	0.04	0.11	0.06	0.03	0.2
<i>RU</i>	0.45	0.05	0.16	0.06	0.02	0.25
<i>UA</i>	0.52	0.05	0.15	0.05	0.02	0.21
<i>UK</i>	0.61	0.04	0.16	0.04	0.03	0.11
<i>US</i>	0.6	0.05	0.14	0.06	0.05	0.11





# GENIUS

- Collection of song lyrics and musical knowledge
- Started as platform for annotating Rap lyrics
- User generated artist and lyrics data and annotations

# Genius dataset

- Crawled via website and API
- 8GB SQLite database
  
- Artists, Songs and relations between them
- 18.138 Artists
- 911.953 Songs
- 859 Tags

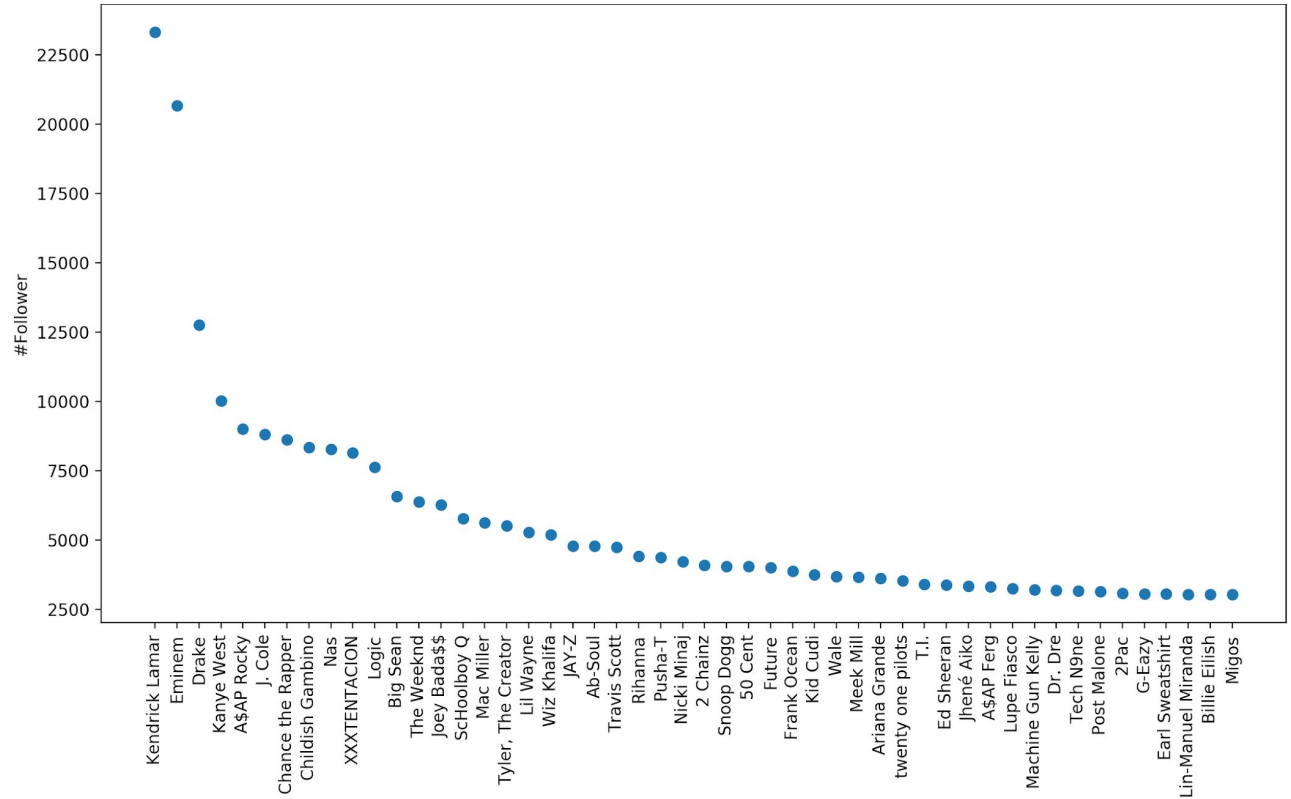
# 18.138 Artists

Most popular artists on  
genius

**#1** Kendrick Lamar  
23,314

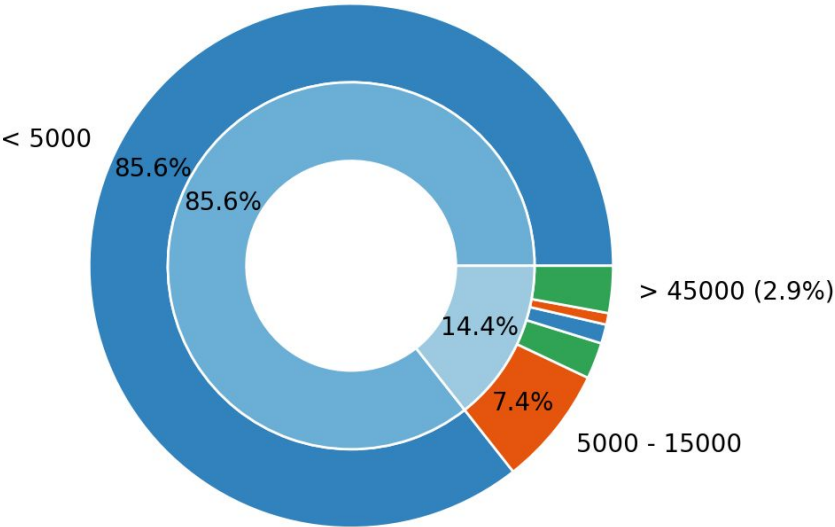
**#2** Eminem  
20,654

**#3** Drake  
12,744

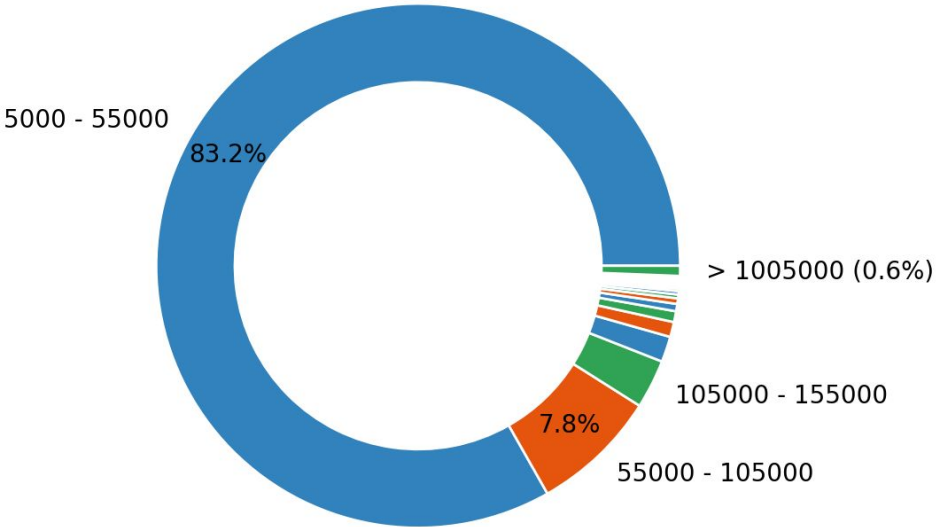


# 911.953 Songs

#Pageviews



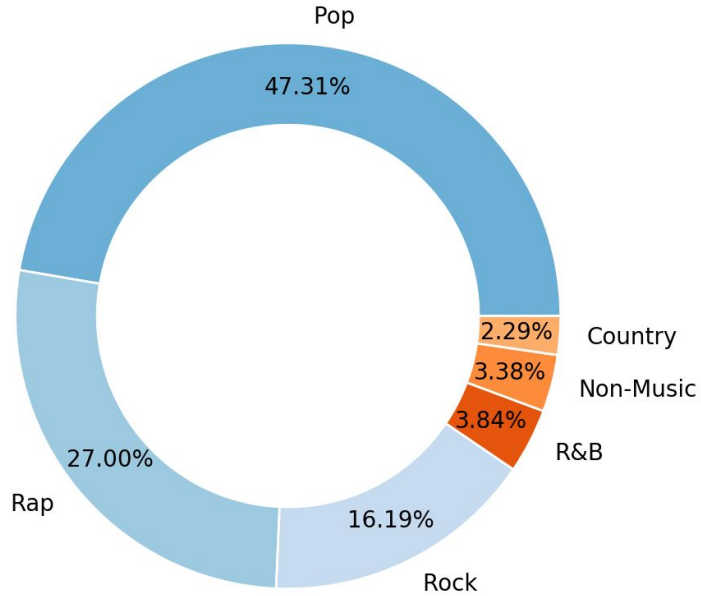
#Pageviews > 5000



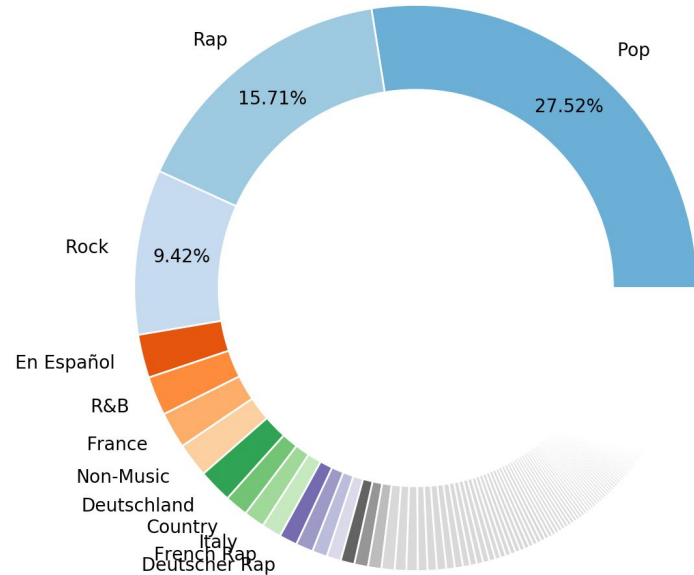
Songs with less than 5000 pageviews have set 0 as value for pageviews.

# 859 Tags

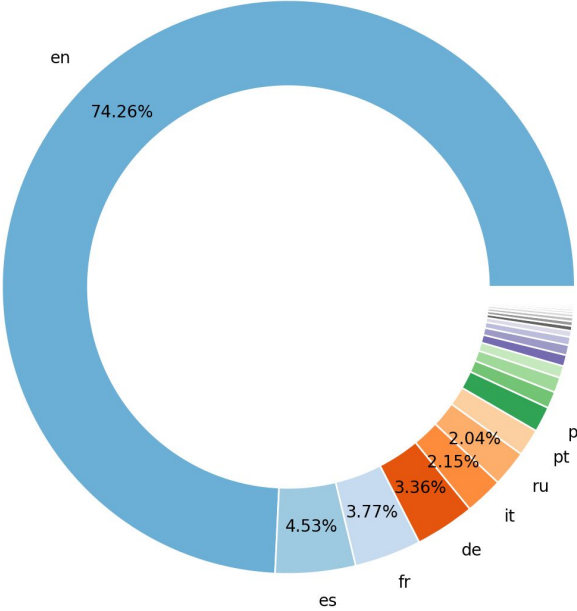
Songs per tag with only primary tag



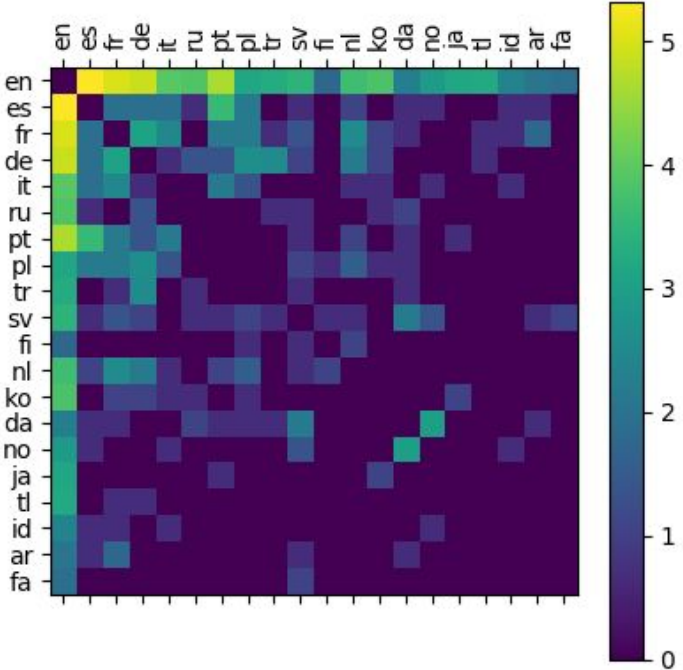
Songs per tag



# 132 Languages



Language distribution in dataset



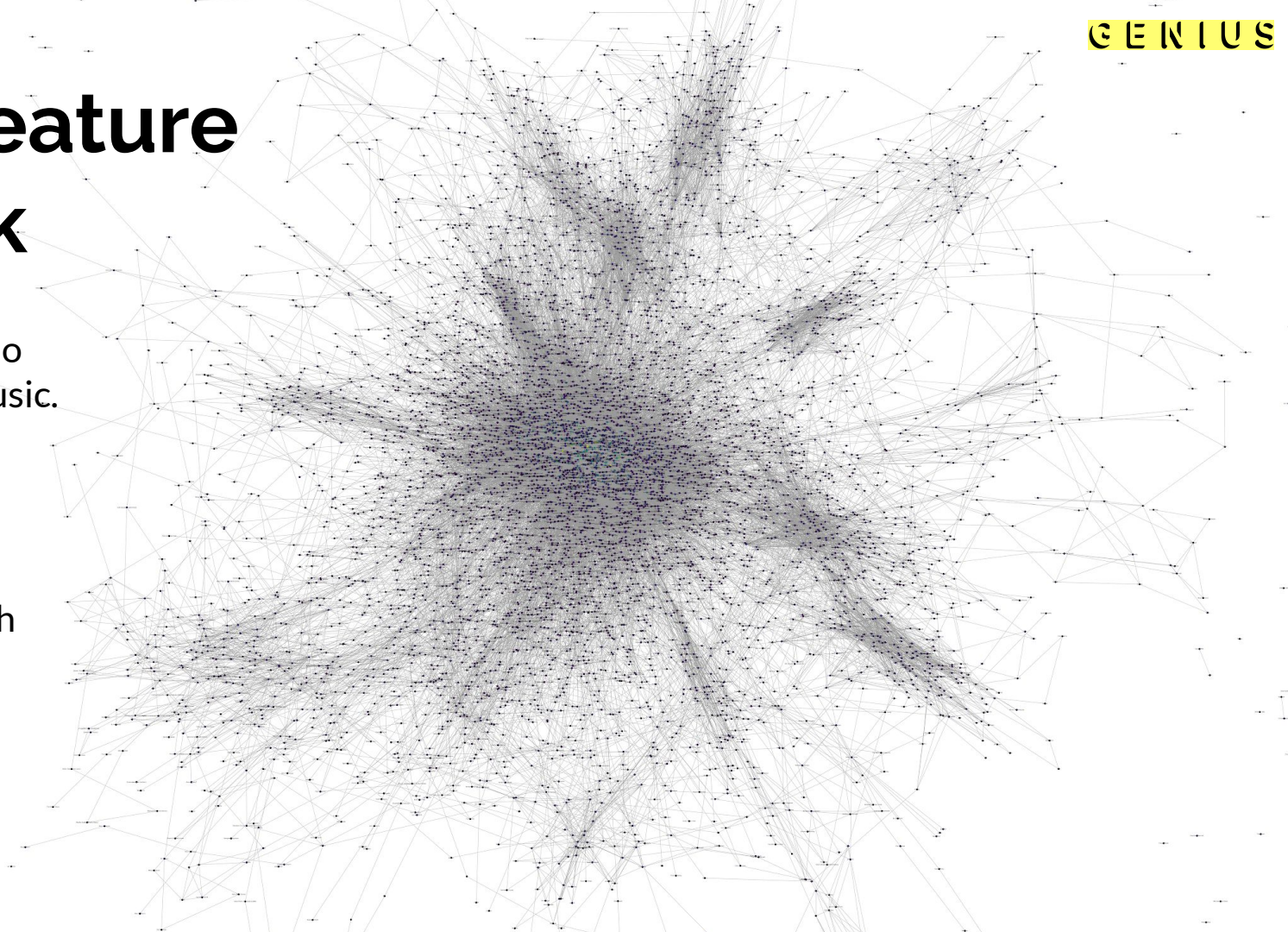
Collaborations across languages

# Artist Feature network

Connect artists who  
collaborated on music.

Average degree  
3.3

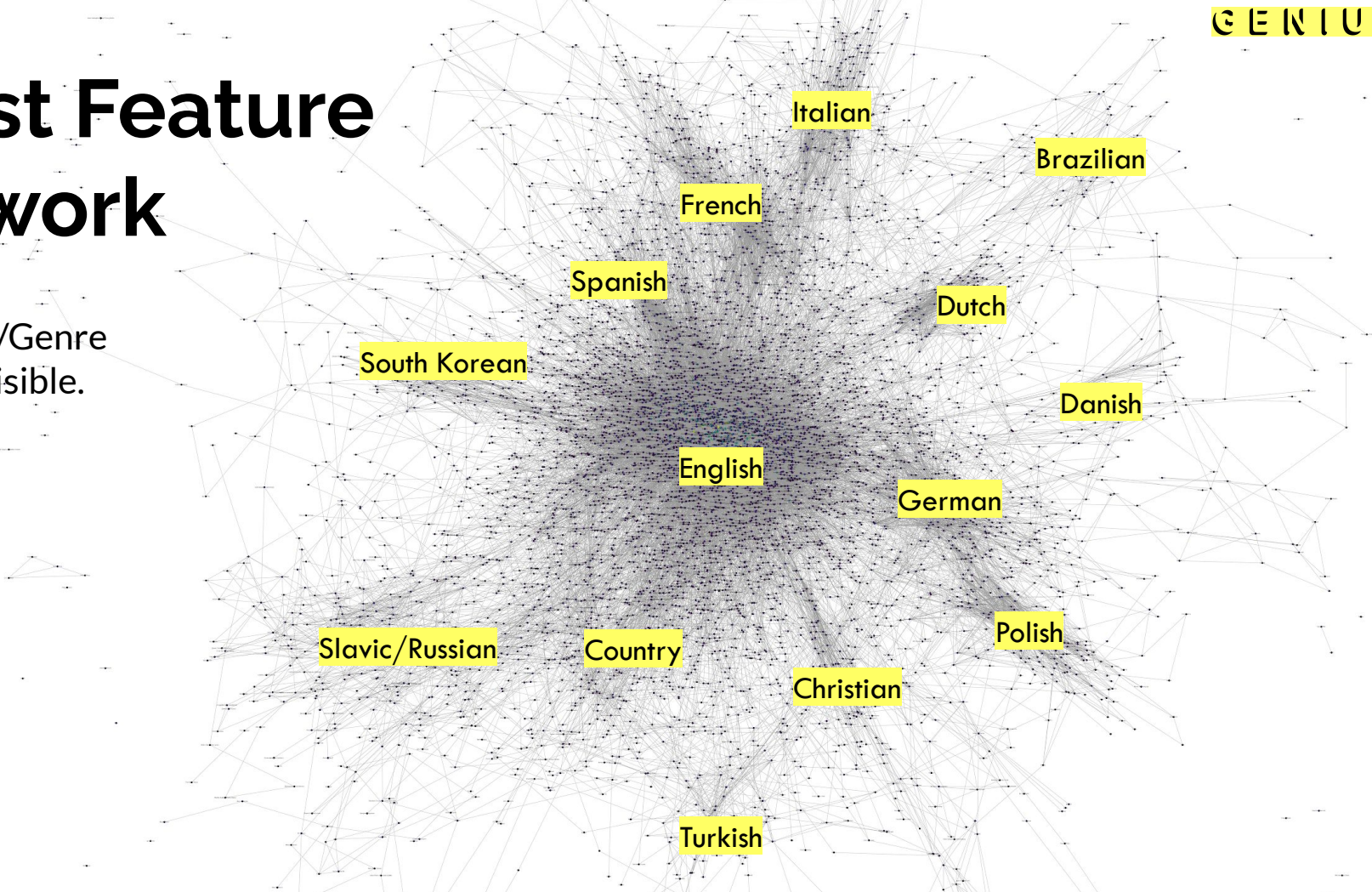
Average Pathlength  
8.2





# Artist Feature network

Language/Genre clusters visible.





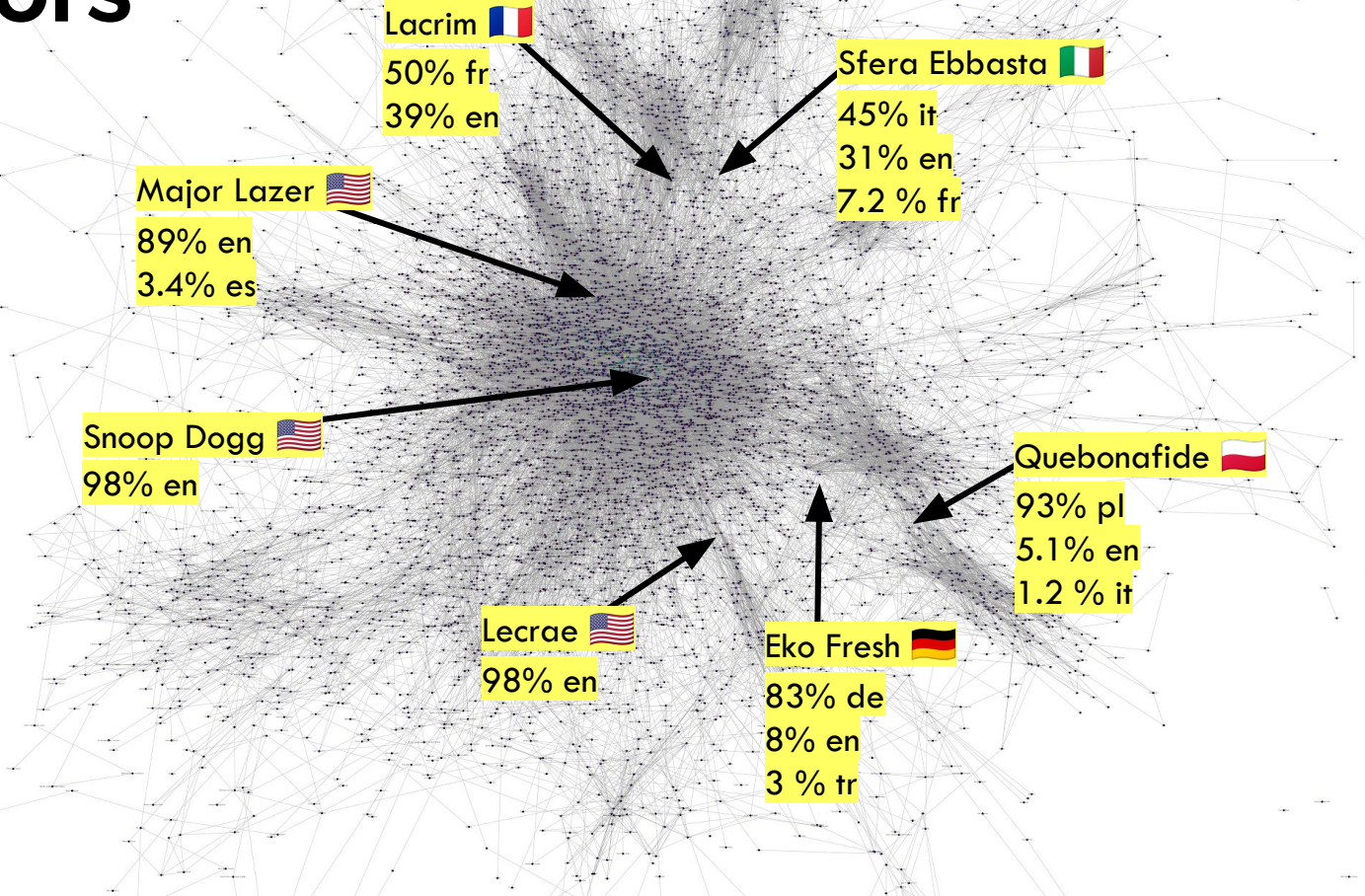
# Connectors

Betweenness score

Highest betweenness:  
Snoop Dogg

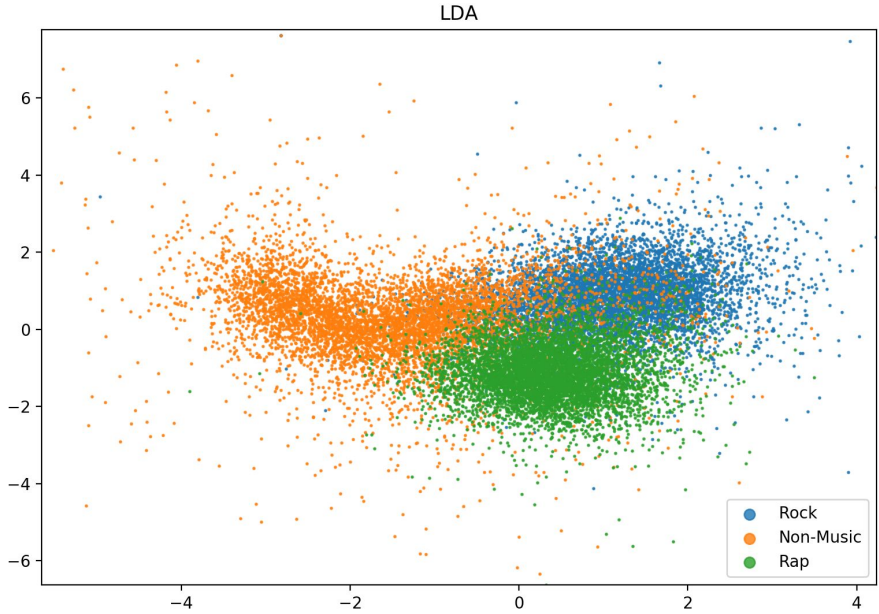
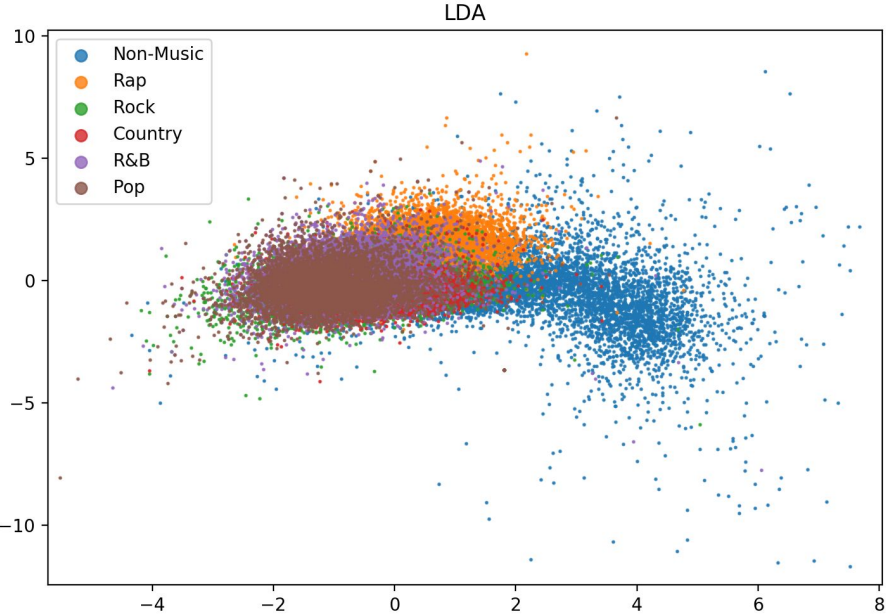
Connectors between  
languages  
Sfera Ebbasta, Lacrim

Connectors between  
genres  
Lecrae



# Genre Prediction

- LDA classification
- Based lyrical features
- Test accuracy: 61%



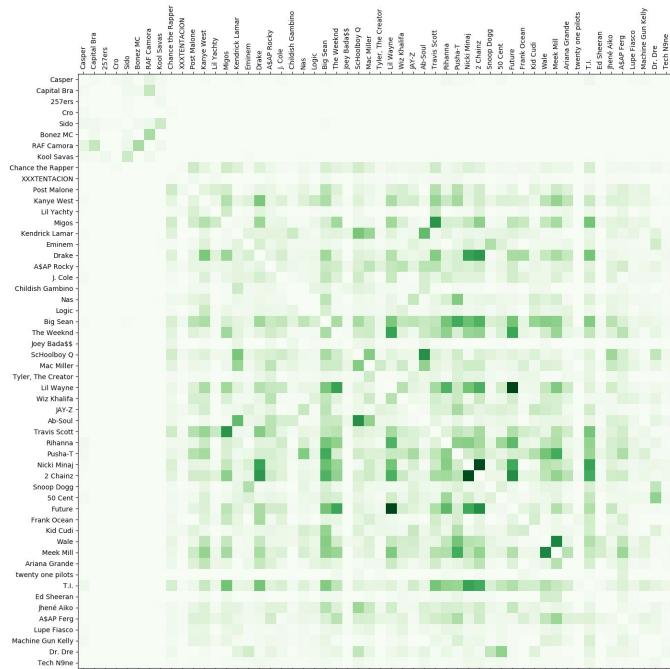
# Genre Prediction

- Neural Network
- Single hidden layer
- 100 neurons with ReLU activation function

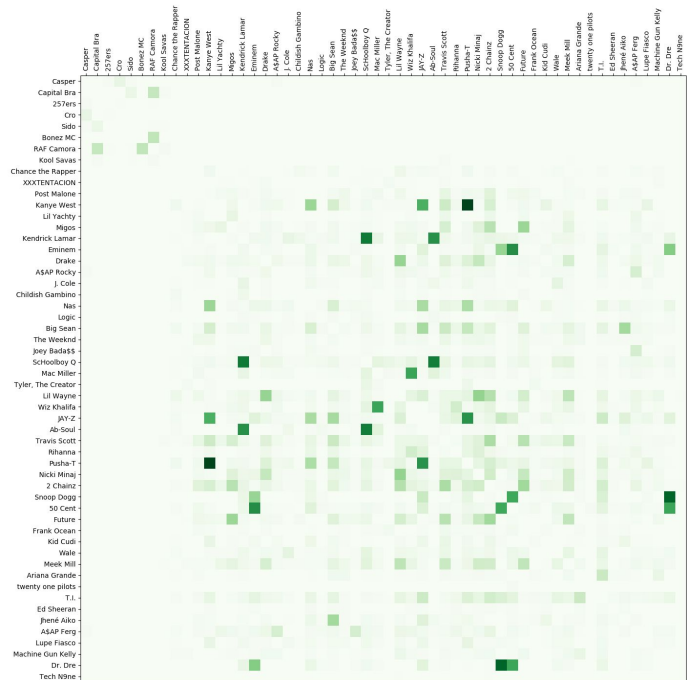
Feature Set	Test accuracy
Lyrical features	63.40%
Metadata and collaborations	57.71%
Combined	68.66%



# Feature-, Producer-similarity



Feature similarity



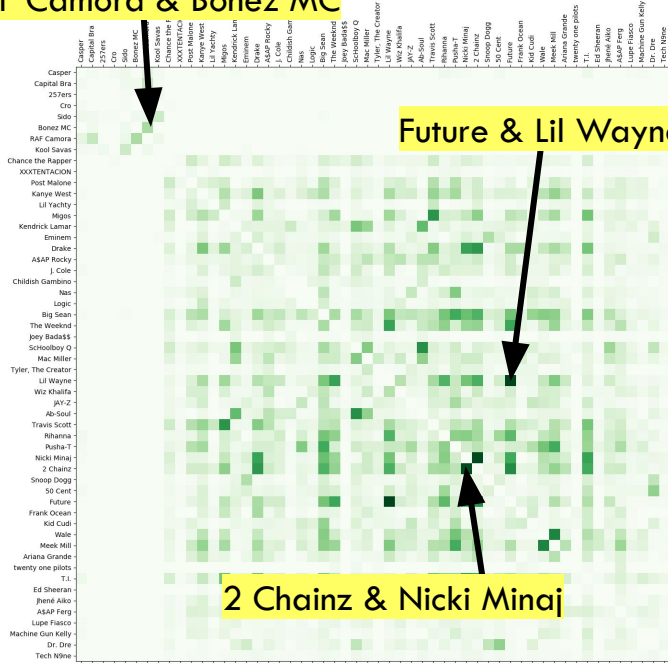
Producer similarity





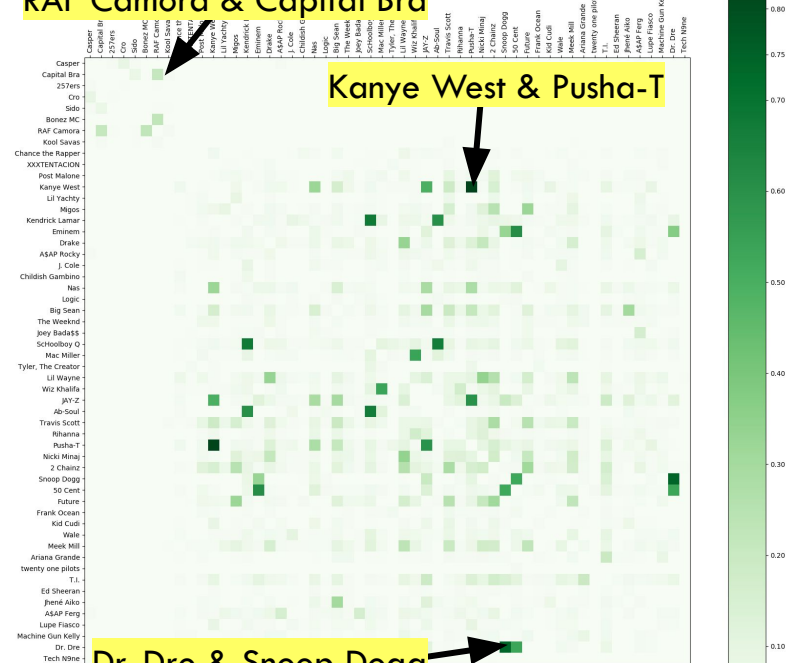
# Feature-, Producer-similarity

RAF Camora & Bonez MC



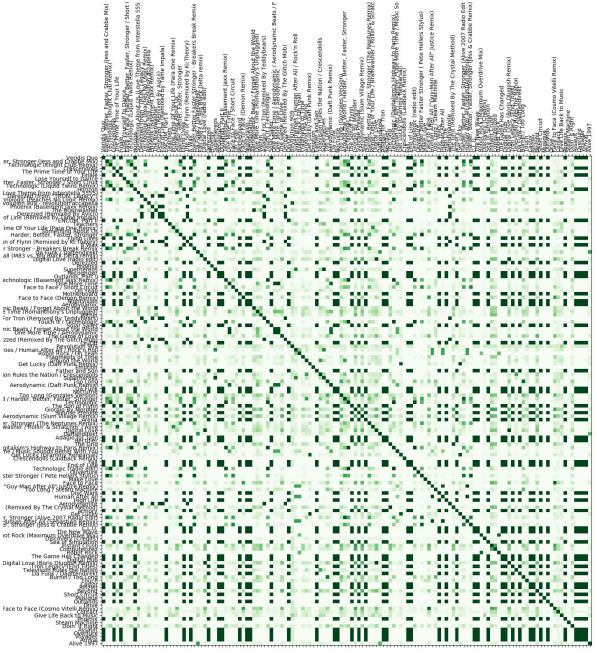
Feature similarity

RAF Camora & Capital Bra

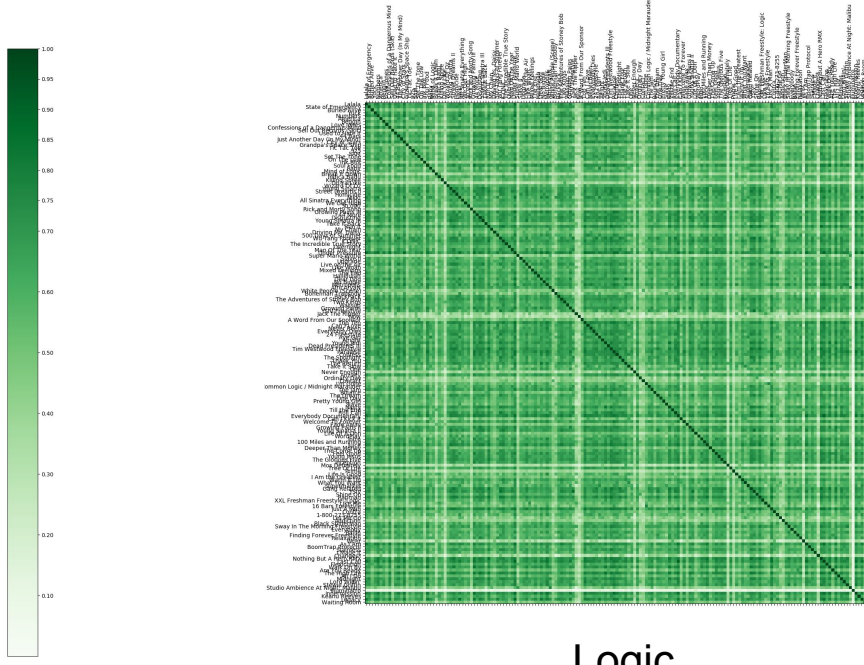


Producer similarity

# Song similarity per Artist



Daft Punk  
Electronic



Logic  
Rap

# Song similarity per Artist

Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world  
 Around the world, around the world

Daft Punk  
 Around the world

Ayy, bitch, I've been goin' and goin' like the Energizer  
 Yeah, I'm supplyin' the wood like Elijah  
 In the cut, smokin' on indica  
 Might fuck around and compartmentalize ya  
 They say, they say life is a bitch  
 And if that is the case then I'm finna surprise her  
 No, I am not an advisor, but I do advise ya  
 To pay attention, never need to mention  
 I'm that motherfucker bustin' heads  
 Finna push it to the ledge  
 Yeah, I been smokin' my meds  
 Ain't got no love for the feds  
 Can't let fame go to yo head  
 Fuck with me, watch where you tread  
 I'm finna kill it instead  
 I'm finna, I'm finna, I'm finna  
 I will, I will, do it like I never done it  
 And I wanna run it and I wanna keep it goin' and goin'  
 Like infinity, be the only entity to ever rip it apart  
 From the start like this, from the heart like this  
 Finna murder it, a million miles a minute, no nitrous  
 Doin' righteous

Logic  
 44 more



# Thank you

Any Questions?

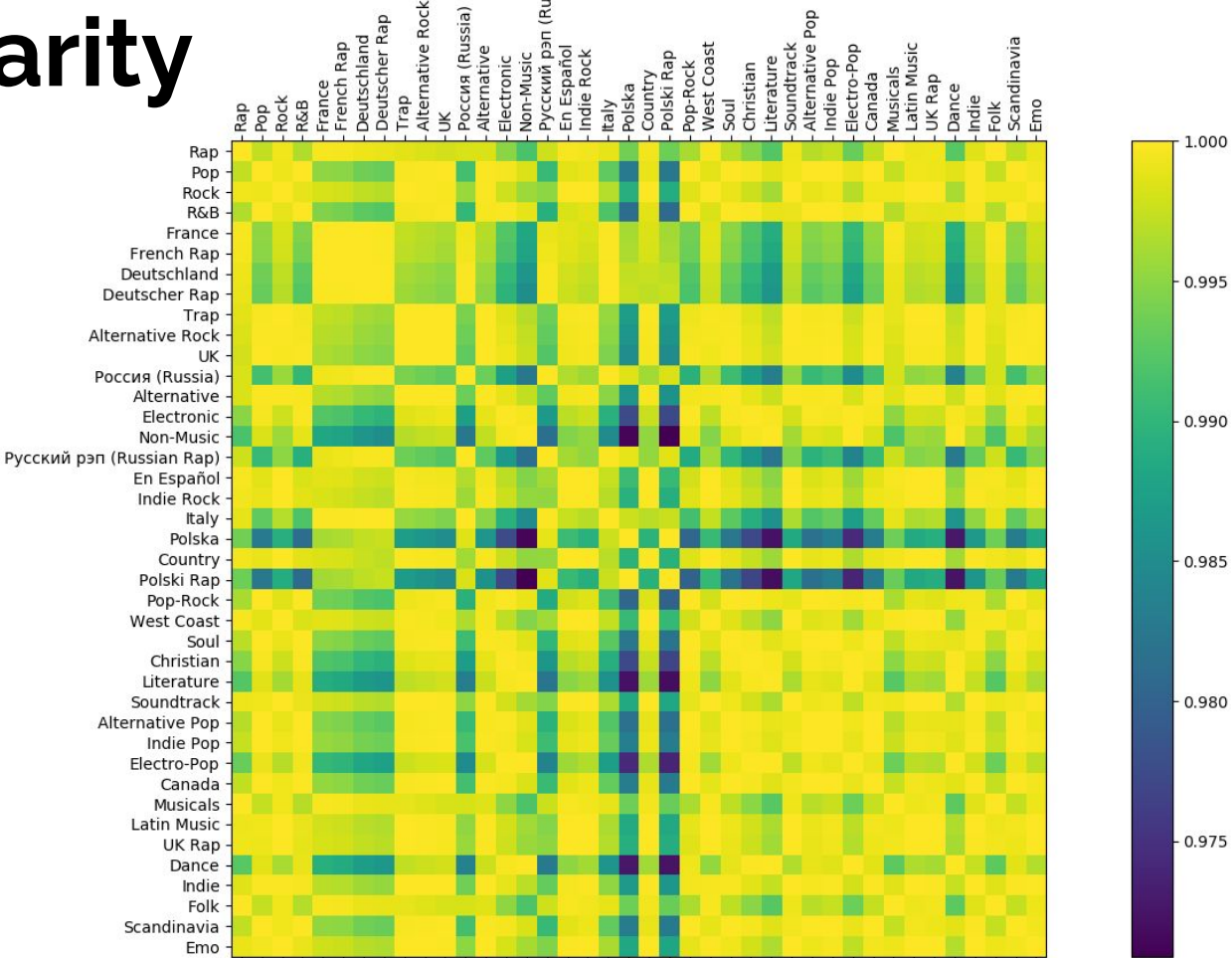
The logo for last.fm, featuring the text "last.fm" in a white, lowercase, sans-serif font centered on a solid red rectangular background.

last.fm

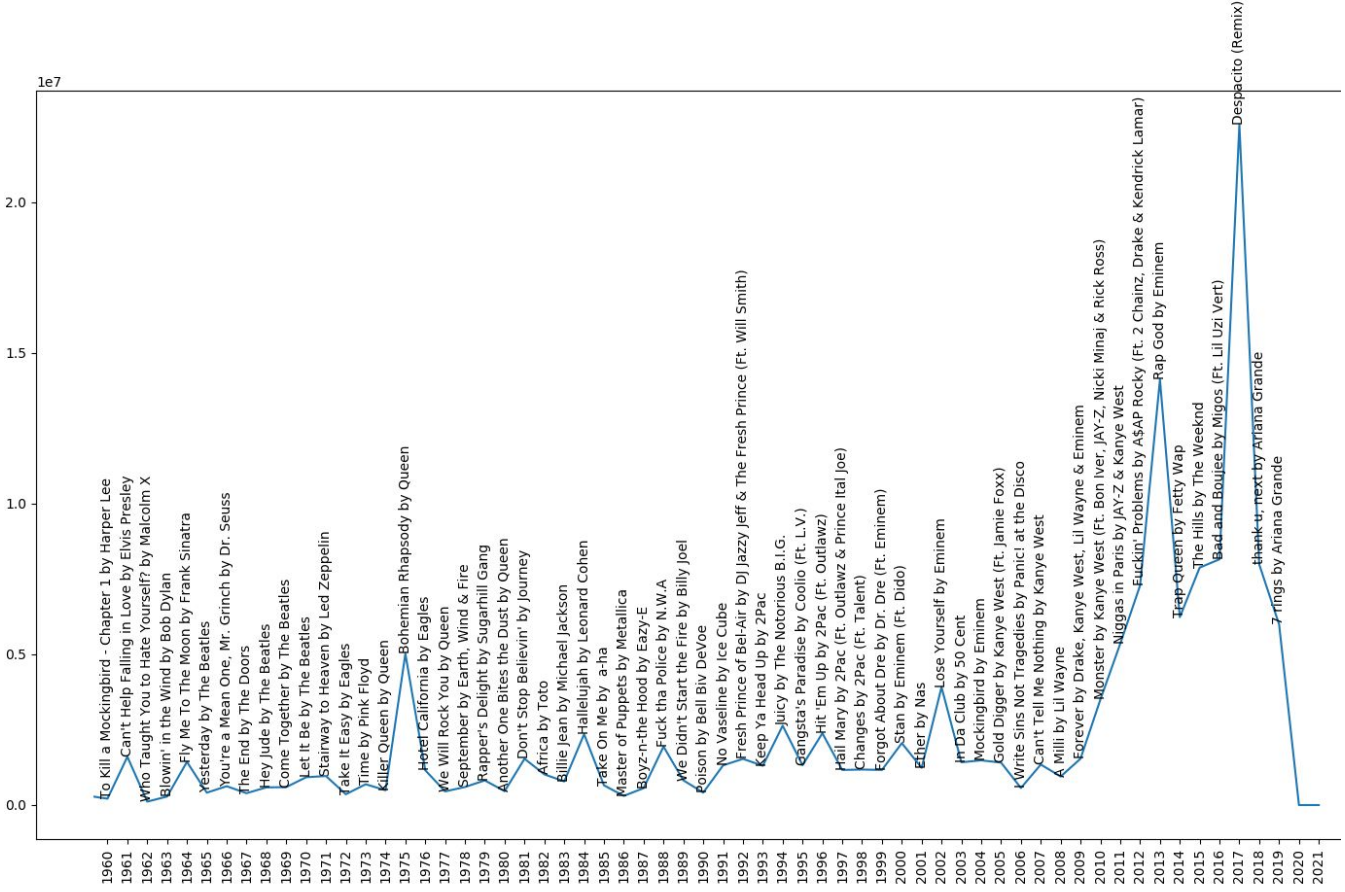
The logo for Genius, featuring the word "GENIUS" in a black, uppercase, serif font centered on a solid yellow rectangular background.

GENIUS

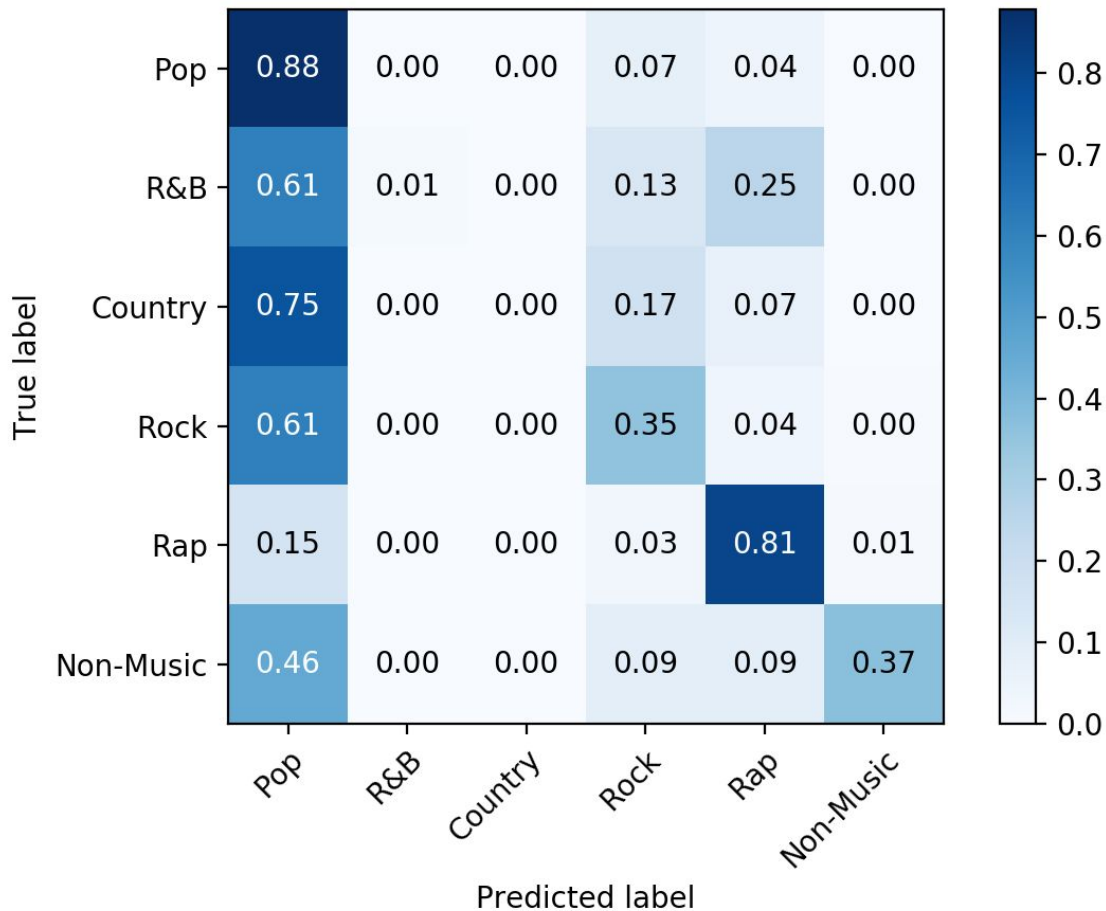
# Genre similarity



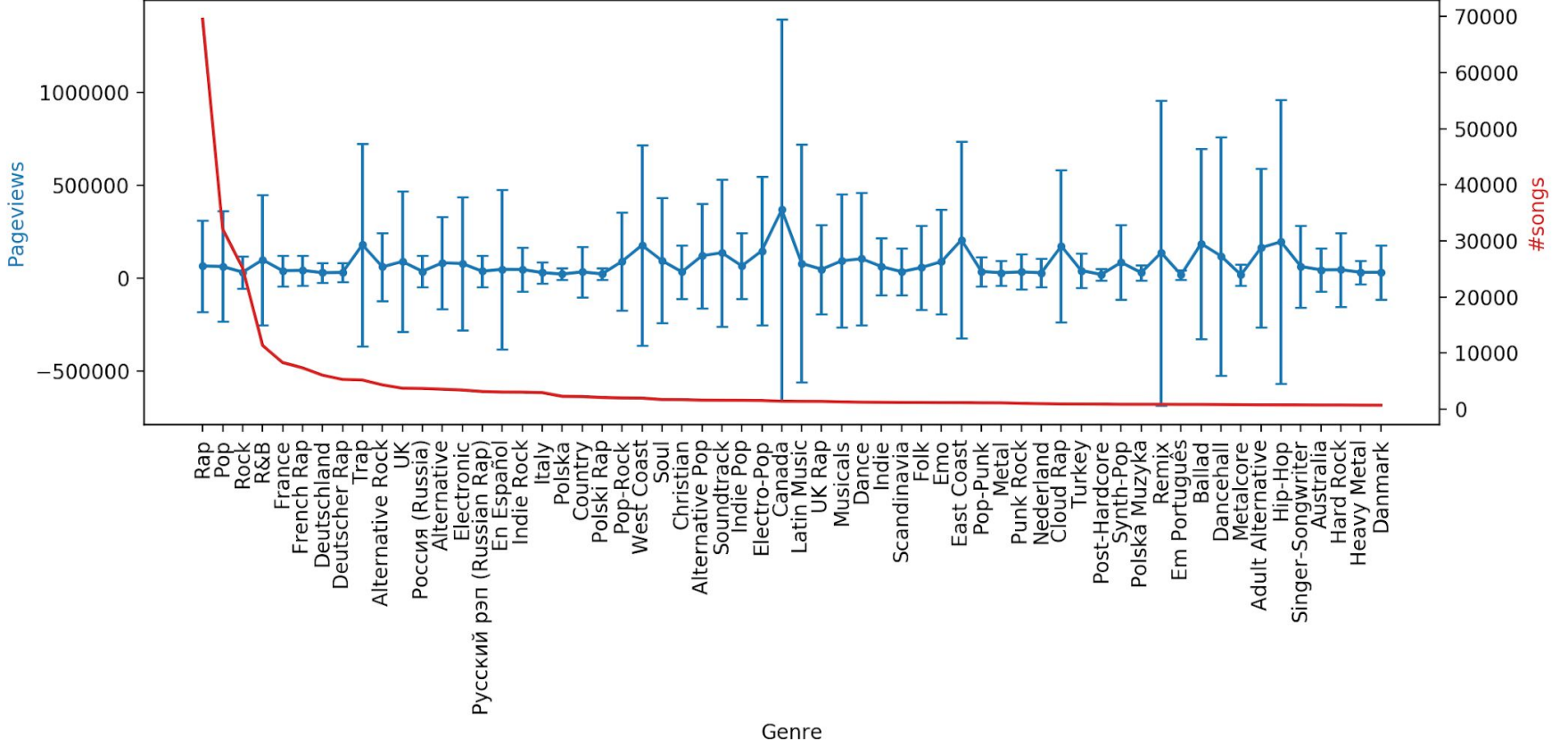
# Most popular song per release year



Confusion matrix



Average Genre Popularity



Average Genre Popularity

