Demystify Trends of Artists with Twitter, Reddit and Spotify API

Ketaki Khade
kkhade1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Nikita Shriram Navkar
nnavkar1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Swarali Patil
spatil32@binghamton.edu
Binghamton University
Binghamton, NY, USA

Tanmayee Kulkarni
tkulkar1@binghamton.edu
Binghamton University
Binghamton, NY, USA

Yash Jawale
yjawale1@binghamton.edu
Binghamton University
Binghamton, NY, USA

ABSTRACT

This project's main objective is to analyze data on currently popular musicians in the context of numerous social trends. Due to web's global presence and accessibility, social media platforms are crucial in boosting the popularity of any kind of music and music-based events that have an emotional impact on people. Reddit and Twitter are social media web platforms that millions of people use to share their ideas and opinions on a variety of topics. As a result, a stream of text data is produced that is made up of a huge number of tweets, posts and comments. In this project, we collected millions of posts tagged with "r/spotify", "r/music", "r/musician", "r/singer," etc. using Reddit API and also related tweets that refer to certain Spotify tunes, music, albums and artists from Twitter API. We analysed those tweets, 'r/spotify', 'r/music', 'r/musician', 'r/singer', posts, and comments and gather track-related data using the Spotify API. Finally, we identified the artists who are now trending by using the information we have acquired in combination with a particular time frame. We gathered a large number of tweets and posts from Twitter and Reddit, then we extracted the specific information, such as the URLs of popular songs or tracks, and we used those URLs to obtain artists from Spotify. We used the Spotify API to analyse those tweets and posts and extract information about the tracks. We plotted different type of graphs in this analysis. Taken together, our results provide a data-driven, analysis of trending artist in particular time frame in music industry.

KEYWORDS

Twitter API, Reddit API, Spotify API, MongoDB, Data Collection, Snowball Sampling

ACM Reference Format:
Ketaki Khade, Nikita Shriram Navkar, Swarali Patil, Tanmayee Kulkarni, and Yash Jawale. 2022. Demystify Trends of Artists with Twitter, Reddit and Spotify API. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

People of all ages value music greatly. It is a means of communicating and expressing human feeling. Popular music evolves in response to current social events. For instance, many candidates use music during election season to appeal to voters for a number of reasons, and campaign songs play a significant role. The music industry is currently undergoing ongoing change, with several transitional music genres experiencing popularity swings. There are a number of various platforms and sources for online music streaming, including Spotify, Apple Music, YouTube Music, Reddit, and others. Everybody has different musical tastes, both in terms of genre and artist. Even if it's not always better, if we hear something new, we usually share it with our friends, peers, and relatives. This eventually contributes to the creation of the music that is now popular or trendy. Using various hashtags, we frequently have a tendency to tweet, post, or share about it on social media, making it widely apparent. This unlimited, never-ending flow of knowledge about music from many sources could be used in creative ways in academia and the music industry. We've acquired real-time music-related data for this project from social networking sites like Twitter and Reddit as well as other music streaming services like Spotify, which offers a lot of music-related data. By collecting the data from these three sources over a specific period of time, we conducted analysis to discover which artists are popular and trending.

2 DATASET DESCRIPTION

2.1 Data Source

Here, in this project we used Twitter and Reddit data from their respective API's to collect tweets and posts. Using that data, we collected information about top trending songs. For this we used Spotify web API to fetch data from a Spotify URL. We extracted track ID from the URL and sending it as a parameter to the Spotify API to retrieve details of any available track. For Twitter we used streaming API and for Reddit, we collected data using $https://www.reddit.com/api/v1/access_token$ endpoint.

2.2 API METHODS

We will access data resources through standard HTTPS requests to an API endpoint. Web API uses appropriate HTTP verbs like; GET, POST, PUT and DELETE methods and returns all response data as a JSON object. As Twitter and Reddit APIs allow us to find, retrieve and engage or create different resources, for this it uses two HTTP

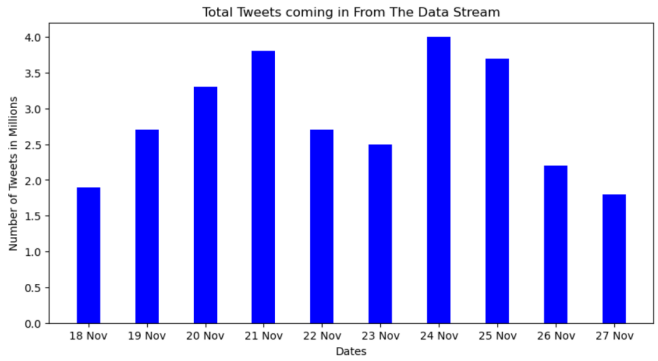

Figure 1: Total Tweets Coming in from Data Stream

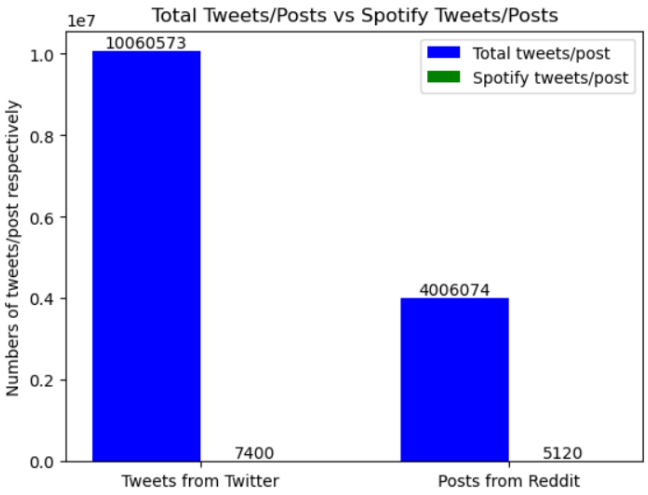

Figure 2: Total Tweets/Posts vs Spotify Tweets/Posts

methods: POST and GET. In this case using GET method we can retrieve data from Twitter and Reddit APIs. As Spotify offers us API methods based on REST principles. We begin receiving a stream of data and then the data will be in JSON format in response.

2.3 DATA EXPLORATION

We are able to process approximately 4 million tweets and 20 thousand posts daily after successfully scheduling the cron job on the VM for the same. Out of these, we are saving in the database around 0.0625 percent (2500) of the tweets/posts that have been filtered out based on a "Spotify" match. We filtered away the tweets and posts that have Spotify URLs before we examined the data from Twitter and Reddit. To extract the track id and other valuable information that can be used to fetch and store track, artist, and genre information from the Spotify API into our database, we used the URLs stored within those specific objects in the database. We got enough data to do our research after removing the tweets and posts and retrieving and saving the Spotify data in our database. We were able to discover relevant facts and insights concerning the three major purposes of this project with the aid of the combined data from two

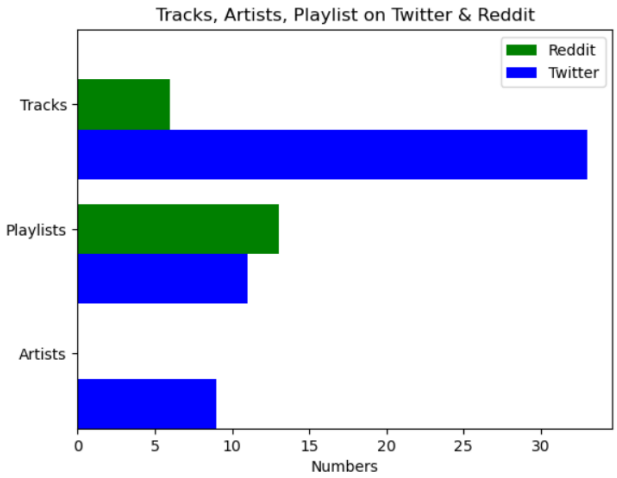

Figure 3: Tracks, Artists, Playlist on Twitter Reddit

separate sources and certain data science tools and approaches. We used MongoDB as database tool for storing the tweets from Twitter and posts from Reddit. We mainly used matplotlib for plotting graphs. For data cleaning and analysis we used NumPy and Pandas, 'request' library to fetch data from Twitter, Reddit and Spotify API's. Along with these we also used date, pymongo, tabulate, urlparse etc libraries for overall analysis.

3 BACKGROUND AND RELATED WORK

- Due to the growth of the web, new distribution networks of music sites facilitated an enormous amount of music that can be listened regularly. These new channels for distribution, like music streaming services, produce and maintain useful information about users and their listening habits. Most of the time, the data that these companies gather is confidential. These private data collections have led to datasets that are easily accessible and regularly used for implementation and evaluation. There are certain databases, such as the million-song dataset (MSD) are available, although they are usually out of date. To address the issue of a lack of current and openly accessible data for the development and assessment of recommended systems relating to the most recent artists in the music industry, we took advantage of the fact that a lot of users uses music streaming platforms to tweet about what they are listening to on the micro-blogging site like Twitter. Therefore, we used the free Twitter and Reddit API to gather real-time trading information about the latest music and musicians of album. Using a dataset derived from such tweets, we implement and evaluate a filtering based music analysis system and show that this is a promising approach.
- Earlier we were using r/artist for fetching the post related to trending artists (singers/musicians/bands etc) from Reddit. Although the data we received was insufficient for making a strong analysis regarding the trending artist in the field of music, r/artist consists of more post related to the artist from other fields like drama, portraits, dance, etc, and no posts

related to a music artist. Hence we decide to drop the r/artist subreddit and focus on other subreddits which will provide us with more robust data for trending artists in music. We used various cloud generators like word cloud, cloudly, and abcya to find the words related to music and on those, we tried to find the related subreddit. We used r/music, r/spotify, r/musician, r/songs, and r/singer subreddits to support our analysis.

4 DATA FLOW

As we can see in Figure 8, we first fetch data from Twitter API using the Twitter Stream API. We have used python's request library which provides us an easy way of recursively sending API requests to the twitter Stream API. We then filtered the text field of the received data for relevant key words like 'spotify' and saved those tweets in our MongoDB database. We further filtered the tweets in the MongoDB database to extract the URL and check the URL is in a valid format, as the URL will provide us with the Spotify track ID, which used to fetch the track information using Spotify API to get details like the artist name, genre, etc. We then saved these details in a separate database in MongoDB and used these in our analysis and visualization going ahead.

Artists	Twitter Tweets
BTS	1510
V	718
Jung Kook	565
Various Artists	352
Cher Lloyd	133
BE:FIRST	102
Charlie Puth	97
JIN	85
Mark Tuan	61
Jong Ho	33

Figure 4: Trending Artists on Twitter

5 OBJECTIVES

The main objective of our project is to get more clarity and information on what type of music people listen to, which artist and albums are more trending and how these trends change with time. Moreover, we are focusing on to analyze top trending artists on Twitter and Reddit with respect to our data collected on Spotify. We will take top artists/music on Spotify, and compare whether they are trending on Twitter and on Reddit. (RESEARCH QUESTIONS)

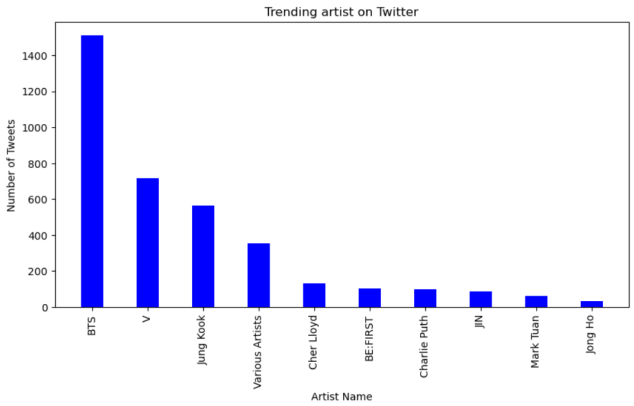

Figure 5: Trending Artists on Twitter

Artists	Reddit Post
Merkwood Music	13
Miles French	13
Various Artists	12
Audiomachine	11
Susumu Hirasawa	9
Mark Petrie	8
Ben Berkenbosch	7
Garrett Weyenberg	6
Hans Zimmer	6
Deftones	6

Figure 6: Trending Artists on Reddit

6 METHODOLOGY

We will pull track ID from Spotify URL which we get from our data and use this Track Id to send GET HTTP request to Spotify API. GET request returns the track details such as the track name, the artist name, album name etc. which will be saved into database. We will find the popularity of artist by analysing data we collected in database. In the same way, we will collect data from Reddit and Twitter API, using which will be used to determine the top trending artists and albums on twitter and Reddit. By collecting data related to popular artists we will show change in popularity of artist with time by using plots, which help to determine change in trade with time.

7 DISCUSSION/OBSERVATION

Using the Twitter stream API, we displayed the total amount of incoming tweets from the twitter stream as shown in Figure 1.

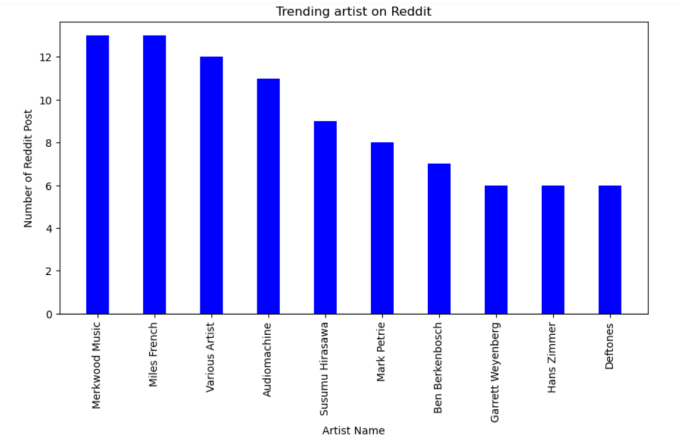

Figure 7: Trending Artists on Reddit

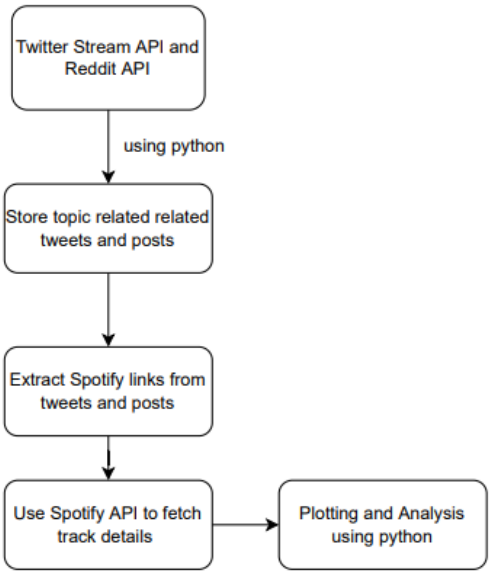

Figure 8: Data Flow Diagram

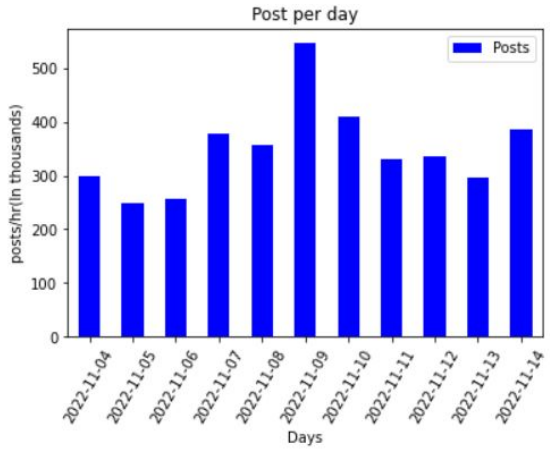

Figure 9: Political Posts per day on Reddit

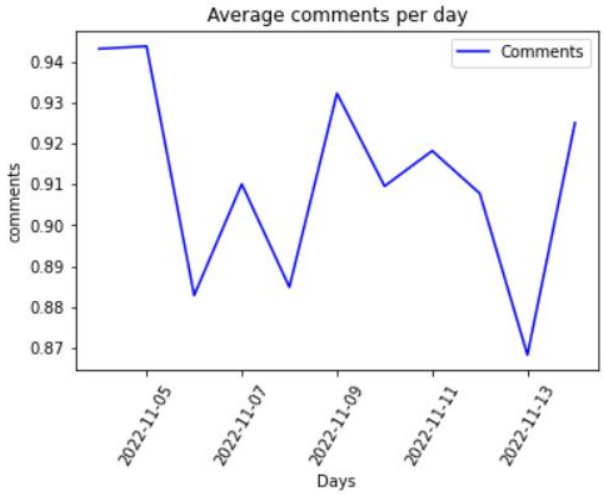

Figure 10: Average comments per day

We observed the daily variations in the amount of tweets arriving, which are brought on by the intermittent disconnection of our stream due to various circumstances. The repeated crashes and manual restarts required by the data gathering program also contributed to the graph's high level of fluctuation. Figure 2 depicts the correlation between the total number of tweets processed so far and the percentage of those tweets that match the keyword "spotify" to filter out tweets that might be of use. Along with that, the graph shows the overall number of Reddit posts and the proportion of posts relevant to Spotify. Since individuals typically communicate their musical preferences on private chat platforms rather than on Twitter or Reddit, there is a significant discrepancy between the overall number of tweets and the final useful tweets. We used these potentially helpful tweets/posts later on in our data processing framework to determine the music-related information contained in the tweet object's 'URLs' field, such as tracks, playlists, artists, albums, etc. i.e Snowball Sampling technique. Figure 3 shows the frequency of tweets and posts mentioning various spotify content characteristics, such as playlists, tracks, artists, albums, and podcasts. This offers us a general sense of the information being gathered and how we might utilize it to present our research of popular artists over a specified time period. The top trending artist on Twitter is shown in Figure 4 table along with the number of tweets that are specifically about that musician. Similarly for Reddit, the analysis is shown in Figure 5. The table shown in Figure 6 lists the top trending artists on Reddit together with the number

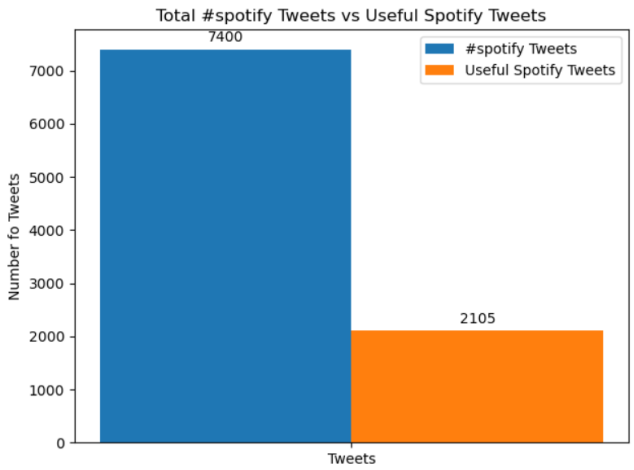

Figure 11: Total spotify Tweets vs Useful Spotify Tweets

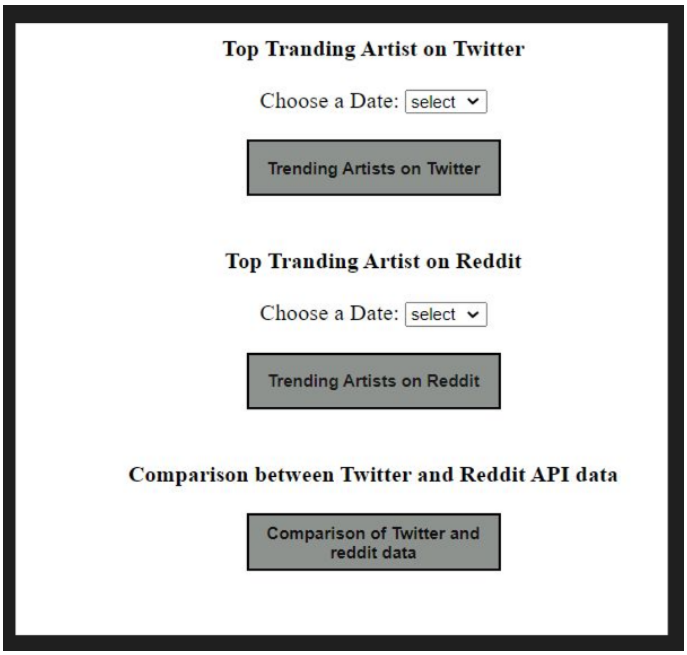

Figure 12: User Interactive UI

of times their music has been played throughout the specified period of time. This analysis has been plotted as shown in Figure 7. The underlying MongoDB database is used to retrieve this Twitter and Reddit data. Figure 9 shows the peak in posts for the r/politics subreddit between November 4 and November 14. Elections for the US Senate were held during this time. We can see that there was a rise, particularly on November 10, which was the day after the results of the US Senate election. Figure 10 shows the results of our analysis of the comments made for these postings using the snowball sampling technique together with this data for the word "election". Figure 11 shows the Total no. of Tweets vs The Tweets which contained Useful track URLs which we then used to plot our top trending artists. Finally, Figure 12 shows user interactive UI design using which user can get information about Trending artists on Twitter, Trending artists on Reddit and comparison between Twitter and Reddit API data.

RESEARCH QUESTION ANSWERED:

- find relationship between Social Media(Twitter and Reddit) mentions of top trending artists and trending artists on Spotify Music platform.

ANALYSIS QUESTIONS :

- Top trending artist in a given time-frame on Twitter and Reddit
- Comparison of Datasets on Twitter and Reddit
- Analyze the variation in the most popular artists on Twitter and Reddit.

8 PROJECT 3 IMPLEMENTATION

For the dashboard, we ended up making a web-app using Python flask and HTML, which takes two dates as input in MM-DD format and plots the requested plot between those two dates. When we execute the flask python script, as seen in Figure 12, the web-app is hosted on our local-host:9090 port, and we can input the dates and generate plots there. When we input the dates and click the generate plot button, in the backend, a function is called with the two dates as input parameters, and the function returns a list of the artists/songs being mentioned in tweets/posts between those dates. We then calculate how many times the artist/song has been mentioned in the time period and return the data-frame to the flask app, which generates the plot for the same using matplotlib. We used X2GO client to access the VM and make the dashboard and access our local-host. The X2GO client needs a host IP and a password, we used our server IP, which is 128.226.29.11 as our host IP, and connected as one of the users of the VM, using their password. Also note that we had to be on our university VPN to access the VM. We then had GUI access to the VM, which we used to make and deploy our web-app.

9 DASHBOARD

As we can see in the Figure 12, it represents the home page of our flask web-app,where we can enter the dates in MM-DD format to generate the two plots. The range for the dates is 11-13 to 12-14 Figure 5 and Figure 7 show us the top trending songs and the artists on both platforms respectively for the time period 11-19-2021 to 11-28-2021. We have used two separate plotting mechanisms and made it such that only one plot is plotted at a time, in order to respect the spotify rate limits, which is a 30 second rolling window. Spotify does not mention a particular rate limit, but we encountered issues periodically because of this.

10 BUG-REPORT

The plots, especially the comparison between Twitter and Reddit APIs is taking some time to produce since they use the Spotify API to send requests, retrieve data in real time from the MongoDB, parse the answer, and then provide the dataframe. Despite this, we do not currently have major bugs.

11 RESULTS

The top trending artists for the input time period are shown in Figures 5 and 7. According to data gathered from Twitter and Reddit, BTS is the top trending artist over Twitter platform, whereas Merkwood Music is the top trending artist according to data gathered from Reddit API. Since not everyone makes their opinions public, data from Twitter and Reddit varies. Additionally, Figure 3 compares the information gathered from the Reddit and Twitter APIs. Overall, this analysis demonstrates that there is no connection between top trending artists' social media (Twitter and Reddit) mentions and top trending artists on the Spotify Music platform.

12 CONCLUSION

As per our observation, we observed that, trending artists on Twitter are different from that of trending artists on Reddit. Different people have different preferences to express their music interests. And as per our analysis, Twitter is most used platform as compared to Reddit. We also observed that people ask about song suggestions on Twitter than on Reddit.

13 REFERENCES

- Gerlitz, C. Rieder, B. (2013). Mining one percent of Twitter: collections, baselines, sampling. M/C Journal, vol. 16, no. 2.
- Alina Campan, Tobel Atnafu Joseph Nolan (2018). Is Data Collection through Twitter Streaming API Useful for Academic Research? 2010 IEEE Tenth International Conference on Peer-to-Peer Computing (P2P). Spotify – Large Scale, Low Latency, P2P
- Abel, C. Hauff, G.-J. Houben, R. Stronkman, and K. Tao. Semantics+Filtering+Search=twitcident. Exploring Information in Social Web Streams. In Proc. of the 23rd ACM Conference on Hypertext and Social Media, pages 285–294. ACM, 2012.
- Dooms, T. De Pessemier, and L. Martens. MovieTweetings: a Movie Rating Dataset Collected From Twitter. In Workshop on Crowdsourcing and Human Computation for Recommender Systems, 2013.
- T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere. The million song dataset. In A. Klapuri and C. Leider, editors, Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011), pages 591–596. University of Miami, 2011.
- M. Schedl, P. Knees, and G. Widmer. Investigating web-based approaches to revealing prototypical music artists in genre taxonomies. In Proceedings of the 1st International Conference on Digital Information Managemen