Music Tweet Map: A Browsing Interface to Explore the Microblogosphere of Music

David Hauger Markus Schedl
Johannes Kepler University
Department of Computational Perception
Linz, Austria
Email: {david.hauger,markus.schedl}@jku.at

Abstract—**In this demo paper, we present the "Music Tweet Map" interface for browsing music listening events on a global scale. These events have been extracted automatically from a large set of microblogs harvested from Twitter. We showcase the major functionalities offered by the interface, i.e., browsing music by time, specific locations, topic clusters learned from tag information, and music charts. Furthermore, music can be explored via artist similarity. To this end, we present a music similarity measure, based on co-occurrence analysis of items in users' listening histories.**

I. Introduction and Motivation

Nowadays, many people are eager to share a wide range of aspects of their daily lives via social media. These aspects of course also include information about their music preference and music consumption. Exploring music listened to by others, even in different parts of the world, represents an interesting opportunity to get to know new music, e.g., [1], [10], [7], [6], [4]. While music recommender systems [9] also have the same goal, the browsing paradigm implemented in this demonstrator particularly addressed the human desire to joyfully explore and encounter new stuff, in visual and auditory ways.

Targeting this desire, we present the "Music Tweet Map" (MTM) interface[1] that allows to interact with a repository of listening events shared on the microblogosphere. In contrast to an earlier introduction of the MTM [2], the current version — while still a research prototype — has a completely reworked and tidied up user interface and further provides a few new functionalities. The most remarkable ones are the extended ways to create charts for artists and genres, as well as the possibility to aggregate tweets, i.e., listening events. While the previous version could only display all individual tweets separately, which considerably slowed down the interface when a large number of tweets were displayed, in the current one, the user can decide to aggregate all tweets within a certain radius.

The remainder of this demo paper is structured as follows. Section II first details the dataset we used to build the browsing interface. Subsequently, the main functionalities of the MTM are presented, in particular focusing on the computation of *topic clusters* to explore music by genre or style and the *co-occurrence similarity measure* to discover artists similar to a given seed. We round off the paper in Section III by summarizing our work and pointing to possible extensions of the interface.

[1]http://www.cp.jku.at/projects/MusicTweetMap

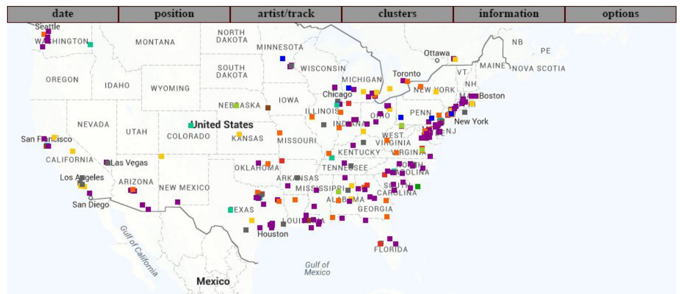

Fig. 1. Screenshot showing all identified music listening events for one day in the USA. In the upper part, the top-down menus can be seen. Different colors indicate different topic clusters (styles or genres).

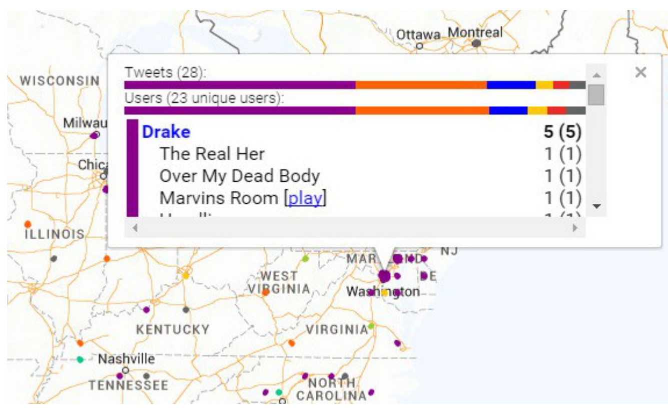

Fig. 2. Illustration of the aggregated listening events for a certain area. Tweets shown in the selection are grouped by artist. In the top of the selection tool, the distribution of listening events over topic clusters (encoded in different colors) is shown.

II. Music Tweet Map

In the following, we present the dataset used to create the MTM interface, before explaining its main functionalities.

A. Dataset

We exploited the "Million Musical Tweets Dataset" [3] of geo-tagged listening events mined from Twitter.[2] To create this dataset, we retrieved all posts provided by the Twitter Streaming API[3] in the time period September 2011 to April 2013 and filtered them to include only those with location information. Subsequently, we applied a pipeline of pattern matching rules to identify potential listening events, e.g., using patterns of the form `#nowplaying [song] by [artist]`. To increase accuracy of such detections, e.g., excluding posts like `#nowplaying Diablo 3 by Blizzard` (a popular computer game), we matched the potential song and artist names with the MusicBrainz[4] database and only retained tweets in which both artist and track name pointed to valid database entries. Eventually, this process resulted in more than one million geo- and time-annotated listening events. To be able to play the actual music, we further mapped the extracted listening events to music snippets offered by Amazon[5] and 7digital.[6]

B. Music Search and Exploration in the MTM

Basic functionalities and visualizations: A screenshot of the main interface is shown in Figure 1. In this image, listening events are not aggregated, but shown separately, each as square. In Figure 2, on the other hand, listening events are grouped by area and illustrated by discs whose size equal the number of tweets they encompass, and whose color encodes the main topic cluster, i.e., genre or style (see below). To aggregate listening events, we use an adaptive grid whose grid size is adjusted based on the zoom level. Upon selecting a disc, a pop-up window is shown, in which not only the artists and songs, but also the distribution of the songs among topic clusters can be seen. Songs we could identify on Amazon or 7digital can be previewed via a "play" link. In addition to the specific features of the MTM browsing interface, which are explained below, it also allows to search for particular artist and track names.

Temporal and location-based filtering: As illustrated in Figure 3, the user can opt to display only listening events within a certain time range, shown here for April 2013. Location-based filtering is possible either by specifying a window via longitude and latitude coordinates or by selecting an item from lists of countries, states, and cities. Figure 4 depicts the result when the user filters by country Portugal.

Exploration by genre and style via topic clusters: In order to enable filtering according to certain kinds of music, we first downloaded from Last.fm[7] all collaborative tags available for each artist. We then filtered these tags by a list of 1,944 genre and style names from Freebase.[8] Subsequently, we applied non-negative matrix factorization [5] on the artist-tag-occurrence matrix, which is used to categorize each artist into a number of $k = 10 \ldots 20$ topic clusters, where k can be adjusted by the user. In the interface, each topic is assigned a color and is described by its most important genres. Figure 5 illustrates the filtering tool for genre clusters. As can be seen, this functionality may serve to identify which kind of music is popular in which regions of the world. While people in North America foremost listen to cluster 5, which is characterized as hip-hop and rap, South Americans prefer cluster 2, i.e., rock, hard rock, and alternative music.

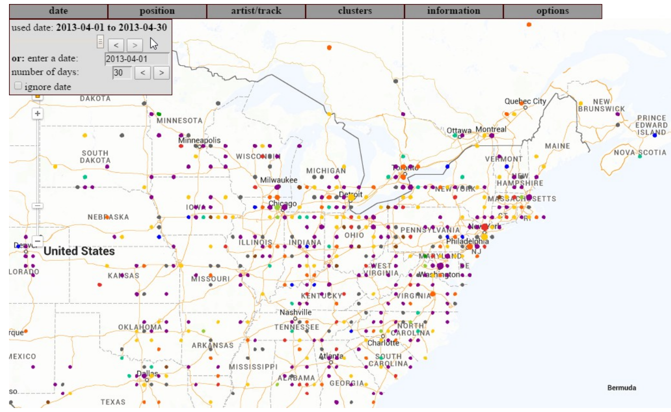

Fig. 3. View on music listening events in the Eastern part of the USA, filtered by date; in this case, April 2013.

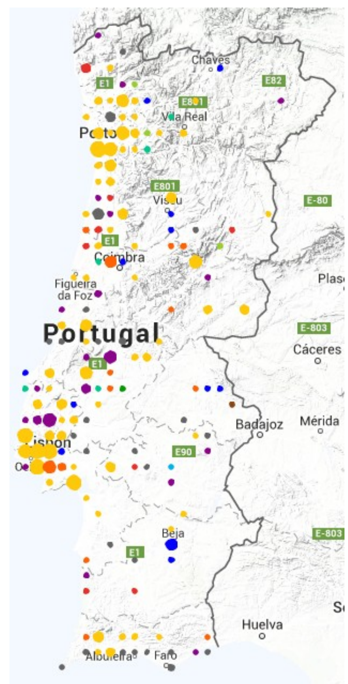

Fig. 4. View on music listening events posted in Portugal, between September 2011 and April 2013.

[2]https://twitter.com
[3]https://dev.twitter.com/docs/streaming-apis
[4]https://musicbrainz.org
[5]https://www.amazon.com
[6]https://www.7digital.com
[7]http://last.fm
[8]http://www.freebase.com

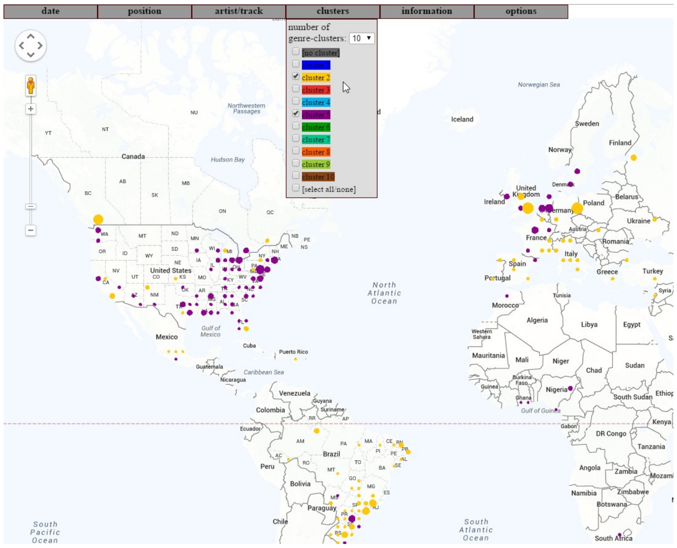

Fig. 5. Filtering based on topics that correspond to genres and styles.

Exploration by similar artists: Discovering music by similarity to the user's preferred genre or artist is a key feature of most music retrieval and recommendation systems. To provide this kind of functionality, we implemented a co-occurrence similarity measure based on the listening histories of the users in the dataset, which was already proven useful for this task [8]. More precisely, we computed artist similarity according to Formula 1, where o_i denotes the total number of listening histories[9] in which artist i occurs and $co_{i,j}$ is the number of listening histories in which artists i and j co-occur.

$$sim_{i,j} = \frac{co_{i,j}}{\sqrt{o_i \cdot o_j}} \tag{1}$$

In the user interface, similar artists are displayed using a color mapping that expresses the degree of similarity with respect to the selected target artist. Figure 6 shows a corresponding screenshot. The target artist in this case is Xavier Naidoo.[10] On the map shown in the left part of the figure, listening events to the target artist are depicted as black squares, while shades of red are used to indicate different degrees of similarity to the target, darker shades meaning higher similarity. On the right side of the figure, a list of most popular artists is shown, sorted in descending order of total playcounts. The similarity coding on the very right shows that the artists most similar to Xavier Naidoo are Silbermond, Rosenstolz, Glasperlenspiel, and Unheilig. While they make music of different genres, they are all popular German artists with foremost German lyrics. This fact also illustrates the performance of the similarity measure, because content-based similarity approaches would most likely not have been able to infer such aspects of similarity from the audio.

[9]A listening history is defined as the entirety of listening events of a user.
[10]https://en.wikipedia.org/wiki/Xavier_Naidoo

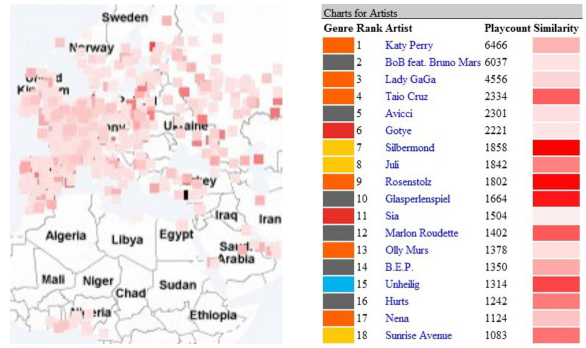

Fig. 6. Exploring music similar to Xavier Naidoo. Darker shades of red indicate higher similarity to the target artist. The left figure shows the corresponding map, while the right one lists the artists most similar to Xavier Naidoo, sorted by popularity.

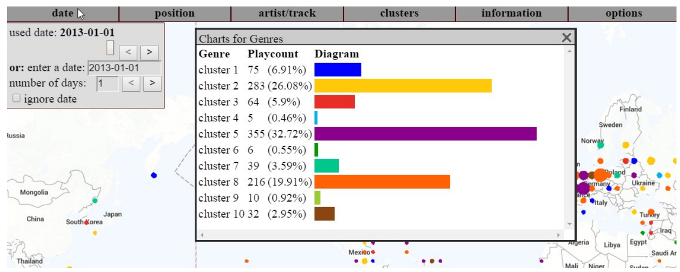

Fig. 7. Global genre charts for January 1, 2013.

Artist and genre charts: To explore which artists or genres are particularly popular in a given time frame or geographic area, the MTM interface can create respective charts. Examples for genre and artist charts are shown in Figures 7 and 8, respectively. The former illustrates the global distribution of listening events over 10 genre clusters, for January 1, 2013. Clusters 2 and 5 were already explained above; cluster 8 includes pop and rnb music, cluster 1 electronic, and cluster 3 indie music. Cluster 10 encompasses death and black metal, while cluster 7 represents folk and country music. Clusters 4, 6, and 9, represent ambient, punk, and jazz music, respectively. Figure 8 visualizes the top 20 artists in Brazil for the time period covered in the dataset, i.e., September 2011 to April 2013. In this snapshot, most of the artists in the charts belong either to cluster 2 (rock), 8 (pop and rnb), or 5 (hip-hop and rap).

III. Conclusion and Future Work

We presented the Music Tweet Map interface to visually and aurally explore a dataset of over one million geo-localized listening events mined from tweets. The web interface provides a wealth of functions to access the collection: among others, meta-data based search, listening to audio previews, and filtering by time ranges, areas defined by coordinates or countries, states, or cities. Furthermore, the collection is automatically organized into a number of topic clusters learned from tag occurrences. These clusters roughly correspond to

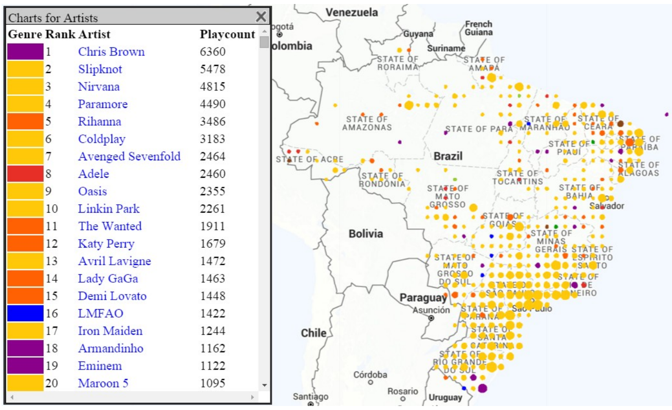

Fig. 8. Artist charts for Brazil, considering the entire time period covered by the dataset, i.e., September 2011 to April 2013.

music genres or styles, are color-coded in the interface, and can be used for filtering. To discover artists similar to a selected target artist, we implemented a similarity measure based on co-occurrences of artists in listening histories. The user interface correspondingly provides a view in which only artists similar to the target are shown. They are colored in a way that reflects the degree of similarity to the target. The MTM also supports the creation of artist and genre charts, either on a global scale or temporally or spatially filtered.

Extensions we contemplate include integrating a content-based music similarity measure in addition to our co-occurrence similarity measure. Respective audio-based features, such as descriptors of rhythm, timbre, melody, or key, could further enable a deeper analysis of music preferences around the world. Likewise, exploiting more precise temporal information, e.g., hour of the day, we could presumably relate listening events to certain activities (e.g., working hours versus leisure time) and in turn perform additional analyses.

Acknowledgment

This work was supported by the Austrian Science Fund (FWF): P25655.

References

[1] M. Hamasaki and M. Goto. Songrium: A Music Browsing Assistance Service Based on Visualization of Massive Open Collaboration Within Music Content Creation Community. In *Proceedings of the 9th International Symposium on Open Collaboration (WikiSym)*, pages 4:1–4:10, New York, NY, USA, 2013. ACM.

[2] D. Hauger and M. Schedl. Exploring Geospatial Music Listening Patterns in Microblog Data. In *Proceedings of the 10th International Workshop on Adaptive Multimedia Retrieval (AMR)*, Copenhagen, Denmark, October 2012.

[3] D. Hauger, M. Schedl, A. Košir, and M. Tkalčič. The Million Musical Tweets Dataset: What Can We Learn From Microblogs. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil, November 2013.

[4] P. Knees, M. Schedl, T. Pohle, and G. Widmer. An Innovative Three-Dimensional User Interface for Exploring Music Collections Enriched with Meta-Information from the Web. In *Proceedings of the 14th ACM International Conference on Multimedia*, Santa Barbara, CA, USA, October 2006.

[5] D. D. Lee and H. S. Seung. Learning the Parts of Objects by Non-negative Matrix Factorization. *Nature*, 401(6755):788–791, 1999.

[6] S. Leitich and M. Topf. Globe of Music - Music Library Visualization using GeoSOM. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, September 2007.

[7] A. Lillie. MusicBox: Navigating the Space of Your Music. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, September 2008.

[8] M. Schedl and D. Hauger. Mining Microblogs to Infer Music Artist Similarity and Cultural Listening Patterns. In *Proceedings of the 21st International World Wide Web Conference (WWW): 4th International Workshop on Advances in Music Information Research (AdMIRe)*, Lyon, France, April 2012.

[9] M. Schedl, P. Knees, B. McFee, D. Bogdanov, and M. Kaminskas. *Recommender Systems Handbook*, chapter Music Recommender Systems. Springer, 2nd edition, 2015.

[10] S. Stober and A. Nürnberger. MusicGalaxy - An Adaptive User-Interface for Exploratory Music Retrieval. In *Proceedings of 7th Sound and Music Computing Conference (SMC)*, Barcelona, Spain, July 2010.