Analyzing the Potential of Microblogs for Spatio-Temporal Popularity Estimation of Music Artists

Markus Schedl
Department of Computational Perception
Johannes Kepler University
Linz, Austria
http://www.cp.jku.at

Abstract

This paper looks into the suitability of microblogs for an important task in music information research, namely popularity estimation of music artists. The research questions addressed are the following: To which extent are microblogs used to communicate music listening behavior? Are there differences between different countries of the world? Is it possible to derive a popularity measure from user's microblogging activities?
We found that microblogging does indeed represent an important communication channel for revealing music listening activities, although the intensity of its use vary considerably from country to country. Motivated by this finding, we took first steps towards a *geo-aware, social popularity measure* for music artists. To this end, we analyzed user posts mined from the microblogging service Twitter over a period of five months. Addressing the problem of determining the popularity of music artists, we employed a gazetteer on extracted posts relevant for particular music artists. The presented approach aims at extracting time- and location-specific artist popularity information. We evaluated the performance of the approach by comparing the popularity rankings derived from Twitter posts against the popularity rankings provided by last.fm, a popular music information system and recommender engine.

1 Motivation and Context

The emergence of microblogging services date back to 2005. However, they gained greater popularity not before the years 2007 and 2008[1]. Today's most popular microblogging service is Twitter[2], where millions of users post what they are currently doing or what is currently important to them [Kazeniac, 2009].

The work at hand tackles a problem from music information research (MIR), a field that is concerned with the extraction, analysis, and usage of information about any kind of music entity (for example, a song or an album) on any representation level (for example, an audio signal, a symbolic MIDI representation, or an artist's name) [Schedl, 2008]. Figuring out which artists/performers of music are popular is an interesting research question, not at last for the music industry, but also for the artists themselves and for the interested music aficionado.

In MIR we can distinguish three broad categories of modeling music items with respect to the underlying data source, namely *music content*-based [Casey *et al.*, 2008], *music context*-based [Schedl, 2011], and *user context*-based [Göker and Myrhaug, 2002] approaches. Feature vectors describing aspects from one or more of these three categories can be constructed, and similarity measures can be applied to the resulting vectors of two pieces of music or two music artists/performers. Elaborating such musical similarity measures that are capable of capturing aspects that relate to perceived similarity is one of the main challenges in MIR. Such measures are a key ingredient of various music applications, for example, automatic playlist generators [Aucouturier and Pachet, 2002; Pohle *et al.*, 2007], music recommender systems [Celma, 2008], music information systems [Schedl, 2008], semantic music search engines [Knees *et al.*, 2007], and intelligent user interfaces [Pampalk and Goto, 2007] to music collections.

For all applications mentioned, popularity information can be of particular value. For example, a music recommender system or a playlist generator will benefit from popularity information in that it will allow adapting its recommendations or playlists to different types of users. It is a well-known fact that different kinds of users (in terms of their music understanding and background) require different music recommendations, cf. [Celma, 2008]. In particular, music experts in certain styles or genres get quickly bored if such a system keeps on recommending popular artists that are already known to this sort of user. Incorporating popularity information – in this case, including "long tail" artists in the recommendations – is likely to yield serendipitous results for such users.

Addressing popularity as an important category of music-related information, the work at hand was driven by two research questions: To which extent are microblogs used to express music preferences and listening activities in different

[1] http://www.sysomos.com/insidetwitter (access: March 2010)

[2] http://www.twitter.com (access: January 2011)

places around the world? Is it possible to derive a popularity measure from user's microblogging activities? The first question will be answered by an analysis of microblogs about music listening in Section 3.1. Prior to that, Section 2 describes the data acquisition and popularity estimation steps. Here, we present first steps towards deriving *location- and time-specific music popularity information* from posts of `Twitter` users. The second question is addressed in Section 3.2, where we report on the results of quantitative experiments comparing the microblog-based popularity estimates with a reference data set extracted from `last.fm`[3]. Eventually, Section 4 summarizes the work and points out some directions for future research.

1.1 Microblog Mining

With the advent of microblogging, a huge, albeit noisy data source became available. Since millions of `Twitter` users tweet around the world, telling everyone who is interested what is important to them, microblogs are an obvious source to derive popularity information. However, it seems that literature dealing with microblogs mostly studies human factors (e.g., [Teevan *et al.*, 2011]) or describes properties of the Twittersphere (e.g., [Java *et al.*, 2007; Kwak *et al.*, 2010]). A general study on the use of `Twitter` can be found in [Java *et al.*, 2007]. Java et al. report that `Twitter` is most popular in North America, Europe, and Asia (Japan), and that same language is an important factor for cross-connections ("followers" and "friends") over continents. The authors also distilled certain categories of user intentions to microblog. Employing the *HITS* algorithm [Jon M. Kleinberg, 1999] on the network constructed by "friend"-relations, Java et al. derive user intentions from structural properties. They identified the following categories: information sharing, information seeking, and friendship-wise relationships. Analyzing the content of `Twitter` posts, the authors distill the following intentions: daily chatter, conversations, sharing information/URLs, and reporting news.
Scientific work related to content mining of microblogs includes the following: Cheng et al. propose a method to localize `Twitter` users based on cues ("local" words) extracted from their tweets' content [Cheng *et al.*, 2010]. Sakaki et al. propose semantic analysis of tweets to detect earthquakes in Japan in real-time [Sakaki *et al.*, 2010]. A more general approach to automatically detect events and summarize trends by analyzing tweets is presented by Sharifi et al. [Sharifi *et al.*, 2010].

1.2 Popularity Estimation for Music

Determining the popularity of a music artist or song is a relatively new research area. The earliest work in this direction, to the best of our knowledge, is [Grace *et al.*, 2008], where Grace et al. estimate popularity rankings based on user posts mined from `myspace`[4]. The authors apply different annotators to artist pages in order to detect artist, album, and track names, as well as descriptions of sentiments and spam. A data hypercube (OLAP cube) is then used to project the data to a one-dimensional popularity space. Based on a conducted user study, the authors conclude that the list generated by this method is on average preferred to the `Billboard` charts[5].
Using search queries raised in the Peer-to-Peer network `Gnutella` [Ripeanu, 2001],[Koenigstein and Shavitt, 2009] present an approach to predict music charts. The authors demonstrate that a song's popularity in the `Gnutella` network correlates with its ranking in the `Billboard` charts. For their analysis Koenigstein and Shavitt only consider the United States of America, because of its predominance in available data.
The work at hand is probably most related to [Schedl *et al.*, 2010], where popularity estimates of music artists are calculated on the country level using different data sources. Schedl et al. use page count estimates returned by Web search engines as result of music artist-related requests, user posts returned by `Twitter` as result of artist-related queries, information on music files shared by users of the `Gnutella` file sharing network, and playcount data extracted from `last.fm`. The approach proposed here is different from [Schedl *et al.*, 2010] in that Schedl et al.'s work only allow for an overall popularity prediction on the country level. It does neither take into account the *popularity on the level of individual cities*, nor the *time-dependence of the popularity estimate*. Both factors are, in our belief, indispensable for a fine-grained analysis of popularity.
Using content-based audio features and manually assigned labels to predict the popularity of a song is addressed in [Pachet and Roy, 2008]. Pachet and Roy's conclusion is, however, that even state-of-the-art machine learning techniques fail to learn factors that determine a song's popularity, irrespective of whether they are trained on signal-based features or on high-level human annotations.

2 Microblog Mining for Popularity Estimation

Since we are interested in the *spatio-temporal popularity distribution* of music artists, we first extracted in May 2010 from `World Gazetteer`[6] a list of world's largest agglomerations. The data set comprises 790 cities with at least 500,000 inhabitants. We further gathered the corresponding location information (longitude- and latitude-values).
Using these coordinates, we monitored user posts on `Twitter` that include geographic positioning information for a period of five months, more precisely, from May to September 2010. To this end, the geo-localization methods provided by the `Twitter` API[7] were used. We searched for the exact longitude- and latitude-coordinates of the agglomerations in our list and added a search radius of 50 kilometers in order to account for surrounding suburbs. Since we focused our analysis on music listening activities, we restricted the extraction of messages to posts including the `#nowplaying`

[3] `http://last.fm` (access: January 2011)
[4] `http://www.myspace.com` (access: November 2010)
[5] `http://en.wikipedia.org/wiki/Billboard_Hot_100` (access: May 2009)
[6] `http://world-gazetteer.com` (access: October 2010)
[7] `http://apiwiki.twitter.com/Twitter-API-Documentation` (access: January 2011)

hashtag, as this descriptor is commonly used to refer to music currently played by the user. It has to be noted, however, that only a few percentage ($< 5\%$) of tweets come along with geo-local information. Therefore, our results may be biased towards technology-affine users who possess the latest generation of smartphones or other mobile computing equipment.

Having gathered user posts together with spatio-temporal information in this way, we built a word-level index [Zobel and Moffat, 2006] applying casefolding and stopping. We then employed an annotation component, whose knowledge base comprised of $3{,}000$ names of music artists. To this end, we retrieved the overall most popular artists from `last.fm` using their Web API[8]. Since `last.fm`'s data is known to contain a high amount of misspellings or other mistakes due to their collaborative, user-generated knowledge base [Lamere, 2008], we cleaned the data set by first matching each artist name with the database of the expert-based music information system `allmusic.com`[9] and second retaining only those names that were also known by `allmusic.com`. For indexing the tweets we used a modified version of the `lucene`[10] indexer. We adapted the retrieval component for optimized retrieval of ranked artist sets, given a particular day and location. Ranking is simply performed according to the total count of artist occurrences in the respective posts, which corresponds to the *term frequency* $tf_{a,l,t}$ of term a (denoting the artist name) in the posts retrieved for a particular location l at a particular time t.

The list of agglomerations we used can be downloaded from `[omitted due to to double-blind review]`. The set of $3{,}000$ artist names is available at `[omitted due to to double-blind review]`.

3 Statistics and Evaluation

Experimentation was driven by two main questions. On the one hand, we were interested in the *distribution of music-related* `Twitter` *posts* around the world. Assessing whether the quantity of available information varies considerably for different regions of the world was one subtask. The other was investigating if any differences between the general use of `Twitter` and the posting of music-related information can be detected.

The second set of experiments addressed the question whether music-related posts are capable of revealing information on the *current popularity of music artists*. To this end, we compared the location- and time-specific information derived from the `Twitter` posts with artist charts gathered from the music information system `last.fm`.

3.1 Geographical Analysis of the Data

Analyzing the geographical distribution of `Twitter` users who reveal their current music taste (by including the `#nowplaying` hashtag in their posts) was our first objective. Figures 1 and 2 show the cities whose inhabitants are most active in terms of posting listening-related messages. Figure 1 reveals the top 10% of cities with the highest absolute number of music-related `Twitter` posts, whereas Figure 2 shows the cities with highest relative number of posts, normalized by the respective city's number of inhabitants. Taking a closer look at the most active agglomerations in absolute terms, the dominance of Asian metropolises stands out; six out of the top 10 cities are located in Asia, two in Europe, one in North America, and one in South America. When analyzing the activeness of inhabitants relative to their city's size, Brazil dominates the ranking (four out of the top 10 cities). Four cities are located in Asia, one in Europe, and five in South America (one in Chile, in addition to the four in Brazil). This might reflect the ascribed affinity of South Americans to music and their openness to talk about their habbits and activities.

To obtain an estimate of music-related `Twitter` use on the country level, we aggregated the data obtained for individual cities to the respective countries. The results are shown in Figures 3 and 4 as absolute amount of posts and as number of posts relative to the number of inhabitants, respectively.

These figures are largely in line with a report that explores the use of `Twitter` around the world [Evans, 2010] and shows the top 20 countries in terms of total number of posts contributed. However, some interesting outliers can be found. Comparing the list of top-ranked countries in terms of total tweet contributed [Evans, 2010] with the 20 top-ranked countries in terms of music listening-related posts (cf. Figure 3) allows to approximate in which countries `Twitter` is disproportionately often or seldom used to report on listening activities: Countries whose inhabitants are most frequent users of `Twitter`, but not among the top 20 in terms of sharing current listening activities are Australia (ranked 5$^{\text{th}}$ in terms of total tweet contributed according to [Evans, 2010]), Singapore (12$^{\text{th}}$), France (14$^{\text{th}}$), Ireland (15$^{\text{th}}$), New Zealand (17$^{\text{th}}$), Italy (19$^{\text{th}}$), and Iran (20$^{\text{th}}$). On the other hand, a disproportionately high amount of music-related posts in relation to overall `Twitter` usage can be found in China (ranked 6$^{\text{th}}$ in Figure 3), South Korea (7$^{\text{th}}$), Venezuela (8$^{\text{th}}$), South Africa (16$^{\text{th}}$), Colombia (18$^{\text{th}}$), Chile (19$^{\text{th}}$), and the Dominican Republic (20$^{\text{th}}$). These countries occur among the top 20 in our list reporting on music-related posts, but are not included in the list of top 20 countries for total tweet contributed.

3.2 Comparison of `Twitter` and `last.fm` Popularities

In a second set of evaluation experiments, we compared the set of popular artists extracted from `Twitter` posts with a reference set. Since traditional music charts, such as the "Billboard Hot 100" released weekly for the United States of America by the `Billboard Magazine`, are neither available on the level of individual cities, nor for all countries in the world, we gathered popularity data from `last.fm` as follows.

`last.fm` provides weekly artist charts for selected "metros". Taking as input the list of 790 locations gathered from `World Gazetteer`, we were able to extract artist charts

[8] `http://last.fm/api` (access: March 2010)
[9] `http://allmusic.com` (access: January 2011)
[10] `http://lucene.apache.org` (access: January 2011)

for 84 of the corresponding metropolises. In order to better understand the quality of the data on different temporal levels, various experimental settings were evaluated. The results are summarized in Table 1. First, we performed an exact *day-to-day* comparison experiment. To this end, for all pairs of days and locations for which information was available from both sources (`Twitter` and `last.fm`), we compared the extracted artist sets. Since `last.fm` provides popularity information only on the level of weeks, we interpolated this weekly information to individual days. The results of this day-to-day comparison experiment, averaged over all locations and days under consideration, are depicted in the second column of Table 1, labeled `D2D`. Taking the `last.fm` artist set as reference set, we calculated the following performance measures:

Precision at a specific day t and location l is defined as the fraction of artists found in `Twitter` messages posted at day t and location l that are also reported by `last.fm`'s chart function for l and t, among the total number of artists found in `Twitter` messages for day t and location l. Formally,

$$prec_{t,l} = \frac{\left|A^{tw}_{t,l} \cap A^{fm}_{t,l}\right|}{\left|A^{tw}_{t,l}\right|}$$

where $A^{tw}_{t,l}$ is the set of popular artists predicted by our approach for time t and location l, and $A^{fm}_{t,l}$ is the set of popular artists reported by `last.fm`.

Recall is defined as the percentage of `last.fm` artists for t and l that are also part of the artist set extracted from `Twitter` messages at t for l:

$$rec_{t,l} = \frac{\left|A^{tw}_{t,l} \cap A^{fm}_{t,l}\right|}{\left|A^{fm}_{t,l}\right|}$$

F_1-*measure* is the weighted harmonic mean of precision and recall [van Rijsbergen, 1979]:

$$F_1 = \frac{2 \cdot prec \cdot rec}{prec + rec}$$

Overlap is defined as the number of artists occurring in both sources divided by the maximum number of artists retrieved by either source (at t for l):

$$overlap_{t,l} = \frac{\left|A^{tw}_{t,l} \cap A^{fm}_{t,l}\right|}{\max\left(\left|A^{tw}_{t,l}\right|, \left|A^{fm}_{t,l}\right|\right)}$$

Alleviating the very strict matching requirement of the day-to-day experiment, we further performed *city-to-city* matching by aggregating all `Twitter` posts retrieved for each city (regardless of the date) and comparing them to the city's aggregated `last.fm` charts for the same period (the five months for which we gathered data). The precision on the city level is calculated based on artist sets A^{tw}_l and A^{fm}_l, irrespective of the date t:

$$prec_l = \frac{\left|A^{tw}_{l} \cap A^{fm}_{l}\right|}{\left|A^{tw}_{l}\right|}$$

The definition of recall, F_1-measure, and overlap updates analogously.

The results of this experiment can be found in the third column of Table 1, labeled `C2C`. It can be seen that the average recall increases substantially compared to the day-to-day setting, while the average precision remains almost the same. This can be explained by the disproportionately low number of artists covered by `last.fm` charts, compared to the number extracted from `Twitter`, for this granularity level. In fact, the average number of unique artists in `Twitter` posts exceeds the average number of unique artists covered by `last.fm` charts by a factor of five – cf. first two rows of Table 1. The considerable improvement of the `C2C` setting over the `D2D` setting might also be explained by a temporal lead or lag of the two data sources `last.fm` and `Twitter`, which is smoothed out when temporal aspects are ignored.

Further broadening the scope of matching yields the final experiment conducted. For this overall matching experiment, all extracted `Twitter` posts as well as all retrieved `last.fm` charts were aggregated, and the performance measures were only calculated on the resulting two artist sets. This setting can be thought of as a global popularity prediction. The precision for the overall matching experiment is calculated on artist sets A^{tw} and A^{fm}, irrespective of both date t and location l:

$$prec = \frac{\left|A^{tw} \cap A^{fm}\right|}{\left|A^{tw}\right|}$$

The definition of recall, F_1-measure, and overlap updates correspondingly.

The results of this overall matching experiment are given in the fourth column of Table 1. Please note that in the table average performance values are given for day-to-day matching (averaged over all locations and dates) and city-to-city matching (averaged over all locations), whereas total scores are given for the overall comparison experiment. Hence, for day-to-day matching, average precision is calculated as

$$prec = |T|^{-1} \cdot |L|^{-1} \cdot \sum_{l \in L} \sum_{t \in T} \frac{\left|A^{tw}_{t,l} \cap A^{fm}_{t,l}\right|}{\left|A^{tw}_{t,l}\right|},$$

whereas for city-to-city matching average precision is calculated as

$$prec = |L|^{-1} \cdot \sum_{l \in L} \frac{\left|A^{tw}_{l} \cap A^{fm}_{l}\right|}{\left|A^{tw}_{l}\right|}.$$

L denotes the set of locations, whereas T denotes the points in time (days) for which information is available. The other performance measures are calculated analogously.

Addressing the question if the quality of the popularity estimates is consistent over different cities, Figure 5 depicts the individual precision and recall values for all agglomerations for which `last.fm` provided corresponding data. The cities are sorted according to the F_1-measure. As it can be seen, the results vary strongly over different cities. The standard deviation of the precision values is $\sigma_{prec} = 0.0678$, that of the recall values equals $\sigma_{rec} = 0.2403$.

4 Conclusions and Outlook

We presented an analysis of music-related microblogging activity around the world and a simple popularity measure based on music artists' term frequencies in `Twitter` posts. Investigating the spatial distribution of music-related tweets revealed a considerable dominance of Asian countries (in terms of absolute number of posts) and of South American countries (in terms of number of posts relative to the number of inhabitants). The *location- and time-specific popularity measure* was evaluated in various experiments on different scales of granularity. On the level of individual days, the approach yielded modest precision and recall values, whereas remarkable recall could be achieved when aggregating the location-specific posts for all days under consideration.

Future work will be centered around exploiting the fine-grained day-level rankings. They could be used, for example, to illustrate changes in popularity around the world. For the application scenario of music chart prediction, the rankings could be used to complement traditional music charts, as they are generally biased towards actual music sales and also neither available on the city level, nor for all countries in the world. We will also experiment with data from domains other than music. For example, we are currently investigating different term weighting approaches to predict popularity of movies. Furthermore, it would also be interesting to analyze if certain popularity patterns can be clustered according to properties such as country, continent, or language group. Another direction for future work will be visualizing the derived popularity information. By applying time-series visualization techniques [Few, 2007], changes in popularity could be appealingly illustrated, for example via popularity "Flow Maps" [Phan *et al.*, 2005]. Reconsidering our main research focus on music information retrieval, artist popularity estimates on different geographical scopes and temporal points can help build personalized models of musical similarity and user preferences, which may ultimately yield to better personalized music services and applications, such as automatic playlist generators and music recommender systems.

Acknowledgments

This research is supported by the Austrian Science Funds (FWF): P22856-N23 and L511-N15.

References

[Aucouturier and Pachet, 2002] Jean-Julien Aucouturier and François Pachet. Scaling Up Music Playlist Generation. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2002)*, pages 105–108, Lausanne, Switzerland, August 2002.

[Casey *et al.*, 2008] Michael A. Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. Content-Based Music Information Retrieval: Current Directions and Future Challenges. *Proceedings of the IEEE*, 96:668–696, April 2008.

[Celma, 2008] Òscar Celma. *Music Recommendation and Discovery in the Long Tail*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2008.

[Cheng *et al.*, 2010] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. You Are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users. In *Proc. of the 19th ACM Int'l Conference on Information and Knowledge Management (CIKM)*, Oct 2010.

[Evans, 2010] Mark Evans. Exploring the Use of Twitter Around the World. http://blog.sysomos.com/2010/01/14/exploring-the-use-of-twitter-around-the-world (access: October 2010), 2010.

[Few, 2007] Stephen Few. Visualizing Change: An Innovation in Time-Series Analysis. *Visual Business Intelligence Newsletter*, September 2007.

[Göker and Myrhaug, 2002] Ayse Göker and Hans I Myrhaug. User Context and Personalisation. In *Proceedings of the 6th European Conference on Case Based Reasoning (ECCBR 2002): Workshop on Case Based Reasoning and Personalization*, Aberdeen, Scotland, September 2002.

[Grace *et al.*, 2008] Julia Grace, Daniel Gruhl, Kevin Haas, Meenakshi Nagarajan, Christine Robson, and Nachiketa Sahoo. Artist Ranking Through Analysis of On-line Community Comments. In *Proceedings of the 17th ACM International World Wide Web Conference (WWW 2008)*, Bejing, China, April 21–25 2008.

[Java *et al.*, 2007] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why We Twitter: Understanding Microblogging Usage and Communities. In *Proceedings of WebKDD/SNA-KDD*, 2007.

[Jon M. Kleinberg, 1999] Jon M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5), 1999.

[Kazeniac, 2009] Andy Kazeniac. Social Networks: Facebook Takes Over Top Spot, Twitter Climbs. http://blog.compete.com/2009/02/09/facebook-myspace-twitter-social-network (access: March 2010), 2009.

[Knees *et al.*, 2007] Peter Knees, Tim Pohle, Markus Schedl, and Gerhard Widmer. A Music Search Engine Built upon Audio-based and Web-based Similarity Measures. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*, Amsterdam, the Netherlands, July 23–27 2007.

[Koenigstein and Shavitt, 2009] Noam Koenigstein and Yuval Shavitt. Song Ranking Based on Piracy in Peer-to-Peer Networks. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, Kobe, Japan, October 2009.

Table 1: Summary of the experiments comparing `Twitter` and `last.fm` popularity rankings.

Property	**D2D**	**C2C**	**Overall**
Avg. number of artists in `Twitter` posts	21.97	410.49	2,490
Avg. number of artists in `last.fm` charts	37.49	79.94	1,534
Avg. precision on `last.fm` charts (%)	11.16	12.70	51.68
Avg. recall on `last.fm` charts (%)	6.36	51.80	83.90
Avg. F_1-measure on `last.fm` charts (%)	8.10	20.39	63.96
Avg. overlap between `Twitter` posts and `last.fm` charts (%)	4.43	11.05	51.68

[Kwak *et al.*, 2010] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, A Social Network or a News Media? In *Proc. of the 19th Int'l Conference on World Wide Web (WWW)*, Apr 2010.

[Lamere, 2008] Paul Lamere. Social Tagging and Music Information Retrieval. *Journal of New Music Research: From Genres to Tags – Music Information Retrieval in the Age of Social Tagging*, 37(2):101–114, 2008.

[Pachet and Roy, 2008] François Pachet and Pierre Roy. Hit Song Science is Not Yet a Science. In *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, Philadelphia, PA, USA, September 2008.

[Pampalk and Goto, 2007] Elias Pampalk and Masataka Goto. MusicSun: A New Approach to Artist Recommendation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, September 23–27 2007.

[Phan *et al.*, 2005] Doantam Phan, Ling Xiao, Ron Yeh, Pat Hanrahan, and Terry Winograd. Flow Map Layout. In *Proceedings of IEEE Information Visualization 2005 (InfoVis 2005)*, Minneapolis, Minnesota, USA, October 2005.

[Pohle *et al.*, 2007] Tim Pohle, Peter Knees, Markus Schedl, Elias Pampalk, and Gerhard Widmer. "Reinventing the Wheel": A Novel Approach to Music Player Interfaces. *IEEE Transactions on Multimedia*, 9:567–575, 2007.

[Ripeanu, 2001] Matei Ripeanu. Peer-to-Peer Architecture Case Study: Gnutella Network. In *Proceedings of IEEE Peer-to-Peer Computing*, 2001.

[Sakaki *et al.*, 2010] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In *Proc. of the 19th Int'l Conference on World Wide Web (WWW)*, Apr 2010.

[Schedl *et al.*, 2010] Markus Schedl, Tim Pohle, Noam Koenigstein, and Peter Knees. What's Hot? Estimating Country-Specific Artist Popularity. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, Utrecht, the Netherlands, August 2010.

[Schedl, 2008] Markus Schedl. *Automatically Extracting, Analyzing, and Visualizing Information on Music Artists from the World Wide Web*. PhD thesis, Johannes Kepler University Linz, Linz, Austria, 2008.

[Schedl, 2011] Markus Schedl. *Music Data Mining*, chapter Web- and Community-based Music Information Extraction. Taylor-Frances/CRC, 2011.

[Sharifi *et al.*, 2010] Beaux Sharifi, Mark-Anthony Hutton, and Jugal Kalita. Summarizing Microblogs Automatically. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT)*, Jun 2010.

[Teevan *et al.*, 2011] Jaime Teevan, Daniel Ramage, and Meredith Ringel Morris. #TwitterSearch: A Comparison of Microblog Search and Web Search. In *Proc. of the 4th ACM Int'l Conference on Web Search and Data Mining (WSDM)*, Hong Kong, China, Feb 2011.

[van Rijsbergen, 1979] C. J. van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, London, UK, 2nd edition, 1979.

[Zobel and Moffat, 2006] Justin Zobel and Alistair Moffat. Inverted Files for Text Search Engines. *ACM Computing Surveys*, 38:1–56, 2006.

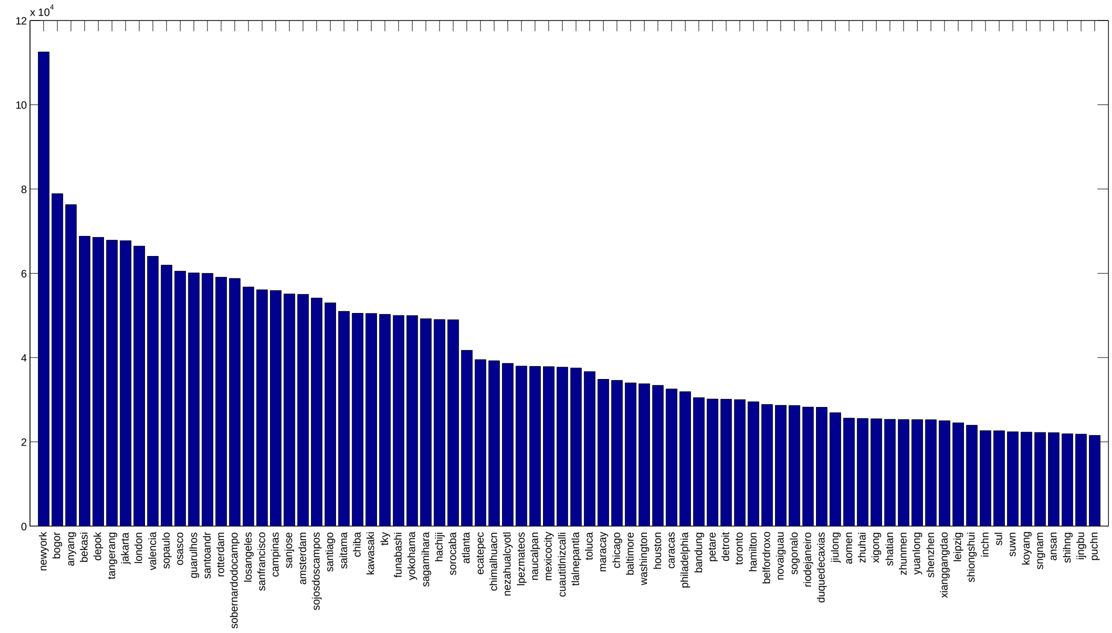

Figure 1: Cities with largest total number of #`nowplaying` posts (> 90th percentile).

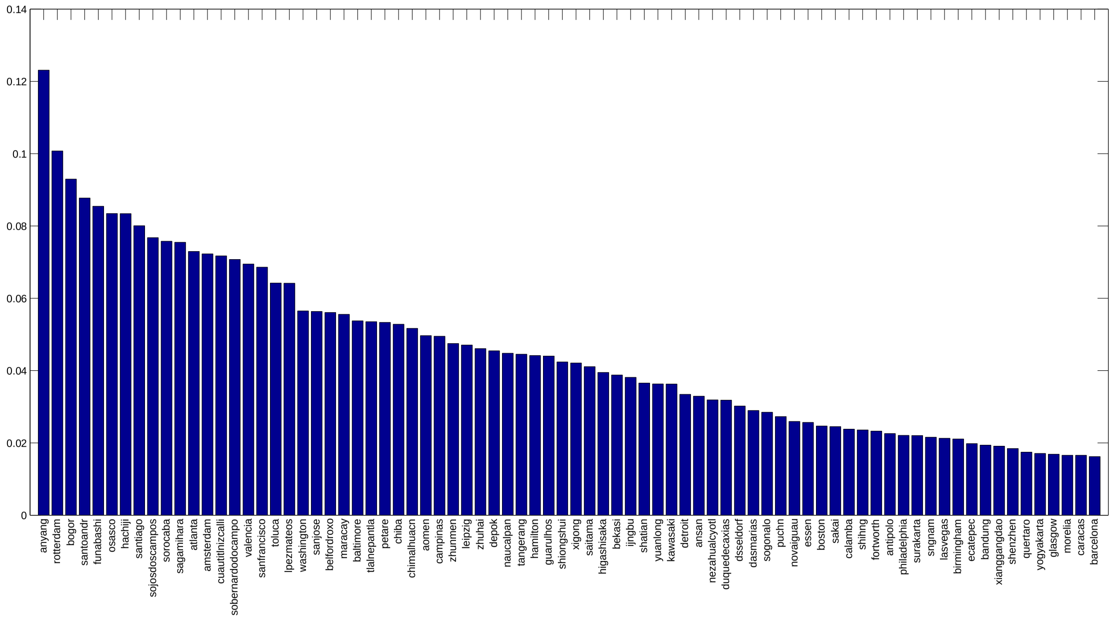

Figure 2: Cities with largest relative number of #`nowplaying` posts (> 90th percentile).

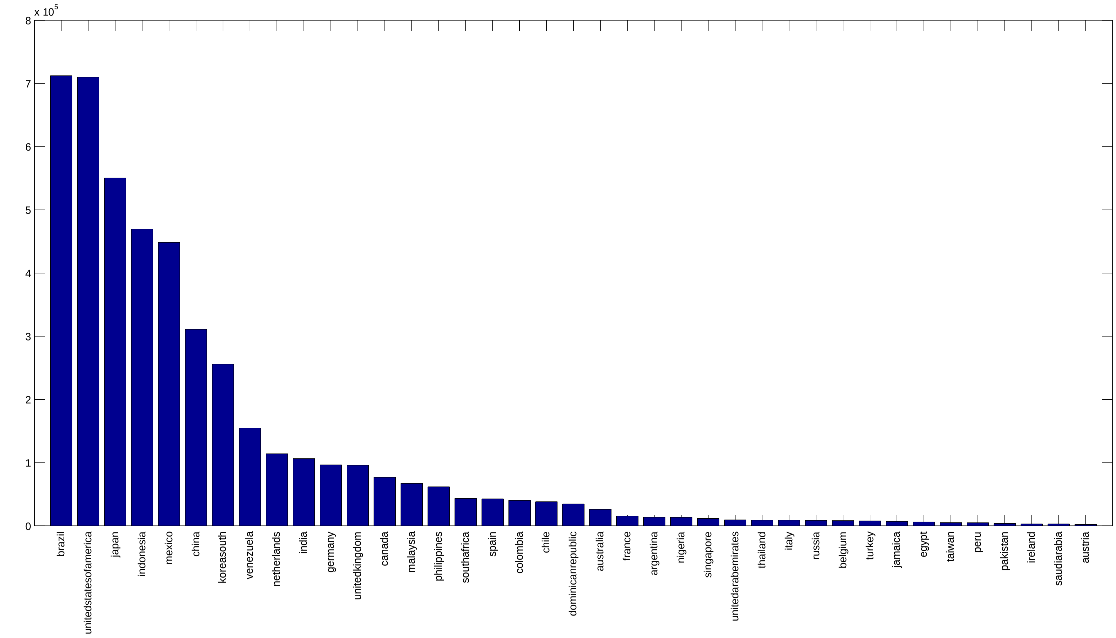

Figure 3: Countries with largest total number of `#nowplaying` posts (> 70th percentile).

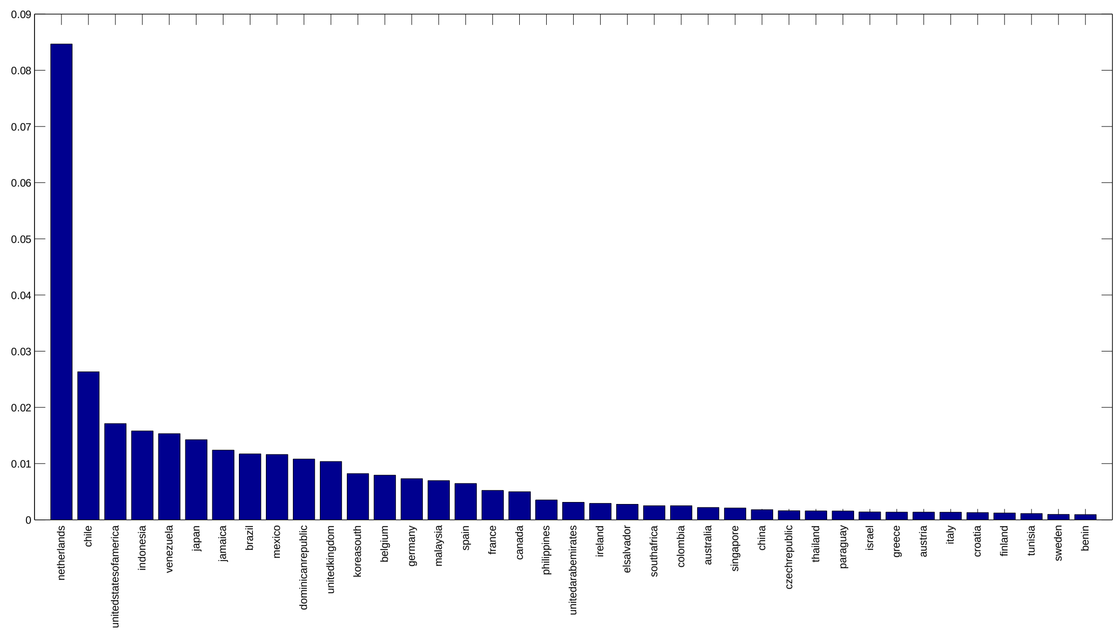

Figure 4: Countries with largest relative number of `#nowplaying` posts (> 70th percentile).

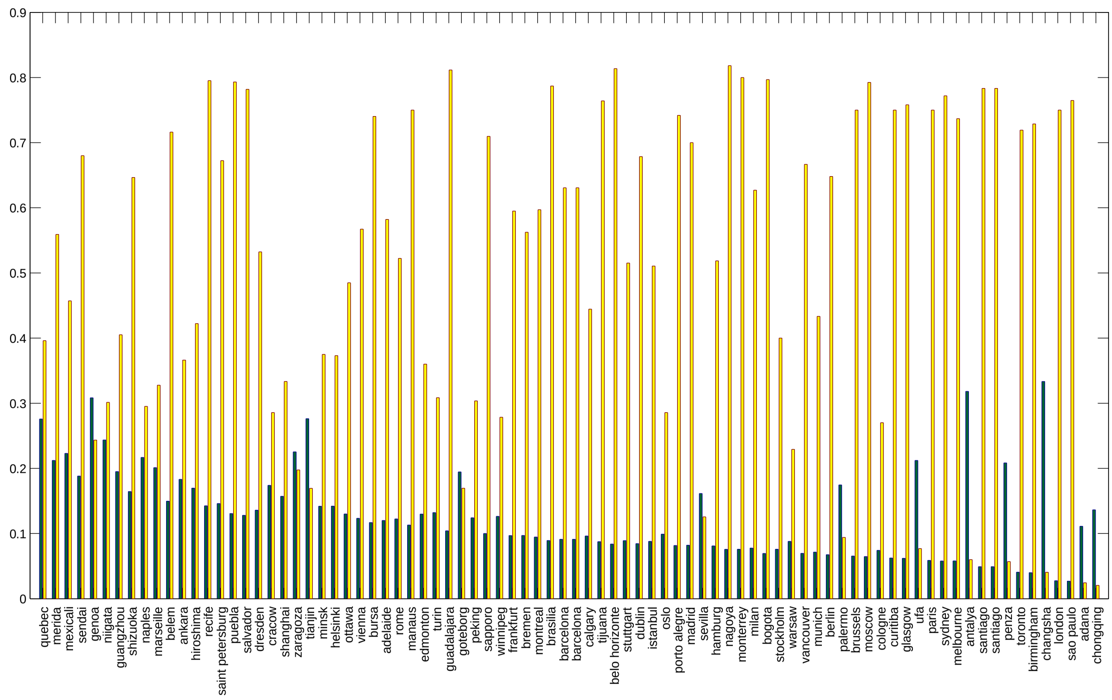

Figure 5: Precision (dark bars) and recall (bright bars) of the `Twitter`-based popularity estimation evaluated on `last.fm`, sorted by F_1-measure.