



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

**Markus Schedl**

**Department of Computational Perception  
Johannes Kepler University (JKU)  
Linz, Austria**

## Typical Application Scenarios of MIR

- personalized Web radio
- automatic playlist generation
- intelligent user interfaces

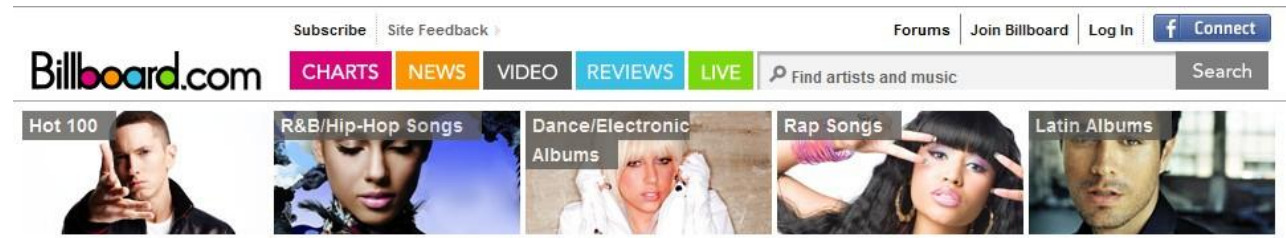




## Motivation

Goal: Determine **artist popularity** on the level of individual countries.

Although music charts do exist...



...they suffer from various shortcomings.

## Motivation (II)

Drawbacks of traditional charts, such as the "Billboard Hot 100":

- not available for all countries (or not publicly accessible)
- cover only certain channels of music distribution/consumption
- bias towards record sales / radio plays
- heavy distortions caused by (illegal) music sharing channels
- inhomogeneity between countries (in terms of channels) if data available

→ Look into microblogging activity to derive a measure of artist popularity



## Research Questions

**To which extent are microblogs used to communicate music listening activity? Are there differences between different countries of the world?**

**Is it possible to derive a spatio-temporal music popularity measure from microblogging activity?**



## Data Acquisition

twitter



*preprocessing:*

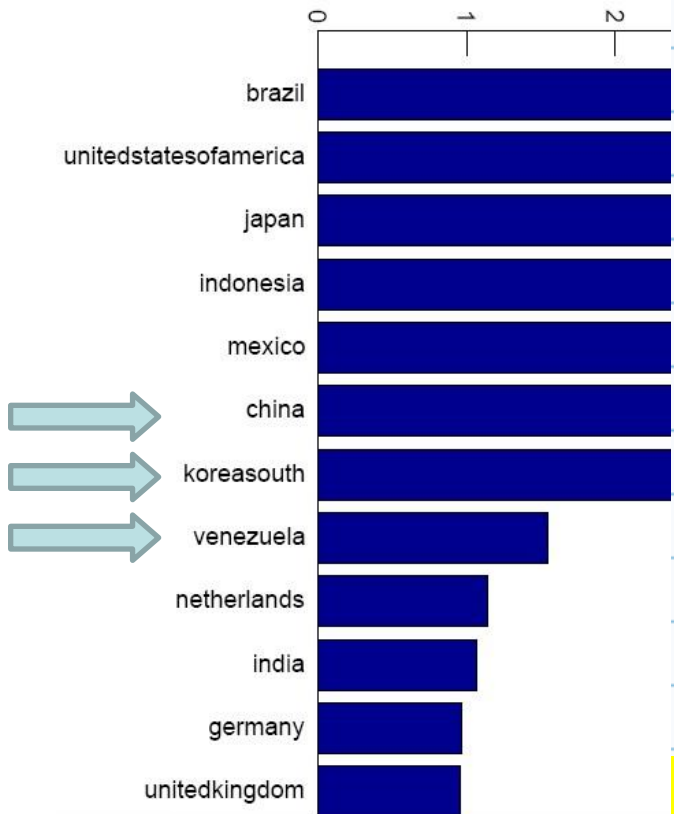
extract major cities of the world (>500,000 inhabitants) from *World Gazetteer* and geolocate them → [city, country, long., lat.]

- search for *twitter* posts including *#nowplaying* around the cities' coordinates (up to 1,500 posts per city)
- aggregate posts for each country
- calculate term frequency of each artist name within a country's posts (artist names extracted from most popular *last.fm* artists)

data acquisition took place from May to September 2010

## Statistics and Evaluation

Distribution of music-related tweets

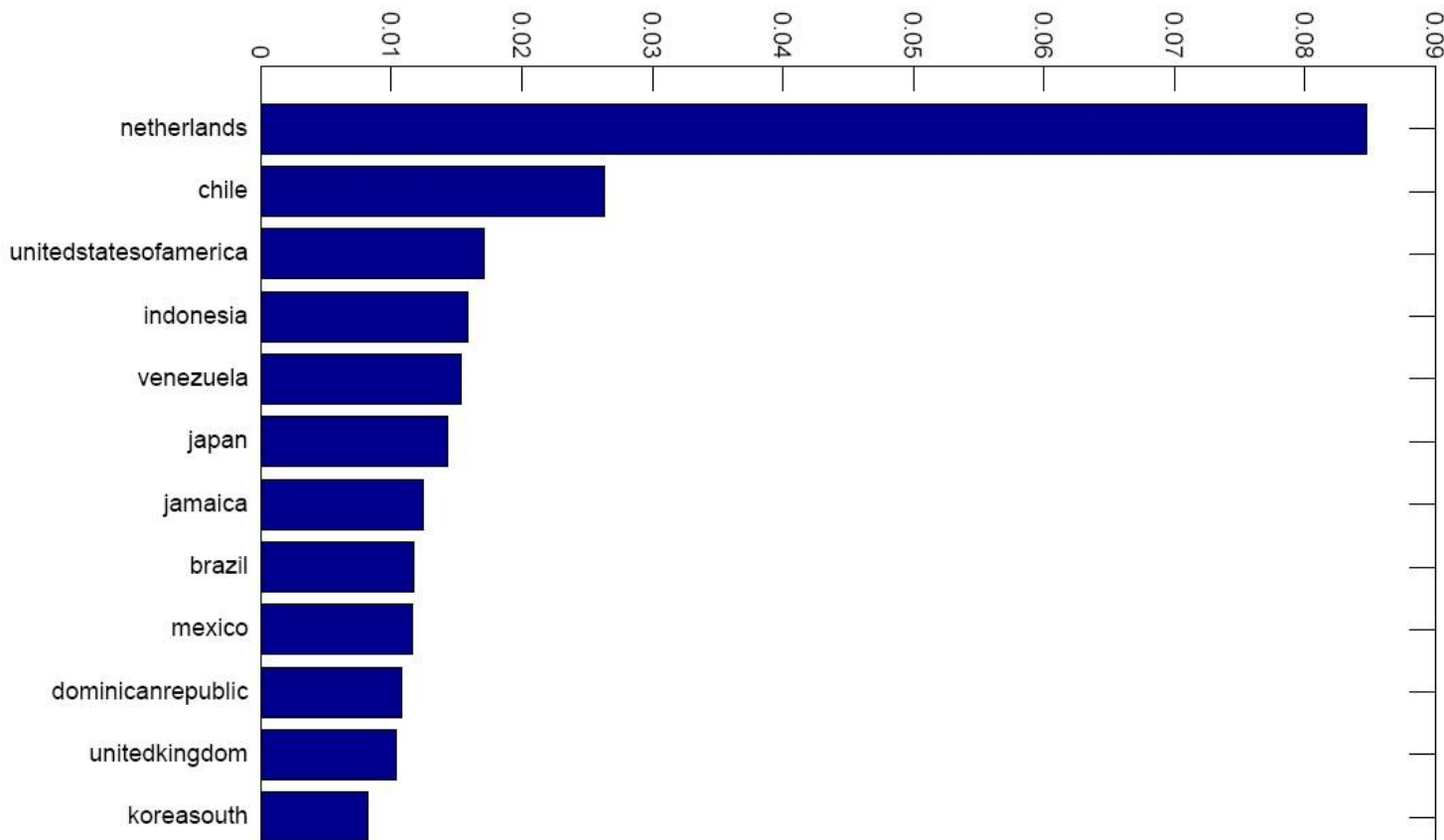


Country	% total tweet contributed
USA	56.59
UK	8.09
Brazil	6.73
Canada	4.36
Australia	2.63
Indonesia	2.34
Germany	1.58
Japan	1.47
Netherlands	1.10
India	0.97
Mexico	0.90
Singapore	0.88

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: July 2011)

## Statistics and Evaluation

Distribution of music-related tweets around the world (relative to no. of inhabitants)





## Statistics and Evaluation (II)

Comparison between *twitter* and *last.fm* data:



ground truth:

weekly charts by *last.fm* for “metros” (for only 84 out of 790 data available)

analysis on different scopes: day-to-day (D2D), city-to-city (C2C), overall

quality measures: precision, recall, F1 measure, overlap

$$prec_{t,l} = \frac{|A_{t,l}^{tw} \cap A_{t,l}^{fm}|}{|A_{t,l}^{tw}|}$$

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

$$rec_{t,l} = \frac{|A_{t,l}^{tw} \cap A_{t,l}^{fm}|}{|A_{t,l}^{fm}|}$$

$$overlap_{t,l} = \frac{|A_{t,l}^{tw} \cap A_{t,l}^{fm}|}{\max(|A_{t,l}^{tw}|, |A_{t,l}^{fm}|)}$$

## Statistics and Evaluation (II)

Comparison between *twitter* and *last.fm* data:

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

- almost no temporal correspondence between predictions
- ignoring time, modest recall between countries

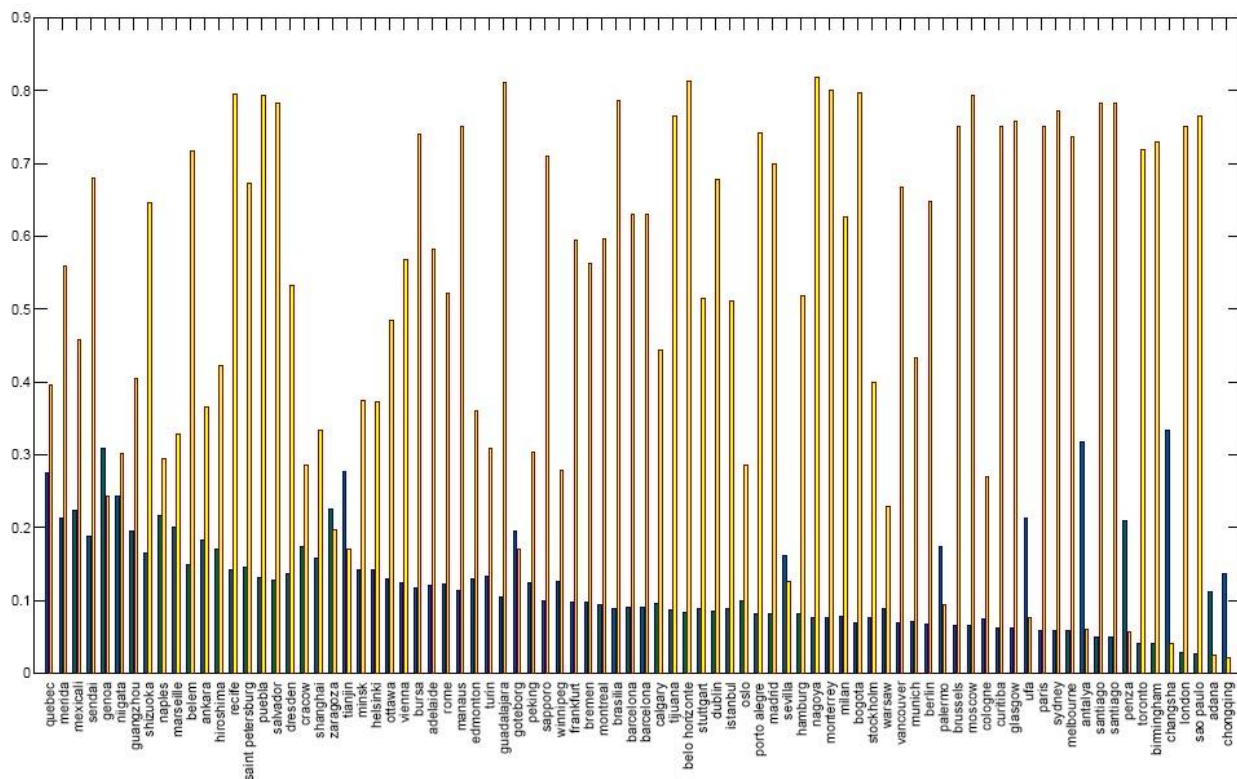
## Statistics and Evaluation (II)

Comparison between *twitter* and *last.fm* data:

Differences between cities?

$$\sigma_{\text{prec}} = 6.78$$

$$\sigma_{\text{rec}} = 24.03$$





## Conclusions

- *twitter* used to different extent to communicate music listening activity (although corresponding to overall *twitter* usage)
- more data available on *twitter* than what is offered by *last.fm*'s API
- little overlap between different data sources

Future Work  
**last.fm**



- hybrid approaches that combine different data sources
- methods to account for different quantity and quality of information
- charts for subgroups of artists (e.g., different genres)
- charts on track level

Google



overall overlap measure ov

las	1.00	0.57	0.51	0.54	0.53
p2p	0.57	1.00	0.53	0.67	0.58
exa	0.51	0.53	1.00	0.53	0.51
gog	0.54	0.67	0.53	1.00	0.56
twi	0.53	0.58	0.51	0.56	1.00
	las	p2p	exa	gog	twi



## Thank you !

