Depict or Discern ? Fingerprinting Musical Taste

Kristina Matrosova, Manuel Moussallam, Thomas Louail, Olivier Bodini

DEPICT OR DISCERN? FINGERPRINTING MUSICAL TASTE

K. Matrosova [1, 2], M. Moussallam [3], T. Louail [1], O. Bodini [2]

1. Géographie-cités, CNRS, France 2. LIPN, USPN, France 3. Deezer, France

In this work, we try to build a musical taste fingerprint based on declared preferences of users from a music streaming service.
We propose two distinct methods, that turn out to be conflicting: one that captures the uniqueness, and one that pictures the very essence of one's preferences.

Background

Research on musical taste spans over a large body of scientific disciplines. From a sociological standpoint, musical taste has been long identified as an important, self-claimed, differentiating feature among individuals (*Bourdieu, 1987*). Psychological studies have been investigating positive correlations between musical preferences and personality traits (*North, 2010*). More recently, the concept has been used in the music recommender system literature (*Schedl, 2015*), as the distinctive part of the musical space from which a user is likely to enjoy a recommendation. In its general understanding, musical taste is an individual's set of musical preferences. In that sense, it is a highly personal trait, that uniquely characterise each and every one of us. The purpose of this work is to explore different representations of musical taste through preferences declared on a music streaming service.

Dataset

- Collected from the **Deezer streaming app**
- **1M users** sampled randomly
- Users' **favourite items (artists and songs)**
- A **genre tag** is associated to most items
- For artists - an **embedding** and a list of **similar artists** are pre-computed based on co-listening

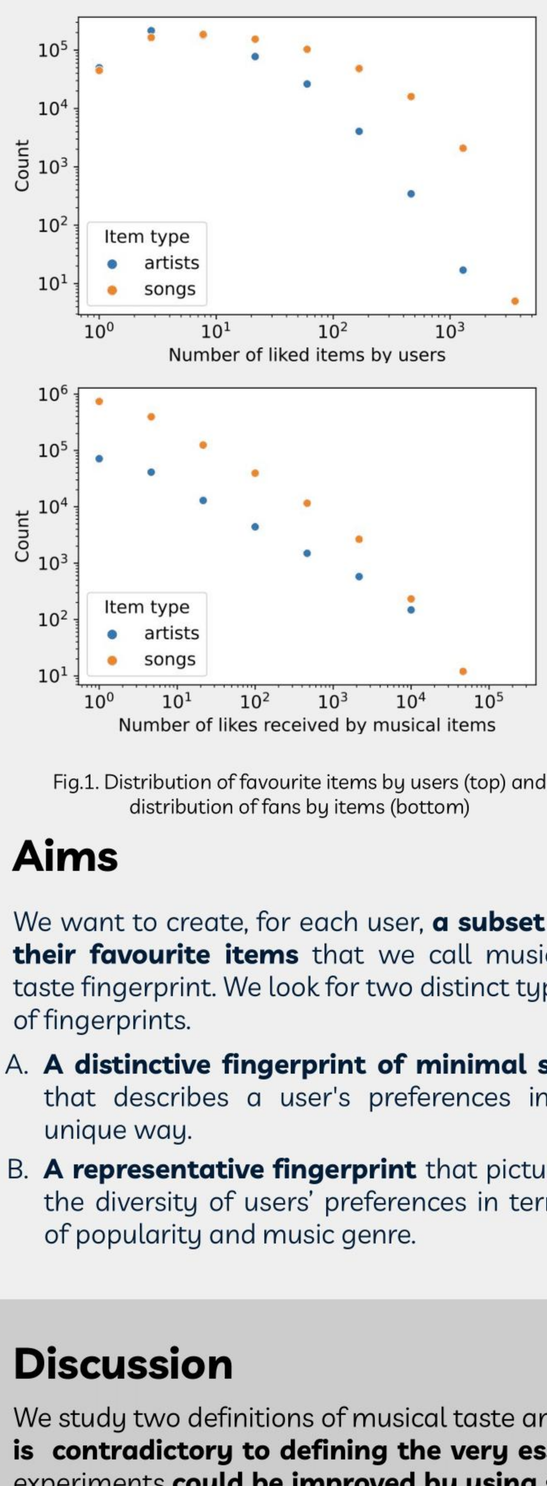

Fig.1. Distribution of favourite items by users (top) and distribution of fans by items (bottom)

Aims

We want to create, for each user, **a subset of their favourite items** that we call musical taste fingerprint. We look for two distinct types of fingerprints.

A. **A distinctive fingerprint of minimal size** that describes a user's preferences in a unique way.
B. **A representative fingerprint** that pictures the diversity of users' preferences in terms of popularity and music genre.

A. Unique fingerprints

- **Uniqueness**. We want to describe each user with a unique set of songs. Like an actual fingerprint, this set of items can not be the same for several users.
- **Non-inclusiveness**. Stronger than uniqueness, non inclusiveness guarantees that a fingerprint of one user is not included in the favourite items of another user. This means that if a user's fingerprint is composed of song a and b, this user is the only one in the dataset to like both song a and b.
- **Minimal size**. We want to minimize the size of the fingerprints to show how just a few likes can reflect individuality of one's taste and differentiate from others.

Fingerprinting method

Let U be a set of users. For each user u in U we sample a fingerprint $F(u)$ from u's favourite items $V(u)$ by:

1. Appending to $F(u)$ the least co-liked item i in $V(u)$
2. Removing all users that have liked i except u
3. Re-iterate while u is not the only user left

Results

In a set of 1M users:

- ➔ **30%** of the users can be identified in a unique way by their favourite **artists**.
- ➔ **70%** of the users can be identified in a unique way by their favourite **songs**.
- ➔ A user is identifiable with an average of **1.8 songs and 2.3 artists**.

The genre composition of the fingerprints differs from the one in the users favourite items (Figure 4), which brings up the question: **are these fingerprints representative of the users musical taste?**

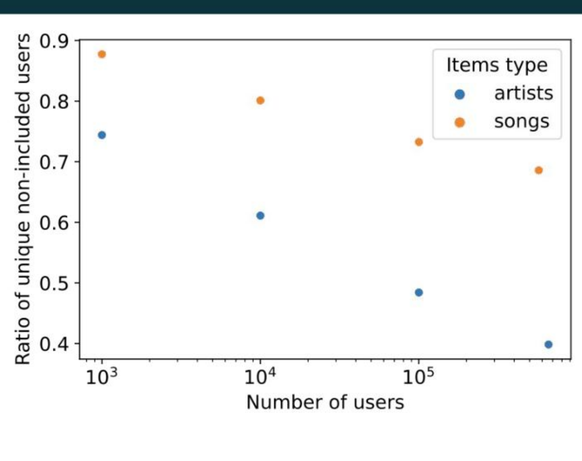

Fig. 2. Number of identifiable users in different size samples. For instance, in a sample of 10^4 users, an average of 60% can be identified with their favourite artists.

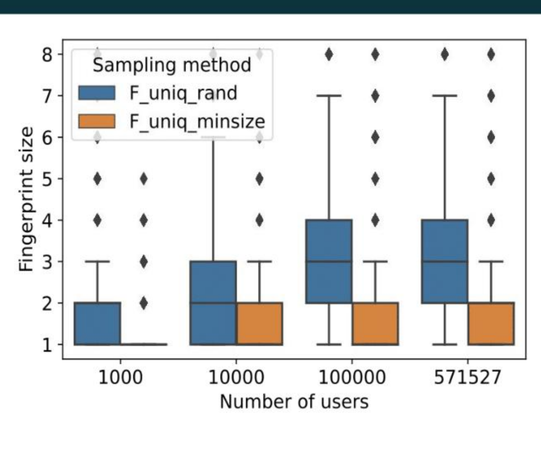

Fig. 3. Size of fingerprints based on favourite songs for different size user samples using unique random vs unique minimum size sampling.

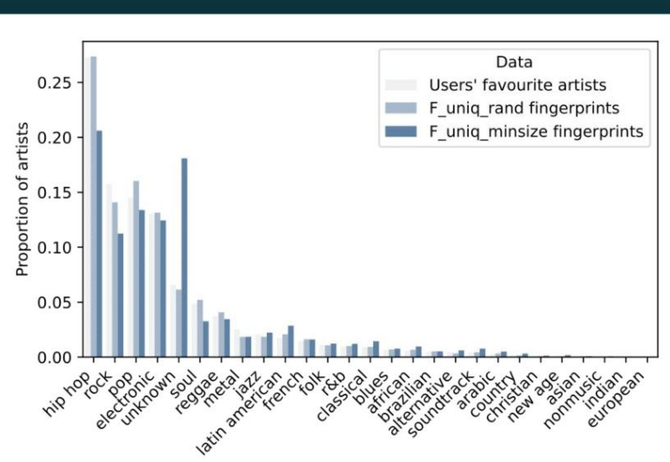

Fig. 4. Distribution of genres in unique minimum size fingerprints, unique random fingerprints and overall users' favourite artists.

B. Representative fingerprints

We want to find, for each user u, a subset of $V(u)$ that is representative of their musical taste. We consider that a subset $F(u)$ is representative if there exists a method $F\bigstar$ such as $\forall u \in U, V(u) \approx F\bigstar(F(u))$. In other words, we consider that a fingerprint is representative if there is a method that allows to recover the initial set of items from it.

Recovering function $F\bigstar$

$F\bigstar$ takes as input: a fingerprint $F(u)$, a weight w_i associated to each item i in $F(u)$, representing the number of items we need to recover from i and a list of 150 similar artists for each artist. As an output, $F\bigstar$ returns a set of predicted items $P(u)$ by simply taking, for each item i in $F(u)$, the w_i closest neighbours of i from the list of i's similar artists. Recovery score is then the ratio of correctly predicted items.

Fingerprinting method

Experiments have shown that the **minimal size unique fingerprints** described in the previous section **are not representative of the users' preferences** (Figure 5), thus we propose another method to sample representative fingerprints. As mentioned in aims, we have, for each artist, an embedding of size 128. We propose to split $V(u)$ in k clusters using the **k-medoids** algorithm, and use the medoids as representative artists of each cluster to build our fingerprint.

Results

- ➔ Average recovery is **15%** of the **users' favourite artists**, against 9% with a random sampling (Figure 6).
- ➔ Average recovery is **50%** and up to **100%** of the **genre ratio** in the users' preferences (Figure 7).

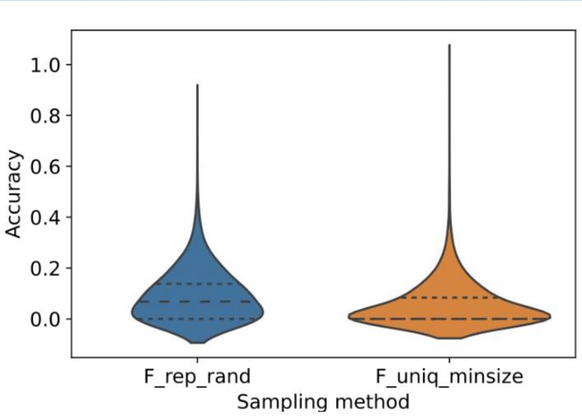

Fig. 5. Recovery score based on randomly sampled fingerprints vs unique minimum size fingerprints.

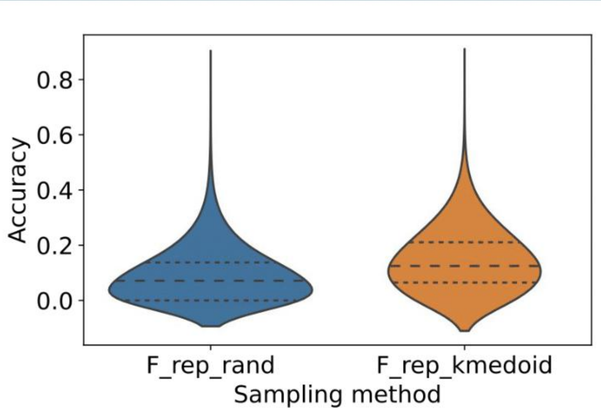

Fig. 6. Recovery score based on randomly sampled fingerprints vs k-medoids fingerprints.

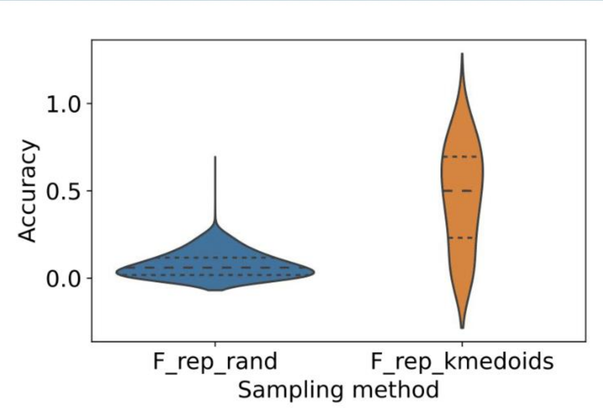

Fig. 7. Genre ratio recovery score based on randomly sampled fingerprints vs k-medoids fingerprints.

Discussion

We study two definitions of musical taste and propose a method to illustrate these definitions using explicit preferences collected from a music streaming service. We show that **capturing the uniqueness is contradictory to defining the very essence** of one's likings. We can underline that **the definition of musical taste should be more specific** when using it in scientific work. The accuracy of our experiments **could be improved by using streaming data** instead, or in addition to explicitly declared preferences, to counterbalance the lack of information about the users that do not usually use the like button. Finally, we show that **70% of the users can be easily identified** by a couple of liked songs, which brings up the question of **anonymity.** In a situation when the user's likes are publicly available on the platform, sharing personal data combined with information about favourite items can be compromising even if the users' IDs are hashed.

References

Bourdieu, P. (1987). Distinction: A social critique of the judgement of taste. Harvard university press. 33-38.
North, A. C. (2010). Individual differences in musical taste. The American journal of psychology, 123(2), 199-208.
Schedl, M., Knees, P., McFee, B., Bogdanov, D., & Kaminskas, M. (2015). Music recommender systems. In Recommender systems handbook (pp. 453-492). Springer, Boston, MA.