Correlating Big Five Primary Personality Dimensions with Musical Preferences

Luiz Delando Santos Moreira Júnior[1]*, Guilherme Ribeiro Eulálio Cabral[2] Giordano Ribeiro Eulálio Cabral[2]

[1]Centro de Informática – Universidade Federal de Pernambuco
Av. Jorn. Aníbal Fernandes, s/n - Cidade Universitária, Recife - PE, 50740-560

[2]Univasf Universidade Federal do Vale do São Francisco)
Av. José de Sá Maniçoba, s/n - Centro, Petrolina - PE, 56304-205

ldsmj@cin.ufpe.br, grecabral@gmail.com, grec@cin.ufpe.br

Abstract. *According to [1] in 2019 an estimated 2.95 billion people were using social media worldwide, with a projection of 3.43 billion in 2023. Social networks increasingly play a fundamental role in the way of expressing ourselves and relating to others, in which it is directly linked to our personality. Several studies relate a person's personality to the most diverse areas, such as predicting political orientation, classifying gender, or even measuring the level of success at work. [2] proposes an algorithm that can allegedly estimate the personality of a Facebook user better than his partner, based on his digital footprint, thanks to AI techniques and the massive number of interactions that users perform on their social media. The focus of this work is to relate the "Big Five" personality theory with musical preferences. Since social media posts register moments and impressions of users' lives, and since music streaming platforms provide material about musical taste [3], we used tools to extract the big five primary personality dimensions from Facebook and audio features from the Spotify Web API, and looked for correlations between musical preferences and personality. Results show some correlations and anticorrelations found, pointing to further research with a broader range of features.*

1. Introduction

The massification of social networks are in line with the growing interest of researchers in relating their use with personality theories. One in particular, known as the Big Five [4], has been showing interesting results about the correlations found between personality and musical preferences. For example, according to [3] research, positive correlations were identified between blues, jazz, classical, and folk music with the openness personality trait. In the work of [5], another correlation was found between mass music and the extroversion factor. And the two surveys mentioned using manual data collection as methodology, applying questionnaires to determine which values of each user's personality trait to use, and relating them to information about musical preferences also extracted via questionnaires. This way of collecting data motivated me to look for ways to obtain information in an automatic and scalable manner.

According to [6] it is possible to estimate, only based on digital footprints in social networks, an individual's personality traits and characteristics, such as political orientation and probability of success at work. In this research, we chose to work with music, as it has a prominent place in our lives, according to [3] in a survey conducted by him, music is the most cited activity when the interviewee is questioned for his hobbies and also that musical preferences are considered more significant than food preference.

We built a platform so that it was possible to collect information from likes and Facebook posts through the API to estimate the Big Five personality traits and the information of musical preferences is retrieved from the APIs made available by Spotify.

2. Big Five Personality Traits

According to [7] the Big Five personality traits theory has been growing in psychology as a way of defining the individual's personality, using the most diverse methods. Big Five model had its origin in a series of researches on personality, coming from other theories and personality traits. [8], for a general review, please see [9]. We chose to use this theory as a basis for its acceptance, but also for its simplicity and the availability of digital tools.

As stated by [5], on Big Five there are 5 major traits that characterize human personality. The first of them is openness. According to [10] people who get a high score tend to be creative people and receptive to new ideas and (as the name implies) are more open to experiencing the new. As [9] explains, these individuals are usually referred to as artistic, curious, and imaginative, whereas those with low scores tend to be the opposite, that is, people with conventional thinking, and with a lesser scale of interest.

The second major trait is conscientiousness, according to [10] this factor brings together characteristics that bring the idea of organization, responsibility, and honesty. Individuals who score high on this factor tend to be trustworthy and dedicated workers, those who score low tend to be more impulsive and relaxed. These people can be described as efficient, organized, and responsible [9].

The third major trait is extraversion, which represents traits linked to sociability, as [8] well-defined extraversion is the amount and intensity of interactions between people and activity level. Individuals who score high on this factor tend to be people who prefer to be in the presence of other people, whereas those who have a low score tend to be more reserved. Second [9] adjectives

*Supported by CNPq.

that qualify these people are energetic, communicative, active.

The fourth major trait is agreeableness, according to [10], it can also mean the level of socialization and characterizes the individual's tendency to be socially pleasant and warm. People who score high tend to always seek the best in everyone, some adjectives that represent him are generous, friendly and accommodating [9]

The fifth and final trait is neuroticism, according to [10], this factor represents personality traits that involve positive and negative affect, anxiety, and emotional stability. Those who score high tend to experience negative experiences more intensely. While those with low scores are more emotionally stable. Adjectives for this factor are anxious, unstable, concerned [9].

2.1. Predicting BigFive Personality from digital footprints

The digital footprint is a term used to track the steps that the individual takes on the network, that is, all actions executed by the user using a device connected to the internet, from computers, smartphones and embedded devices [11]. According to [12] these brands are divided into two types, the first is the passive footprint, which represents the actions they perform involuntarily, for example, the search history, or the time that remains on the determined website, records the location and interactions with network infrastructure and we also have a qualified active footprint in this work. It is blocked when we interact directly, usually used when we want to change samples blocked or posted or liked on Facebook, Twitter, and Instagram on the internet.

In this study, we used a tool developed by the Center for Psychometry at the University of Cambridge, which was developed from a rich set of data, without which model was created based on more than 6 million social network profiles. Its objective is to personalize the Internet, transform the data of the digital activity tracked, for example, Facebook likes IDs, Publications written on Facebook, tweets and texts written by the user such as: emails and personal diaries that they can use as input. This information is sent to the service that returns the individual user profile that is formed by the psychological profile that provides information on Big Five personality, intelligence, life satisfaction, political ideology, religious views, sexuality and professions while the demographic profile used by age, gender and the status of your relationship.

As shown by [6], records of user behavior are easily accessible, for example with a list of Facebook Likes, it is possible to automatically extract some personal attributes. The model created is capable of correctly predicting the individual's origin in 95% of cases, whether it is African American or Caucasian and in 93% of cases it correctly classifies between male and female genders, these are cases of dichotomous variables that have a high hit rate, but there are also numerical variables that showed less precision, but this fact does not rule out the existence of correlations that can be verified in obtaining age with a coefficient of r = 0.75, in addition to personality traits as openness r = 0.43, extroversion r = 0.40 the other personality traits obtained a slightly lower precision r = 0.17 to 0.30.

Even with the accuracy not reaching the values found in the dichotomous variables, studies show that personality judgments made by computer can be more accurate than those made by humans, in a survey of more than 86,000 people who answered personality questionnaires of 100 items from your friends on Facebook demonstrated that the evaluation performed by a computer is more accurate than that performed by your friends. The computer got value for r = 0.56 while friends got a coefficient of r = 0.49. The results presented show that the average evaluation accuracy of the computer grows as the number of likes increases, in trained models the computer needs 10, 70, 150 and 300 likes to surpass the average evaluation of human judges, respectively, as colleague co-worker, roommate and family member or a spouse. This curve can be seen in figure 2 [2].

3. Platformzation of Digital Music

In the early days of the MIR area, studies were almost limited to specialists in the area, where one of the first challenges was the cocktail party. And with the evolution of the area, tools emerged that began to encapsulate tasks such as MFCC, HMM, and Machine Learning techniques, offering scientists a level of abstraction in the development of their studies.[13]

Thinking of a Platformzation of Digital Music scenario, companies like Echo Nest are dedicated to offering a musical intelligence service to students, researchers and people interested in the field. based on more than 35 million of songs where it can automatically extract various information from songs that are divided into two subcategories: One of defined by musicologists which are: tempo, note, tone, signature of tempo volume, among others [14] . And Cultural context features such as textit Energy, Liveness, Speechiness, Acousticness, Danceability and Valence.

In this study, we chose to use cultural characteristics directly related to the way the listener interprets music, in addition to the listener's ease of interpretation on how their personality can interfere in each one. In addition to the accessibility provided by the Spotify Web API. Table 1 shows a list of the characteristics used in this experiment.

4. materials and methods

In this section, we will explain the profile of the research participants and the procedures for developing the experiments.

The research started with the development of the PersonalityMusic-Crawler capable of collecting data from Facebook users invited to test the tool. and sending it to the API Apply Magic Sauce for analysis and then collect data from Spotify's music preferences and store all this data in a non-relational database. Based on this information,

Features Name	Type	Description
Acousticness	float	A 0.0 to 1.0 confidence measure if the range is acoustic. 1.0 represents high confidence that the track is acoustic.
Danceability	float	Describes how suitable a song is for dancing based on combination of musical elements.
Energy	float	It is a measure of 0.0 to 1.0 and represents a measure perceptual intensity and activity.
Instrumentalness	float	Predict whether a track does not contain vocals. The sounds "Ooh" and "aah" are treated as instrumental in this context.
Liveness	float	Detect an audience presence in the recording. A value above 0.8 provides a strong probability that the track is live.
Tempo	float	The total estimated time of a track in beats per minute (BPM). In musical terminology, Tempo is the speed or rhythm of a given piece and derives directly from the average beat duration.
Valence	float	A measurement from 0.0 to 1.0 describing the musical positivity transmitted by a track. Tracks with high valence sound more positive.

Table 1: Audio features from Spotify Web API

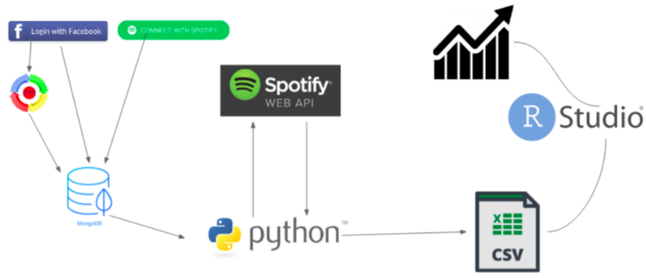

Figure 1: Architecture overview

a python script was created to collect the audio characteristics provided by the Spotify Web API and consolidate everything in a CSV spreadsheet so that it was possible to perform the analyzes.

4.1. Participants

About 65 participants were invited to participate in the research as an application tester. 60 records were obtained from the database, 13 of which were incomplete because users do not have enough data to generate any type of recommendation.

On average, each participant had 32 likes to be analyzed, but, by checking the minimum of likes in the sample, it was possible to verify that there were users with 5 valid likes for analysis, which generated a low degree of assertiveness in the classification of the users' personality. As a way of improving, we used only users who had a minimum of 10 valid likes following the example from [2]. In this paper, with 10 selected likes it is already possible to determine the user's personality better than a colleague of work would judge. Another cutoff point was the amount of music that we managed to recover from Spotify for each user. We chose to select only users with at least 100 songs returned by Spotify. This metric was used to identify users who used the Spotify service most often. After these limitations in the database, a total of 31 users were obtained, of which 16 were female and 15 male, aged between 18 and 30 years.

4.2. Instruments

The data was collected by PersonalityMusic-Crawler, which aimed to orchestrate all the following services AMS-Apply Magic Sauce, Spotify Web API, and Mongo DB, which is a non-relational database, chosen for simply allowing new information to be added in the future.

4.2.1. AMS - Apply Magic Sauce

This service is available free of charge for academic or non-profit purposes. We used them to estimate the personality traits from the likes of Facebook, where each user has their Json or a personality element composed of 5 features: BIG5_Openness, BIG5_Conscientiousness, BIG5_Extraversion, BIG5_Agreeableness, BIG5_Neuroticism, and that information was used for the analyzes. Each feature represented as a float ranges from 0.0 to 1.0 and shows personality traits.

4.2.2. Spotify Web API

To retrieve the user music preferences, we use the method "Get a User's Top Artists and Tracks" to collect 100 songs with which the user has more affinity. This request only returns the metadata of the song, such as name, album, band, and popularity. To collect the rest of the necessary information, we used the endpoint "Get Audio Features" service for all tracks that returned the characteristics that can be seen in table 1.

After retrieving the information, the same script was responsible for extracting the average of these characteristics for each user and then exporting that information together with the personality information to a CSV.

The last instrument used was RStudio, had the CSV containing the entire sample with the average of the audio characteristics and personality traits, Pearson's correlation techniques were applied together with a linear regression of the points that obtained the greatest relevance.

4.3. Procedure

In the first contact with the participants, we presented them with the objective of the research, and with his approval, we inserted them as testers of PersonalityMusic-Crawler, to allow Facebook to authorize the recovery of their posts and likes. When the user accesses the application, a short text was available explaining the search.

After the user's login, the application started the information collection process, retrieving all the posts and likes that the user posted on his timeline, sending this data to AMS to be processed, returning the values of the user's personality.

Then he logs in to Spotify and after collecting his music and storing it. A script retrieves all users in the database. Selecting the 100 songs of each user for analysis of the Spotify service.

After recovering the characteristics of each song, the average of these songs is calculated making a unique value that represents the user. So that we could look for correlations between personality traits and musical preferences. This information was processed and stored in a .CSV file along with gender and personality information. With structured data, we used the R statistics software to perform the correlation and regression calculations.

5. Results and Discussion

In this section we will present the results obtained from the statistical analysis between the Big Five factors and the user's musical preferences and raise discussions about the main associations found.

Through the Pearson correlation, some variations were found with the p-value less than 0.05. They are Extroversion vs Danceability, Overture vs Valencia, Openness vs Danceability, and Opening vs Danceability. In table 2, we have a summary of all the association found, where a * marks a p-value less than 0.05 %.

5.0.1. Extraversion vs Popularity

It is worth mentioning a specific case, between Extroversion and Popularity, which obtained marginally positive correlations, since its p-value was very close to 0.05% for an r = 0.35; p-value = 0.05292.

We had a $R^{2\,2}$ = 0.09 we can see in figure 2 the linear regression model extraversionVSpopularity.

$$P = 43.11 + 23.538 * E \tag{1}$$

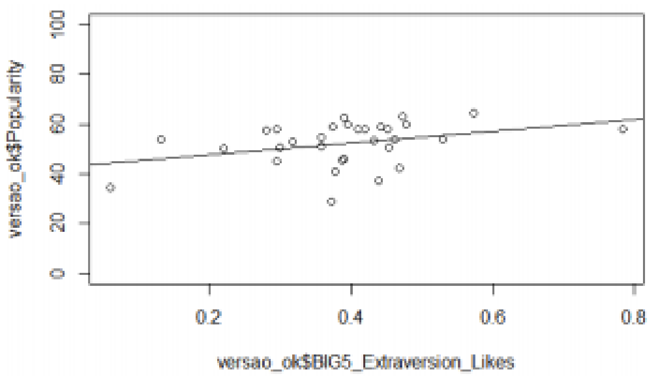

Figure 2: Linear regression - Extraversion vs Popularity

5.0.2. Openness vs Valence

It's found a negative correlation with r = -0.42; p-value = 0.019. The value of $R^{2\,2}$ = 0.147 we can see in figure 3 the linear regression model Openness vs Valence.

$$V = 0.59242 - 0.25512 * O \tag{2}$$

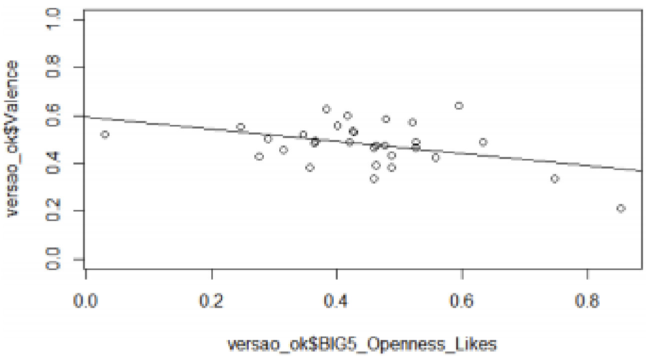

Figure 3: Linear regression - Openness vs Valence

5.0.3. Extraversion vs Danceability

It's Found a positive correlation with r = 0.40 p-value = 0.02. The value of $R^{2\,2}$ = 0.13 was obtained. We can see in figure 4, the linear regression model Extraversion vs Danceability.

$$D = 0.45114 - 0.22402 * E \tag{3}$$

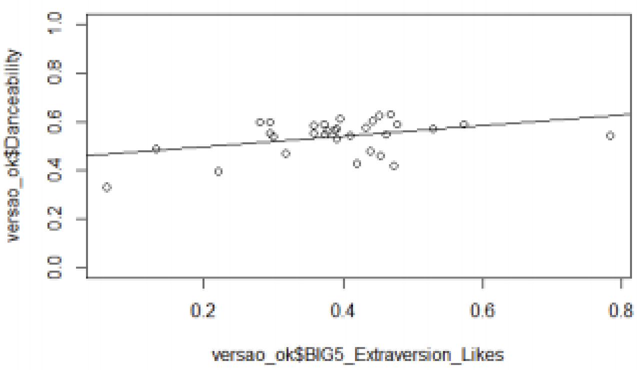

Figure 4: Linear regression - Extraversion vs Danceability

5.0.4. Openness vs Danceability

It's found a negative correlation with r = -0.38; p-value = 0.37. The $R^{2 2}$ = 0.11. We can see in figure 5, the linear regression model Openness vs Danceability

$$D = 0.61902 - 0.18053 * O \tag{4}$$

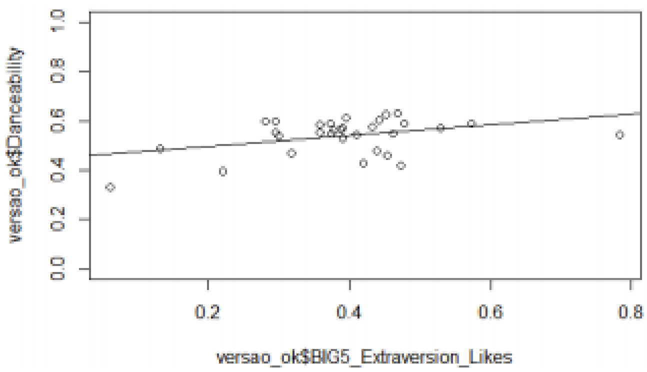

Figure 5: Linear regression - Openness vs Danceability

5.0.5. Agreeableness vs Danceability

It's found positively Association with r = 0.38; p-value = 0.036. The $R^{2 2}$ = 0.11. We can see in figure 6, the linear regression model Agreeableness vs Danceability

$$D = 0.47286 + 0.14903 * A \tag{5}$$

For the characteristic dancability, three personality traits were found that had a relevant level of correlation. In order to find a model that would bring a better result to describe the dance, we performed multiple linear regressions. As a significant result we had the model using the following variables D = Danceability E = Extroversion, O = Openness, A = Agreement. eqnref danceability

$$D = 0.39 + 0.21 * E + 0.13 * A \tag{6}$$

Neste modelo, foram obtidos os seguintes valores de p-value respectivamente p = 0,026 e 0,037. e o valor para R newunicodechar ² assegurama $^{2} = 0,23$.

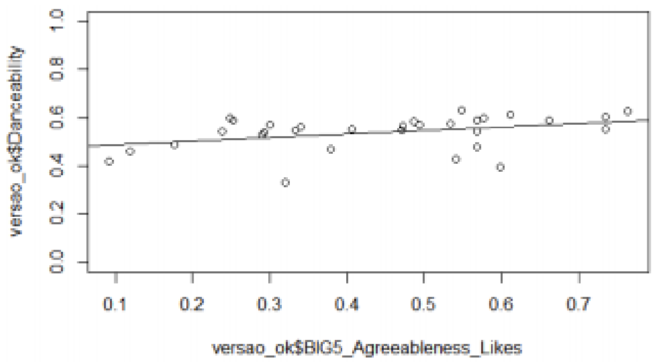

Figure 6: Linear regression - Agreeableness vs Danceability

	Acousticness	Danceability	Energy	instrumentalness
Neuroticism	0.31	0.21	-0.32	-0.20
Conscientiousness	0.02	0.11	0.28	-0.07
Extraversion	-0.30	0.40*	0.20	-0.44*
Openness	0.09	-0.38*	-0.29	0.30
Agreeableness	0.20	0.38*	-0.22	-0.02
	Liveness	tempo	Valence	Popularity
Neuroticism	0.06	0.17	0.09	0.23
Conscientiousness	0.20	0.40*	0.33	0.11
Extraversion	-0.16	0.25	0.35	0.35
Openness	-0.08	-0.22	-0.42*	-0.21
Agreeableness	-0.06	0.05	0.19	0.21

Table 2: Correlation between BigFive traits and Spotify characteristics

5.1. discussions

Of the correlations found, there were three of them that were more relevant and seemed to have an explanation of cause and effect. A marginally significant relationship was found between Popularity and Extroversion, this fact possibly occurred due to the number of users that were in the analysis. In the initial analyzes with fewer users, it was possible to identify this correlation with a p ¡0.05. This was possibly due to some distortion in the data, but according to [15] the correlations increase with the size (and variability) of the sample. With this, it is expected that with the increase of the sample the value of p decreases bringing greater confidence for the relation.

Comparing with the relationships found in the work of[5] we can see that mass music was positively related to extraversion and negatively to openness to experience, which was also the relationships found with the dance characteristic of the music, because when we observe the genres present in popular music we note that they are all genres that tend to be danceable.

The correlation between openness and Valence of music showed a significantly high negative correlation, we can interpret this correlation as people who have a lower value of opening up to new experiences, that is, more closed individuals in their world tend to listen to music that they like, looking for more positive songs that reinforce your feeling. People with a higher value of Openness tend to be more receptive to new ideas, which ends up leading this user to try and listen to songs that have a low Valence, that is, more negative sounds, exemplifying music (sadder, depressive, angry).

The result of the multiple linear regression showed us that it is possible that several personality traits can influence the characteristics of the songs as was the case with danceability. This leads me to think about what other elements of personality we can relate to music and use an approach based on the individual's personality to create initial recommendations for songs and podcasts.

6. Conclusions

The present work promoted research in the areas of personality theory involving the Big Five being obtained from a user's Facebook digital footprint. In which it is possible through the use of the Apply Magic Sauce application to estimate the 5 major factors based on the likes on Facebook pages and correlate this information with characteristics such as Valence, Popularity, Energy, Tempo, Danceability, Liveness, Instrumentalness extracted from the 100 songs that the user has more affinity with one of the biggest streaming services Spotify.

Based on these analyzes, it was possible to identify three main correlations between Valence and Openness to Experience and two other correlations that were found in our analysis: Extroversion and Popularity, Extroversion, and Danceability. These last two were also found similar relations between extroversion and music of very popular Brazilian music that is represented by songs of the genre: sertaneja, pagode, pop music, funk, forró, samba, axé, and Brega, according to [5] which highlights the correlation between these factors.

7. Future works

Add new entries for data collection, incorporate more data in the correlation analysis, as well as insert other characteristics of the music, both metadata such as band, genre, release year among other information and characteristics extracted directly from the audio such as music note, mood that to enhance analysis such as genres, music notes, and humor. In addition to incorporating an analysis of the lyrics of each song, as we well know that the lyrics have a fundamental role in the emotion that the music passes on to the listener and in this work, we do not take this possibility into account so for future works we can do analysis and keywords, and feeling of the lyrics to understand if there are relations between the music, lyrics and personality factors.

References

[1] J Clement. Number of global social network users 2010-2021. *Statistica*, 4(7), 2019.

[2] Wu Youyou, Michal Kosinski, and David Stillwell. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040, 2015.

[3] Peter J Rentfrow and Samuel D Gosling. The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6):1236, 2003.

[4] John M Digman. Higher-order factors of the big five. *Journal of personality and social psychology*, 73(6):1246, 1997.

[5] Carlos Eduardo Pimentel and Edla Daise Oliveira Porto Donnelly. A relação da preferência musical com os cinco grandes fatores da personalidade. *Psicologia: ciência e profissão*, 28(4):696–713, 2008.

[6] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

[7] Felicia Pratto, Jim Sidanius, Lisa M Stallworth, and Bertram F Malle. Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology*, 67(4):741, 1994.

[8] CHSS Nunes and CS Hutz. O modelo dos cinco grandes fatores de personalidade. *Temas de avaliação psicológica. Campinas: IBAP*, 2002.

[9] Robert R McCrae and Oliver P John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.

[10] Claudio Simon Hutz, Carlos H Nunes, Alice D Silveira, Jovana Serra, Márcia Anton, and Luciane S Wieczorek. O desenvolvimento de marcadores para a avaliação da personalidade no modelo dos cinco grandes fatores. *Psicologia: reflexão e crítica*, 11(2), 1998.

[11] Alexandre Neves. E-portefólios para a gestão da identidade digital: estudo de caso da plataforma sapo campus escolas. Master's thesis, Universidade de Aveiro, 2011.

[12] Fabien Girardin, Francesco Calabrese, Filippo Dal Fiore, Carlo Ratti, and Josep Blat. Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive computing*, 7(4), 2008.

[13] Kyle Robert Arcand II. *Stack Music: Spotify and the Platformization of the Digital Music Commodity*. PhD thesis, Duke University, 2019.

[14] Jesper Steen Andersen. Using the echo nest's automatically extracted music features for a musicological purpose. In *Cognitive Information Processing (CIP), 2014 4th International Workshop on*, pages 1–6. IEEE, 2014.

[15] Luiz Pasquali. *Psicometria: teoria dos testes na psicologia e na educação*. Editora Vozes Limitada, 2017.