Predicting Music Success Based on Users' Comments on Online Social Networks

Carlos V. S. Araujo
Computing Institute Federal
University of Amazonas Amazonas,
Brazil cvsa@icomp.ufam.edu.br

Rayol M. Neto
Computing Institute Federal
University of Amazonas Amazonas,
Brazil rayol@icomp.ufam.edu.br

Fabíola G. Nakamura
Institute of Computing Federal
University of Amazonas Amazonas,
Brazil fabiola@icomp.ufam.edu.br

Eduardo F. Nakamura Institute
of Computing Federal University of
Amazonas Amazonas, Brazil
nakamura@icomp.ufam.edu.br

ABSTRACT

In this paper, we aim to determine whether or not we can predict the success of a music album based on comments posted on social networks during the 30 days prior to the album's release. To this end, we focused on the Twitter network to gather user comments. As measures of success, we considered Spotify Popularity and Billboard Units. The reason for these choices is that Spotify represents the most popular type of music consumption today (audio streaming), while Billboard rankings still favor the old school market (physical albums). As a result, we found that the amount of positive tweets (30 days before the album release) can explain 95.5% of the variation in Spotify Popularity with a simple linear model. On the other hand, we could not find statistical evidence that the volume of comments on Twitter correlates with the album success measured by Billboard magazine.

KEYWORDS

Sentiment Analysis, Twitter, Music Success

1 INTRODUCTION

The entertainment industry (cinema, games, music, and television) generates billions of dollars annually worldwide. According to the *Recording Industry Association of America* (RIAA)[1], the music industry alone generated US$ 7.7 billion in the United States in 2016. Of this total, 75.5% relates to digital music (*streaming* or *downloading)*.

Noticing the large increase in users of digital platforms, record labels began to adopt Online Social Networks (OSN), or simply social media, to promote new *singles* (songs promoting an album) or albums. On these RSOs, users can express their personal opinions and perceptions on a wide range of topics [11], thus helping to widely promote these *singles* or albums. Additionally, these opinions can be evaluated through polarity detection, which studies people's opinions and emotions regarding products, services, reviews, and organizations [15].

Based on this, some researchers have attempted to predict the success of an album through the impact of user-generated content, taking into account a set of data obtained from blogs, social networks, and music sales [10]. Others have studied the relationship between blog rumors, the number of times a song is played on the radio, and the sales of these songs at the album and song level [9]. However, the works related to this article do not apply sentiment analysis to user opinions on online social networks to try to predict the success of an album. The success of a *single* or album today can be measured in several ways. One of the most famous is the *ranking* of the American magazine Billboard, which has been publishing *rankings* since 1945 and is one of the pioneers in this type of publication. Since 2014, the magazine has also taken into account the number of sales of individual songs from the album, purchased digitally, as well as the number of *streams* of an album. However, the sale of the complete album carries more weight than the other criteria used to determine the *ranking* score. As such, this *ranking* remains relevant to this day as a means of measuring the popularity of an album (especially physical albums). Another possible way is to analyze the popularity of an album on music *streaming* services (e.g., Spotify, Napster, Apple Music), which are becoming increasingly popular in the music market. This Its relevance is such that, in the United States, *streaming* has been the most popular form of music consumption since 2015. We will therefore correlate this data with messages on social media, together with polarity detection, to achieve the expected contributions.

polarity detection to achieve the expected contributions.

With that, the main contributions of this work are: (i) a statistical approach to the correlation analysis of comments on social media with the future performance of entertainment releases;

[1]www.riaa.com/

[2] http://www.riaa.com/wp-content/uploads/2016/03/ RIAA-2015-Year-End-shipments-memo.pdf.

WebMedia'17, October 17–20, 2017, Gramado, RS, Brazil.
 ISBN 978-1-4503-5096-9/17/10. . . $15.00

(ii) a linear model in which the number of positive tweets in the 30 days preceding the release of an album explains 95.5% of its popularity after release; (iii) a discussion of the representativeness of popularity on *streaming* services (Spotify) vs. physical sales (Billboard) as metrics of a release's success. To validate the proposed approach, we conducted a case study considering fifteen albums released in 2016 and 2017.

This article is organized as follows. In Section 2, we present works related to the topic of this article. Section 3 presents the statistical foundations we use in this work. Section 4 presents the proposed evaluation approach. In Section 5, we present the applied methodology (instantiation of the proposed approach). Section 6 contains the statistical evaluation and a discussion of the results obtained. Finally, in Section 7, we present the conclusions and point out future directions.

2 RELATED WORKS

Dhar and Chang [10] examined the feasibility of using user-generated content on blogs and social networking sites to predict sales in the music industry. The authors tracked changes in what was being said about albums. To do so, they used a sample of 108 albums for four weeks before and after their release dates. As a result, they found that future sales are correlated with the volume of blog posts about an album and traditional factors. It should be noted that these factors are albums released by a major record label and opinions.

from traditional sources, such as Rolling Stone magazine.

Dewan and Ramaprasad [9] drew a parallel between new media (digital downloads), old media (CDs), and sales of

music industry. More specifically, they studied the interaction between what was being discussed on blogs, the number of times a song was played on the radio, and the sales of those songs and
from their respective albums. For analysis, they used the panel vector autoregression (PVAR) methodology. As a result, they demonstrated that the number of times a song is played on the radio is directly related to future sales, both of the song and its album. However, they found that comments about songs on blogs are not related to album sales. However, they do negatively affect song sales, since these songs are shared for free on the Internet.

In the work of Chen et al. [5], the authors employed the same model used by Dewan and Ramaprasad [9], the autoregression
panel vectors, with the aim of investigating the relationship between social media promotions and music sales. To this end, they evaluated the automatic and personal content generated on the social network MySpace and found that social media transmission has a significant effect on sales. More specifically, they showed that this effect was greater when analyzing personal messages rather than automated messages. Finally, they showed that the timing and content of these personal messages influence sales.

Chen and Chellappa [4] evaluated user behavior over time in relation to songs available for free on the social network MySpace Music, with the aim of determining the impact on song and album sales. The authors used

a database with songs by 43 random artists and conducted the experiments from June 2007 to January 2008. The results show that user activity on social networks has a significant impact on music sales. Dewan and Ramaprasad [8] analyzed the relationship between music blogs and tracks with available samples based on theories of online social interaction. To do so, they used the *long tail,* a term used in statistics to identify data distributions, where the volume of data is classified in descending order. In the results, they suggested that the intensity of the music sample is positively associated with the popularity of a blog among its followers.
followers.

In short, the main differences between the proposed approach and the solutions presented (Table 1) are: (i) relating the polarity of comments extracted from OSI to the success of an album; (ii) using music service popularity metrics to analyze the success of an album.

3 FUNDAMENTALS

The correlation coefficient is a widely used tool for observing the association between two random variables in experimental research [2].

Pearson's correlation [12] assesses whether there is statistical evidence of a linear relationship between pairs of continuous variables in the population, and is given by

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \ \sigma_Y}, \tag{1}$$

where $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_{(n)}\}$, $\mathrm{cov}(X, Y)$ is the covariance cia between X and Y, σ_X is the standard deviation of X and $\sigma_{(Y)}$ is the standard deviation of Y.

Spearman's rank correlation coefficient [16] is a nonparametric distribution-free classification statistics,
proposed as a measure of the strength of the association between two variables. It is a monotonic measure of association, which assesses how well an arbitrary monotonic function can describe the relationship between two variables, without making assumptions about the frequency distribution of the variables, which is defined by

$$\mathrm{P} = \frac{6 - d_i^2}{n^3 - n}, \tag{2}$$

where d_i is the difference between each corresponding value position of X and Y, and n is the number of value pairs.

Distance correlation [18] is a measure of dependence between random vectors and is used to verify whether there is any relationship (not necessarily linear) between two variables (X and Y). Furthermore, X and Y can be vectors of different sizes, and is given by

$$\mathrm{dCor}(X, Y) = \frac{\mathrm{dCov}(X, Y)}{\sqrt{\mathrm{dVar}(X)\ \mathrm{dVar}(Y)}}, \tag{3}$$

where $\mathrm{dCov}(X, Y)$ is the covariance distance between (X, Y), defined by the square root of

Table 1: Summary of the main related works and characteristics.

Authors	Database	Relationship
Dhar and Chang [10]	Blogs and Social Networks	Comments vs. Sales
Dewan and Ramaprasad [9]	Blogs	Comments vs. Radio vs. Sales
Chen et al. [5]	Social Networks	Comments vs. Sales
Chen and Chellappa [4]	Social Networks	User Behavior vs. Sales
Dewan and Ramaprasad [8]	Blogs	Blog vs. Free Tracks

$$\mathrm{dCov}^2 = \frac{1}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n} A_{kl} B_{kl}, \quad (4)$$

where $A_{kl} = \| X_k - X_l \|$, $B_{kl} = \| Y_k - Y_l \|$, which are simple linear functions of the distance between pairs of elements in the database [17], and n is the number of elements in the vectors. The variance distances dVar(X) and dVar(Y) are also defined by equation 4, but using the same data vector for A *and* B.

Therefore, the three correlation coefficients provide complementary information.

Pearson quantifies the **linear correlation** between two variables. The Pearson coefficient ranges from -1 (total negative linear correlation) to $+1$ (total positive linear correlation), with a value of 0 indicating that there is no linear correlation, but does not guarantee independence between the variables. Spearman determines whether there is a **monotonic relationship** between the variables (not necessarily linear). The Spearman coefficient ranges from -1 (perfect monotonically decreasing function) to $+1$ (perfect monotonically increasing function). A value of 0 indicates that there is no monotonically decreasing function relating the two variables, but it does not guarantee the independence of the variables. Distance correlation is more robust than the previous ones in that it does not assume linearity or monotonicity. It does not depend on the type of distribution. The value of distance correlation varies from 0 to 1, being **equal to 0 if, and only if, the variables are independent (**unlike Pearson and Spearman).

For these reasons, in this study, we used the three correlation coefficients to evaluate the relationship between comments on social media and music service data to predict the success of an album.

4 PROPOSED APPROACH

The approach proposed in this study, illustrated in Figure 1, seeks to predict the performance/success of entertainment releases based on comments on OSS before release. In particular, we focus on music releases. However, the approach can be applied to the release of films, games, television series, and other entertainment products. The approach consists of the following steps:

(a) **Data Collection**. The first step in the proposed approach is collection, where data is extracted from online social networks (comments) and dissemination and evaluation services (e.g., audio *streaming* services) to generate the database.
(b) **Pre-processing**. In this step, the data is formatted so that only the necessary information can be extracted. such as product popularity and language detection of user comments.
(c) **Polarity detection**. After preprocessing, it is performed sentiment analysis was used to characterize the polarity of comments from social media. This allows us to associate the impact of comments (positive, negative, and neutral) with the success of a product.
(d) **Statistical Analysis**. In this final stage, we calculate the three correlation coefficients (Pearson, Spearman, and Distanic). This allows us to statistically evaluate the correlation between the data extracted from social networks and the data obtained on the success of the product (e.g., album sales success). In addition, the normality test is applied to test whether the sampled data are normal or not. Finally, a Probability-Probability analysis (P-P Analysis) is performed to determine whether the variables, correlated or not, follow distributions with the same behavior. Next, this visual analysis is validated with the Wilcoxon test.

In the next section, we instantiate the proposed approach and detail the specific aspects and choices/decisions for the experimental evaluation of this work.

5 METHODOLOGY

The purpose of this section is to present the methodology used for data collection, data formatting, opinion mining, and graphical representation and statistical significance, which are presented in subsections 5.1, 5.2, 5.3, and 5.4, respectively.

5.1 Data Collection

We collected messages from the social network Twitter using the Tweepy tool, which facilitates the use of the Twitter API, collecting data from the 30 days prior to the release of the albums studied here. This time frame was chosen so that we could track the growth of listeners' expectations as the release date approached.

We chose fifteen notable albums from 2016 and 2017, in a wide variety of musical styles.

For each of these albums, we collected information contained in the Spotify *streaming* service, using its own API[4]. This collection was necessary because this data contains the values popularity of each album. Popularity is defined as an integer value between 0 and 100, calculated from the average of

[3] http://www.tweepy.org
[4] https://developer.spotify.com/web-api/

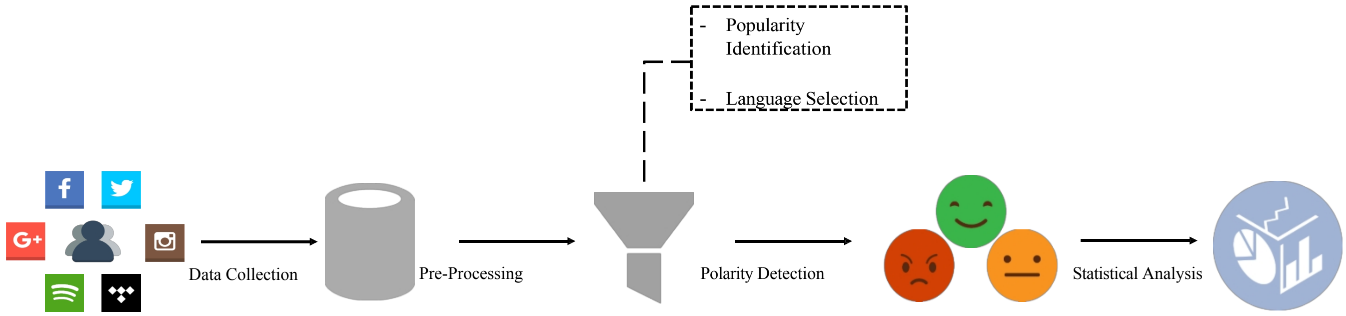


Figure 1: Proposed approach.

popularity of individual tracks on the album, already calculated by the API itself. The higher this value, the more popular the album is. We also collected data on the number of Billboard units in the week the album was released. This index refers to the number of album sales, added to the number of single sales from that album, plus the number of *streams*. Each of these has a different weight in the quantification of units in *the ranking*[5], as seen in Table 2. These values were collected manually, using the weekly news from the magazine itself, which publishes this data in a list of the ten albums that obtained the highest number of units in the week (TOP 10). Therefore, it was not possible to collect data on the albums used in this study that did not appear in the TOP 10 in their week of release.

Table 2: Calculation of Billboard units. The *streaming* weight of a song is almost four orders of magnitude lower than that of an album sale.

Form of Consumption	Billboard Units
Physical Album Sales	1,000
Digital album sales	1,000
Single Music Sales	0.100 (= 1/10)
Streaming of individual songs	$6{,}667\times 10^{-4}$ (= 1/1,500)

5.2 Data Formatting

In order to perform sentiment analysis, it was necessary to exclude tweets that were not in English. To do this, we used polyglot[6], a library for Python, one of whose functions is to identify the language of a text. Thus, only tweets in English were used in this study, due to the fact that it is a universal language, in addition to the fact that the artists mentioned in this study are known worldwide.

To collect only the value related to Spotify popularity, we searched the information returned by the query in its API for *the popularity tag* and its value. This way, we only collected the value of interest in this article.

[5] http://www.billboard.com/articles/columns/chart-beat/6320099/billboard-200-makeover-streams-digital-tracks

[6] http://polyglot.readthedocs.io/en/latest/

5.3 Polarity Detection

For this step, we used SentiStrengh[7], which is a document-level sentiment analysis system, due to its efficiency in analyzing large volumes of data. It performs well with short, informal texts, achieving 72.8% accuracy for positive texts in the article that defined it [20]. SentiStrengh uses a *lexicon* in which words are associated with a specific polarity, and is even capable of identifying the polarity of misspelled or abbreviated words [20]. This *lexicon* was constructed based on the LIWC dictionary [19] with the addition of new features. For example, a set of words that give greater intonation to the sentiment (e.g., "very"), a set of emoticons, and perceptions of repeated punctuation (e.g., "nice!!!") [1].

For each tweet in our database, we applied SentiS-trengh to define its polarity. Each sentence received a value representing its popularity, ranging from -4, higher negative polarity, to $+4$, higher positive polarity. Thus, for each database referring to an album, two sums were calculated, whose values represent the negative and positive polarity scales. The first was defined as the sum of the values of the sentences with negative polarity, while the second was given by the sum of the values of the sentences with positive polarity.

In this way, our work also takes into account the expressiveness that a given phrase conveys. That is, in addition to detecting whether the message is negative or positive, we also detect how expressive that polarity is. For example, a message stating *"This album is good"* has positive polarity ($+1$), but less expressiveness than *"This album is very good!!!"*, which has polarity equal to $+3$. Thus, by summing the polarities of the phrases, rather than simply counting the number of negative or positive tweets, we take this type of expressiveness into account.

5.4 Statistical Analysis

Finally, we created graphical representations, relating the total number of tweets and the values of the negative and positive scales to Spotify popularity and Billboard unit sales. These representations were created using the R language.

[7] http://sentistrength.wlv.ac.uk/

We created scatter plots and probability-probability plots. The first used Pearson, Spearman, and distance correlation coefficients to find correlations between the data studied. The second complements and validates the first. We also performed normality tests to confirm our conclusions.

6 RESULTS

After collecting, formatting, and analyzing the data, we obtained the values shown in Table 3.

To verify whether there is a correlation between the success of a music album and the volume of *tweets* during the 30 days prior to the album's release, we performed:

(1) **Normality Test**. Initially, we present the results of the Shapiro-Wilk [14] Normality test, as they confirm the conclusions about the other analyses presented.
(2) **Dispersion Analysis with Hypothesis Testing**. In this case, we calculated the Pearson and Spearman correlation coefficients and Distance coefficients. Next, we computed the t-test to determine the statistical significance of each test. In addition, the quality of the models is also evaluated using the coefficient of determination r^2, the residual variance σ_r^2, the *mean square error* (MSE) and the F-test [3] for model adequacy.
(3) **Probability-Probability Analysis**. In this case, compa- We calculated the *cumulative distribution functions* (CDF) of the two contrasted variables and evaluated them according to their values of EQM and σ_r^2. This analysis complements the previous one, verifying the similarity between the FDAs of the two variables and validated with the Wilcoxon test.

6.1 Data Normality Test

To test whether the sampled data is Normal or not, we apply the Shapiro-Wilk test, with the **null hypothesis that the distributions are normal**. Using $\alpha = 0.05$, we find that the hypothesis that the data follow a normal distribution is accepted only for Spotify popularity and positive tweets, since p-value > $\alpha = 0.05$, accepting the hypothesis.

6.2 Spotify Popularity vs. Tweets

Figure 2a shows the scatter plot of Spotify Popularity vs. Number of Tweets. The graph includes the linear regression curve of the points (in red) with the respective confidence interval (in gray). In this case, when considering only Positive Tweets, the points are closer to the regression curve.

Statistical tests consider the **null hypothesis that there is a correlation between the variables**. The results show that, considering Positive Tweets, the p-value of all tests is below $\alpha = 0.05$. Therefore, the three calculated correlation coefficients are statistically significant. Consequently, there is a positive correlation between the number of Positive Tweets published in the 30 days prior to the album's release and its Spotify Popularity after release.

Pearson's coefficient of 0.665 > 0.5 indicates a strong positive linear correlation, as explained by Cohen [7]. In addition, the t-test with a p-value of 0.007 < $\alpha = 0.05$ shows that this result is statistically significant.

When comparing the FDAs (Figure 2b), we observed that the distribution of Positive Tweets is the most similar to the distribution of Spotify Popularity. In this case, the mean square error is 0.019 with a residual variance of 0.004 (Table 6a). These results are statistically supported by p-values greater than $\alpha =$ r 0.05, accepting the Wilcoxon test hypothesis that the distributions are equivalent.

The same observations are valid considering All Tweets. However, the statistics for All Tweets are slightly less significant than those for Positive Tweets. Considering Negative Tweets, the hypothesis that there is a correlation is rejected for all tests performed, as their p-value is greater than $\alpha =$ r 0.05 (Table 6a).

Considering positive tweets, the linear regression model is given by:

$$P_{pos}(p) = 56.700 + 6.389 \times 10^{-4}\, p, \tag{5}$$

where $P_{pos}(\cdot)$ is the Spotify Popularity prediction, considering the number p of positive tweets. If we consider the total number of tweets, the model is given by

$$P_{tot}(t) = 57{,}930 + 2{,}441 \times 10^{-4} t, \tag{6}$$

where $P_{tot}(\cdot)$ is the Spotify Popularity prediction, considering the quantity t of total tweets.

The quality of the two models can be assessed by considering the coefficient of determination r^2, the residual variance $\sigma_r^{(2)}$, and the EQM of the models. In this case, the best model is the one in which r^2 is closest to one, σ_r^2 and EQM are close to zero. Therefore, the model $P_{pos}(\cdot)$ is the best, given that the number of Tweets Positives explain 95.5% ($r^2 = 0.955$) of the variation in Spotify Popularity (Table 5). On the other hand, the Total number of Tweets explains slightly less of the variation in Spotify Popularity, in this case 86.4% ($r^2 = 0.864$). The p-values for the F-test are less than $\alpha = 0.05$ (Table 5), indicating the adequacy of the models and rejecting **the hypothesis that the model is not adherent.**

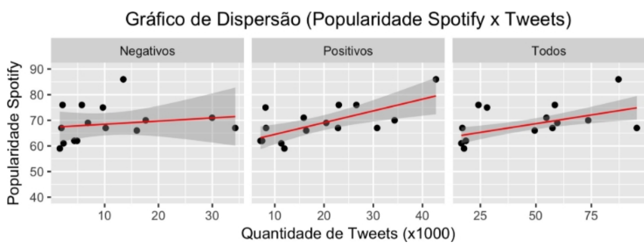
6.3 Billboard Units vs. Tweets

When analyzing the P-P graphs in Figure 2d, we observe that the FDAs of Billboard Units and Tweet Quantities approximate the theoretical behavior. This visual analysis is confirmed by the mean square error and residual variance values, all less than or equal to 0.003 and 0.001, respectively (Table 6b).

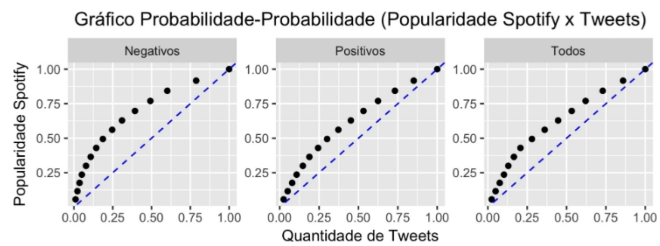
On the other hand, the correlation between the variables (Billboard Units vs. Tweets) is less pronounced, especially for Positive Tweets and All Tweets, when compared to the Spotify Popularity analysis. This fact is confirmed by Wilcoxon tests, whose p-values are all greater than A= R 0.05, accepting the hypothesis that the behavior of the distributions is similar (in terms of FDA). In this case, there is insufficient statistical evidence to support the claim that there is a correlation between the number of tweets published 30 days before the release of an album and its success in the Billboard ranking.

Table 3: Data obtained after collection and sentiment analysis (descending order of Spotify Popularity).

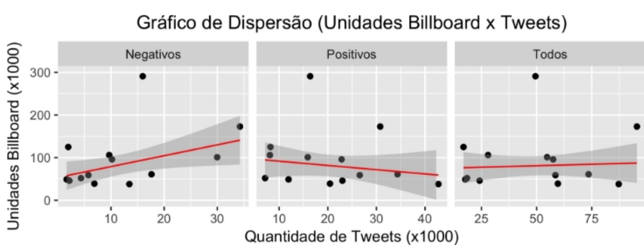
Artist	Album	# Tweets Negative	# Tweets Positive	# Total Tweets	Popularity Spotify (*)	Units Billboard
Chance the Rapper	Coloring Book	13,469	42,749	87,214	86	38,000
Lukas Graham	Lukas Graham	5,758	26,580	58,547	76	59,000
The XX	I See You	2,175	22,995	24,143	76	46,000
Rick Ross	Rather You Than Me	9,672	8,109	27,986	75	106,000
Childish Gambino	Awaken, My Love!	29,970	15,878	54,759	71	101,000
Macklemore	This Unruly Mess I've Made	17,621	34,331	73,564	70	61,000
The Chainsmokers	Collage	6,883	20,447	59,629	69	39,000
Radiohead	A Moon Shaped Pool	34,262	30,785	95,458	67	173,000
Dj Khaled	Major Key	10,188	22,844	57,500	67	96,000
The Lumineers	Cleopatra	1,974	8,217	16,865	67	125,000
Metallica	Hardwired... to Self-Destruct	15,971	16,381	49,469	66	291,000
Deadmau5	W:/2016ALBUM/	4,882	7,431	18,446	62	-
Incubus	8	4,366	7,106	18,453	62	52,000
Tom Odell	Wrong Crowd	2,354	11,321	16,408	61	-
Lindsey Stirling	Brave Enough	1,665	11,937	17,644	59	49,000



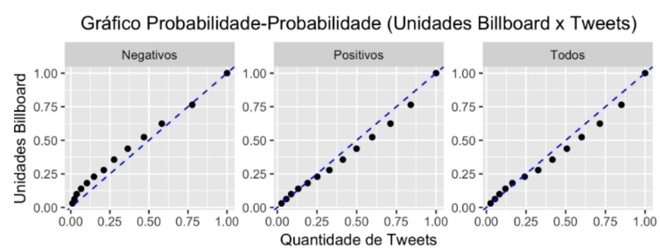
(a) Dispersion Analysis of *Spotify Popularity vs. Number of Tweets* (negative, positive, and total) with the regression line in red and confidence interval (α= 0.05) in dark gray.



(b) Probability-Probability Analysis of *Spotify Popularity vs. Number of Tweets* (negative, positive, and total). The dashed line represents the behavior in the case of equivalence.



(c) Dispersion analysis of *Billboard Units vs. Number of Tweets* (negative, positive, and total) with the regression line in red and confidence interval (α= 0.05) in dark gray.



(d) Probability-Probability Analysis of *Billboard Units vs. Number of Tweets* (negative, positive, and total). The dashed line represents the behavior in the case of equivalence.

Figure 2: Graphical analyses: (a) Dispersion for Spotify Popularity vs. Tweets, (b) Probability-Probability for Spotify Popularity vs. Tweets, (c) Dispersion for Billboard Units vs. Tweets, (b) Probability-Probability for Billboard Units vs. Tweets.

6.4 Discussion of Results

An explanation for why positive tweets explain the popularity of an album can be found in psychology. Cialdini and Garde [6] state that the actions of those around us are important in making decisions. According to the authors, consumers follow a simple heuristic: if it is popular, it is good.

For example, if many people are buying the same product, there is a good chance that this product deserves our attention. The influence of those around us is explained by Rowlands [13], who states that civilization advances as the number of operations we can perform without thinking about them increases. That is, as we live in a complex world, we make decisions based on others.

Table 4: Shapiro-Wilk test for the collected data. Null hypothesis H_0: the distributions are normal.

Sample Evaluated	statistic	p-value	Normal
Spotify Popularity	0.938	**0.358**	yes
Billboard Units Negative	0.738	0.001	no
Tweets	0.827	0.008	no
Positive Tweets	0.919	**0.184**	yes
Total Tweets	0.879	0.046	no

Table 5: Quality of Spotify Popularity prediction models given the number of Positive Tweets and Total Tweets. Hypothesis H_0: model is not adherent (F-test).

Model	r^2	σ_r^2	EQM	p-value
P_{pos} x	**0.955**	**2.277**	**2.125**	**3.814×$^{10-10}$**
P_{tot}	0.864	6.919	6.458	5.462× $^{10-7}$

as heuristics, or mental shortcuts, to help us get on with our lives.

In our context, when many people speak highly of a particular album, common sense will indicate that the album is good, causing more people to listen to it and thus increasing its popularity. This fact is particularly strong in the virtual environment, described by Spotify Popularity and RSOs, whose consumption and interaction profile is much more instantaneous than in the physical environment, in this case described by Billboard Units.

One possible reason for the correlation with Spotify popularity, but not with Billboard units, is that the weight of the number of *streams* in the unit value is much lower than that of an album sale. When collecting the data, we observed that the number of album sales continues to be the most influential factor in the unit value. Albums released only on digital services have never managed to reach the top of *the ranking*. For example, Chance the Rapper's album *Coloring Book* was released only on *streaming* services*,* obtaining 38,000 Billboard units in its first week*, equivalent to 57*.3 million streams. This is the most popular album on Spotify (in this study) and has the highest value on the positive scale.
Billboard in its first week, equivalent to 57.3 million *streams*[8]. This is the most popular album on Spotify (in this study) and has the highest value on the positive sentiment scale. Even artists who prioritize *streaming* releases,
only managed to reach the top by adding up sales. This scenario occurred with rapper Kanye West[9] , whose album was the first to reach the top of the Billboard *charts* primarily based on the number of *streams*.

Another interesting point is the performance of the album *Hardwired... to Self-Destruct* by the band Metallica, highlighted in Table 3. This album follows the linear model between Positive Tweets and Spotify Popularity, but appears as the largest *outlier* in the analysis that considers Billboard Units. This can be explained by the fact that the band Metallica has always fought, including in court, strongly against music sharing on the Internet. Thus, the band Metallica remains more popular in the physical media market (captured by Billboard) than in the *streaming* market (captured by Spotify).

[8] http://www.billboard.com/articles/columns/chart-beat/7378360/ drakes-views-no-1-billboard-200-album-chart-meghan-trainor-thank-you

[9] http://www.billboard.com/articles/columns/chart-beat/7326493/ kanye-wests-the-life-of-pablo-debuts-at-no-1-on-billboard-200

7 CONCLUSION

In this study, we sought to answer the question: "*Is it possible to pre-predict the success of an album based on the volume of comments on Twitter (about this album) during the 30 days leading up to its release?*"

To this end, we collected data on fifteen albums released between 2016 and 2017. We performed a statistical evaluation using Pearson, Spearman, and Distance correlation coefficients. All evaluations were validated with statistical significance tests. In addition, we performed a counter-analysis
by combining the cumulative probability functions to determine
how similar the distributions are.

As a result, we show that we can estimate an album's performance, considering Spotify Popularity, based on the volume of comments on Twitter during the 30 days prior to
its release. However, considering the Billboard ranking, we were unable to identify a significant statistical correlation.

It is interesting to note that Spotify Popularity refers directly to the number of times an album (or its songs) are played on this increasingly popular and significant *streaming* service. On the other hand, the more traditional and closed Billboard ranking gives little value to the digital *streaming* environment. In this context, Spotify Popularity tends to be a more relevant metric because it is transparent and more popular.

Another relevant result is that, when using sentiment analysis, we found that the correlation of Positive Tweets is more significant in determining the success of an album, given the Spotify Popularity metric. Therefore, sentiment analysis has a positive impact on the evaluation presented.

As future work, we are expanding the database to include more albums and evaluate other popularity metrics, such as digital and physical media sales. In addition, we are collecting data on movie and television series releases to assess whether it is possible to predict the success or failure of a movie, given comments on social media before its release.

REFERENCES

[1] Matheus Araújo, Pollyanna Gonçalves, Fabrício Benevenuto, and M Cha. 2013. Methods for sentiment analysis on Twitter. In *Proceedings of the 19th Brazi lian symposium on Multimedia and the Web (WebMedia'13)*.
[2] Atanu Bhattacharjee. 2014. Distance correlation coefficient: An application with Bayesian approach in clinical data analysis. *Journal of Modern Applied Statistical Methods* 13, 1 (2014), 23.
[3] George Casella and Roger L. Berger. 2001. *Statistical Inference* (2nd ed.). Cengage Learning. 660 pages.
[4] Conny Chen and Ramnath Chellappa. 2009. Music sales and online user activity at social networking sites. In *fifth symposium of the Heinz College's iLab Center for Digital Media Research, Carnegie Mellon University on Statistical Challenges in Electronic Commerce Research, Pittsburgh*. 30–31.
[5] Hailiang Chen, Prabuddha De, and Yu Jeffrey Hu. 2015. IT-enabled broadcasting in social media: An empirical study of artists' activities and music sales. *Information Systems Research* 26, 3 (2015), 513–531.
[6] Robert B Cialdini and Nathalie Garde. 1987. *Influence*. Vol. 3. A. Michel.
[7] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. 400 pages.

Table 6: Statistical correlation assessment between: (a) Spotify popularity vs. Tweets and (b) Billboard units vs. Tweets. P-P analysis verifies the similarity between the cumulative probability functions of the two opposing variables, where EQM is the mean square error and $\sigma_r^{(2)}$ is the residual variance. For Pearson, Spearman, and Distance, the null hypothesis is $H_{(0)}$: the variables are not correlated. For Wilcoxon, the null hypothesis is $H_{(0)}$: the distributions are equivalent.

(a) Spotify Popularity vs. Tweets

	Pearson		Spearman		statistical		P-P Analysis		Wilcoxon EQM	
	distance	p-value	statistical	p-value	statistical	p-value		σ_r^2	Statistics	p-value
Negative Tweets	0.501	0.189	0.186	0.361	0.433	0.227	0.053	0.011	160.00	0.050
Positive Tweets x	0.665	**0.007**	0.592	**0.020**	0.636	**0.021**	**0.019**	**0.004**	**140,000**	**0.267**
All Tweets	0.519	0.048	0.596	0.019	0.580	0.032	0.022	0.004	144,000	0.202

(b) Billboard Units vs. Tweets

	Pearson		Spearman		Distance p-		P-P Analysis		Wilcoxon EQM	
	value	p-value	p-value					σ_r^2	Statistics	p-value
Negative Tweets	0.429	0.144	0.407	0.170	0.545	0.168	0.003	0.001	96.500	0.555
Positive Tweets	−0.176	0.566	−0.253	0.404	0.365	0.727	0.002	0.001	82.500	0.939
All Tweets	0.113	0.712	−0.093	0.765	0.341	0.822	0.003	0.001	84.000	1.000

[8] Sanjeev Dewan and Jui Ramaprasad. 2012. Research Note—Music Blogging, Online Sampling, and the Long Tail. *Information Systems Research* 23, 3-part-2 (2012), 1056–1067. DOI:http://dx.doi.org/10.1287/isre.1110.0405
[9] Sanjeev Dewan and Jui Ramaprasad. 2014. Social media, traditional media, and music sales. *Mis Quarterly* 38, 1 (2014), 101–121.
[10] Vasant Dhar and Elaine A. Chang. 2009. Does Chatter Matter? The Impact of User-Generated Content on Music Sales. *Journal of Interactive Marketing* 23, 4 (2009), 300–307. DOI:http://dx.doi.org/10.1016/j.intmar.2009.07.004
[11] Rayol Mendonca Neto, Bruno Ábia Souza, Thaís Gomes Almeida, Fabíola Guerra Nakamura, and Eduardo Freire Nakamura. 2017. An Approach to Identifying Influential Entities in Events Commented on Online Social Networks. In *Proceedings of the 14th Brazilian Symposium on Collaborative Systems*.
[12] Karl Pearson. 1895. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* 58 (1895), 240–242.
[13] Mark Rowlands. 2011. *Body language*. MIT Press.
[14] S. S. Shapiro and M. B. Wilk. 1965. An analysis of variance test for normality s (complete samples). *Biometrika* 52, 3-4 (1965), 591–611.
[15] Bruno A. Souza, Thais G. Almeida, Alice A.F. Menezes, Fabíola G. Nakamura, Carlos M.S. Figueiredo, and Eduardo F. Nakamura. 2016. For or Against?: Polarity Analysis in Tweets About Impeachment Process of Brazil President. In *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web (Webmedia '16)*. ACM, New York, NY, USA, 335–338. DOI:http://dx.doi.org/10.1145/2976796. 2988216
[16] C h a r l e s Spearman. 1904. The proof and measurement of association between two things. *The American Journal of Psychology* 15, 1 (1904), 72–101.
[17] Gábor J Székely, Maria L Rizzo, and others. 2009. Brownian distance covariance. *The Annals of Applied Statistics* 3, 4 (2009), 1236–1265.
[18] Gábor J Székely, Maria L Rizzo, Nail K Bakirov, and others. 2007. Measuring and testing dependence by correlation of distances. *The Annals of Statistics* 35, 6 (2007), 2769–2794.
[19] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54.
[20] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61, 12 (2010), 2544–2558.