Integrated Visual Analysis for Brand Perception on Different Social Networks*

Joana P. Scherer, Pedro Henrique Morais Sanvido, Milene S. Silveira, Isabel H. Manssour
Pontifical Catholic University of Rio Grande do Sul, PUCRS. School of Technology.
joana.scherer@edu.pucrs.br, pedro.sanvido@acad.pucrs.br, {milene.silveira,isabel.manssour}@pucrs.br

Abstract

Nowadays, social media fosters billions of users and has become an important communication source. Social media empowered the users' opinions supplying quick and less costly feedback compared to traditional survey methods. However, analyzing a brand through its social media is not trivial and raises challenges. To benefit from its data, brands need easy-to-use tools to help them understand the data gathered. Thus, the main goal of this work is to provide a visual analysis approach composed of several interactive visualization techniques to support brands to obtain insights from their social media. Besides the analyses allowed by integrating different social media data, the proposed approach has a well-defined pipeline that can be extended and used without programming knowledge. We conducted two case studies to explore these analyses, one for Netflix and one for Amazon Prime Video data. The results help us to highlight the approach's potential and possible future investigations.

1. Introduction

Since its creation, social media has experienced exponential growth. Nowadays, social media fosters billions of users over different platforms and has integrated itself into our daily lives, and has become an important communication medium. Over time, it became usual for users to externalize opinions regarding various subjects, like politics, products, brands, among others. This behavior has increased the proximity among brands and users while also empowering users' opinions. This proximity is useful for brands given the quick [1] and less costly feedback when compared to traditional survey methods. Thus, the feedback obtained through social media has now become a rich source for marketing campaigns and strategies.

Likewise, social media has also empowered the customer side [2]. A recent example of the power of social media was the release of the Sonic movie trailer, which garnered so many negative comments regarding the appearance of its main character, that the studio took the decision to spend 5 million dollars redesigning it. Redesigning the character was costly for the studio, but has probably avoided more losses by preventing a box office flop. This episode illustrates the importance of social media vigilance for brands and users' empowerment. Moreover, the dynamic nature of social media can cause a range of problems for a brand, given its public perception can quickly change. An inadequate product launch, or a wrongly worded announcement, for example, can generate backlash very fast, causing damage and prejudice for the brand.

Given the present situation, the importance of social media monitoring can be attested, as brands can then be able to provide a proper response as soon as a problem arises. It is worth mentioning that, for this work, we are using the term brand in a more broad sense. We consider that a person can also be referred as a brand, considering that many artists, celebrities, athletes, and YouTubers are monetized through their social media accounts. For example, Lionel Messi[1], Gisele Bündchen[2] and Taylor Swift[3], who have 213, 17.2, and 161 million followers on Instagram, respectively, are some examples of celebrities that can be considered a brand.

However, monitoring the social media of brands is not a trivial task and poses several challenges. At first, data gathering can be problematic, given the vast amounts of data generated by many different social media, which needs to be collected and properly stored, and prepared for analysis and interpretation. Second, regarding the analysis of data itself, we propose a

*The results of this work were achieved in cooperation with HP Brasil Indústria e Comércio de Equipamentos Eletrônicos LTDA. and Hewlett Packard Brasil LTDA. using incentives of Brazilian Informatics Law (Law nº 8.2.48 of 1991), and was partially supported by UolEdTech. Manssour thanks the financial support of the CNPq Scholarship-Brazil (308456/2020-3). The authors also would like to thank the financial support of the PDTI Program, financed by Dell Computers of Brazil Ltd. (Law 8.248/91).'

[1] https://www.instagram.com/leomessi/
[2] https://www.instagram.com/gisele/
[3] https://www.instagram.com/taylorswift/

URI: https://hdl.handle.net/10125/79712
978-0-9981331-5-7
(CC BY-NC-ND 4.0)

solution of combining several visualization techniques in a dashboard, allowing brands to obtain significant insights into the collected data. While some works with similar purposes have already been researched [3, 4], to the best of our best knowledge, they are limited to Twitter and Glassdoor (for [4] only) analysis. Given this limitation, they do not provide means to compare and manage public perception among multiple social media, which is desirable as brands look to reach wider audiences across many different platforms. Moreover, those works did not provide analysis regarding time series, and comment chains, which can be useful features to asses the repercussion of controversial posts.

In this context, the main goal of this work is to provide a visual analysis approach with several interactive visualization techniques, with the objective of support brands in obtaining insights about the collection of data in three social platforms: Twitter, Instagram, and Youtube. Our approach allows the brand to inspect public perception, identifying which posts got more attention (good and bad), the most successful ones, and how the comments of a post behave, besides providing comparisons of how it performs in different social media. Thus, the focus of our approach is the provided features and visualizations for brand perception analysis and management. Our main contribution is the visual analysis approach to explore and compare three social media, which has a well-defined pipeline that can be easily extended and used without programming knowledge.

The remainder of this paper is organized as follows. In Section 2 we present some related work for this research. Section 3 describes the developed visual analysis approach, including the research methodology, the development environment, the procedures for data collection and processing, and the interactive visualizations. The case study used to validate our approach is described in Section 4. Section 5 discusses the obtained results as well as research opportunities. In the last section, we outline our conclusions.

2. Related Work

Microblogging platforms, such as Twitter and Sina Weibo, are frequent targets of data analysis and consequently covered by several works. For example, Singh et al. [5] provide a methodology that can be used to extract relevant information from Twitter. Other works covering microblogging analysis will often focus on live-streamed data gathering and cleaning. At the same time, some may also implement mechanisms, such as event detection during the streaming of data [6, 7, 8].

Specifically considering the focus of the research here presented, the use of social networks to evaluate the perception of a brand [9] has also been investigated. For example, Liu et al. [4] and Arvanitidis et al. [3] provide an analytical platform for monitoring brand social media data[4]. Fernández-Gómez and Martín-Quevedo [10] attempt to discover which types of posts are more likely to attract engagement for brands. Microblogging can be used to assess brands according to brands' personality dimensions [11, 4]. In a different approach [12] the authors identify which variables enable organizations to manage their social network services more efficiently.

Besides microblogging platforms, Facebook and Instagram are social networks that often cover the brands' and public perception topics. The papers exploring these platforms are, in their majority, focusing on attention prediction. Lakkaraju Ajmera et al. [13] focus the attention prediction on general posts (e.g., from individual accounts), and Mazloom et al. [14] center their attention on brand-related ones. The pursuit of efforts on the topics of attracting attention is also seen by Kumar et al. [15]. In their work, the authors attempt to predict the best times of the day to maximize the attention and engagement a post can receive. Another work explores the analysis of a brand's overall popularity, incorporating spatio-temporal features was developed by [16].

We also examined papers that use blogs and websites as data sources. Campos Filho et al. [17] provide a Solar System graphic showing the most relevant brands' topics among the collected blogs. Kucher et al. [18] provide a visual analytics solution that supports analysis for texts from multiple sources. They provide a visualization that shows the brand's posts in a timeline view. An additional contribution was the development of a tool that can compare the contents of different texts.

Observing the selected papers is possible to notice that most of them use only one social media as a data source [17, 7, 9, 11]. No more than 2 papers [4, 18] combine multiple social media with building a general brand perception. However, these papers do not perform any cross-platform comparison among them. In this regard, Kumar et al. [18] did provide a comparison among different social media, yet they only operate on the domain of microblogging texts.

Considering visualization techniques, they were present in several papers. In this regard, two made use of wordclouds [9, 5]. This technique is used to provide an overall of what is being said about a brand (topics, positive and negative words). Another frequently employed technique was bar charts [9, 3, 5],

[4] Although the article [3] mentions http://branty.org/ as Branty web application, this site was not available during the development of this work.

in which they were used to provide rankings among the data, and general value comparison.

Some papers provided more unique and elaborate visualizations. The paper [6] presents a tree view in which the blocks' size corresponds to the most commented subjects. Another work [17] introduces a solar system where brand-related content is spread in different tiers according to how relevant they are. Liu et al. [4] proposes a "brand wheel", which is sliced by the brand's personality trait areas (sincerity, excitement, competence, sophistication, and ruggedness). Likewise, the size of the slice corresponds to how relevant a trait is. Lastly, Kucher et al. [18] organizes textual data from social media in a timeline view, providing means to compare them in regards to sentiment and stance.

Among the analyzed papers, five of them present some form of sentiment analysis. Two papers [17] presented the more direct approach of counting which words were said more often. Some other works [14, 16] perform sentiment analysis in a combined approach using both texts and images. The final two works [4, 11] classified the brands according to the brand's personality dimensions.

Only one paper [16] provides analysis through time. This paper analyzes the popularity of a brand's posts through time, using these to predict the brand's popularity. While some research [6, 7, 8] perform data gathering in real-time, the usual concern is to only filter noisy data and event detection. They do not provide brand analysis perception through time or in real-time. Time series analysis was also only presented by one paper [11].

We can observe how research themes and practices are distributed among different social platforms given the subset of analyzed papers. Papers related to microblogging services, such as Twitter, are generally concerned about monitoring, live data gathering, and event detection subjects. Regarding social media like Instagram and Facebook, the paper concentrates on popularity prediction. For sites and blogs, the papers focus on text analysis to discover brand topics and provide means to compare texts from different platforms. A deeper analysis of related work, including some commercial tools, can be found in [19].

3. Approach Description

Our visual analysis approach was developed using the Python programming language and different libraries and tools: Pandas for data analysis and manipulation; Seaborn, Matplotlib, and WordCloud for data visualization; and, due to its high accuracy for social media analysis, Vader for sentiment analysis [20]. The next sections detail the proposed pipeline, how to collect, pre-process, clean, and persist data, as well as its interactive visualizations and filters.

3.1. Pipeline

The pipeline for our approach was divided into five main steps that correspond to the data flow, as illustrated in Figure 1: (1) Data Collection; (2) Data Preprocessing; (3) Data Storage; (4) Data Cleaning and Enriching; (5) Interactive Filters and Visualizations [5]. The first step encompasses the collection of social media data. Next, this data is pre-processed through the execution of sentiment analysis algorithms. We chose to run the sentiment analysis at this point to minimize performance issues during the interactive visualizations. After that, the data is stored in our database. The next step provides data cleaning for word cloud generation and processes statistical information. Finally, the development of the interactive filters and visualization techniques occurs in the last step.

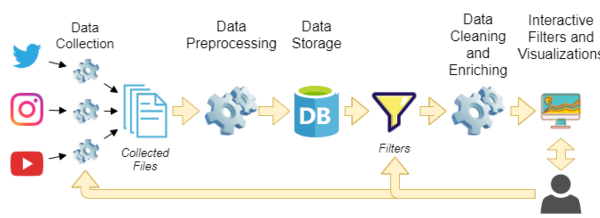

Figure 1. Steps of the approach pipeline.

This pipeline allows the user to update the data without the need to make new implementations. The user only needs to provide new data files and execute the pipeline again. At this moment, the system will update itself, automatically executing preprocessing and sentiment analysis. Afterward, the user will be able to visualize and interact with the new data. Moreover, it can be easily extended because it was designed to allow new filters or, e.g., change to other libraries for sentiment analysis.

3.2. Data Collection

Twitter and YouTube data were collected through their APIs, while Instagram data was collected through a developed script. We chose Twitter because it is the social network people use to comment about several subjects, from daily events to politics. The others were selected because several brands use them for advertising products and launching commercials. Moreover, these three Social Media have millions of users and similar

[5] All the scripts developed for this study are available at GitHub (https://github.com/DAVINTLAB/BrandAnalysis).

features, as the number of likes, shares, and comments, facilitating the comparison among them.

3.3. Data Preprocessing and Persistence

The files with the collected data are uploaded and then stored in a MongoDB database that has been modeled using eight collections: two for Twitter (profile and posts), three for Instagram (profile, posts, and comments), and three for YouTube (profile, videos, and comments). We standardize the input data for the English language to provide sentiment analysis.

The processes to determine a post language and provide its sentiment analysis were executed as preprocessing steps. Two libraries are combined to improve the results: Langid and Langdetect. First, the text is analyzed using both libraries. If both of them return the same language, then this language is assumed as the text language. If the results differ, then the reliability score of Langid is checked: if its score is higher than 80% , the result is assumed to be the text language; otherwise, the text is discarded. Since Langid presented more reliable results during the pre-implementation tests, it was selected when the libraries return different results. Our pre-implementation tests also showed that results below 80% were not reliable.

Vader [20] considers emojis, punctuations, slangs, initialisms and acronyms, and word-shape for sentiment analysis; therefore, data clean-up is not needed at this point. It provides the compound score for sentiment analysis that is a standardized score calculated through the lexicon scores. The compound scores vary between -1 and 1. We considered scores lower or equal to -0.5 as a negative text, greater or equal to 0.5 as positive, and neutral for any value between -0.5 and 0.5.

3.4. Data Cleaning and Enriching

Our approach provides several statistics data, such as the number of posts, the totals, and an average number of likes and replies. These statistics are dynamically calculated according to the user-selected filters. The complete description of all statistics provided by our approach is available in Section 3.5.

For word cloud generation, first, it is necessary to remove the stopwords. For this, we use the WordCloud library's stopword removal functionality, which worked very well in our pre-implementation tests. This process happens dynamically each time a post is selected in the Sentiment Analysis Screen (see Section 3.5), and it only happens for Instagram and YouTube comments.

3.5. Interactive Filters and Visualizations

We decided to use details on demand [21] integrated with coordinated multiple views [22] for the filters and interactive visualizations. In this perspective, the information displayed gets more specific according to the user interaction, while a user selection is reflected in the different charts. Since our target audience is the general public that might not have visualization literacy [23, 24], we chose to use simple visualizations and interfaces. The visualizations are based on common charts, e.g., pie charts, vertical and horizontal bar chars, line charts, heatmaps, and word clouds. We also used the corresponding social media color in our visualizations to easily identify to which social media the data belongs [23, 25].

In this work, we consider that interaction corresponds to any possible reaction for a particular social media, such as likes, replies, and comments. Therefore, we consider likes, retweets, and replies as Twitter interactions; likes and comments as Instagram interactions; and likes, dislikes, views, and comments as YouTube interactions.

Figure 2 shows the main dashboard of our approach. It is divided into two sections: the filter section (Figure 2A) and the charts section (Figure 2B,C,D,E,F). The available filters allow selecting data by the amount of years, social media, or date (year, month, day, and hour). The amount of years filter refers to how many years must be shown on the dashboard. The social media filter enables to choose which social media will be shown on the dashboards. The remaining filters are to select a period of time.

The charts in Figure 2B and 2C are considered general ones because they provide an overview of the three social media together. The pie chart (Figure 2B) shows the amount and percentage of posts on each social media. The bar chart (Figure 2C) displays the number of posts on each social media during the period specified by the filters.

Bellow the general charts (Figure 2B and 2C), the screen is divided into three horizontal sections, one for each social media: Twitter, Instagram, and YouTube, from top to bottom, respectively. The bar charts in Figure 2D present the total number of interactions in each social media in a period of time: likes, retweets, or replies for Twitter; likes or comments for Instagram; likes, dislikes, views, or comments for YouTube. The radio button on the left allows selecting the interaction of interest (e.g., likes or comments for Instagram). These bar charts are important to show the number of interactions increasing or decreasing, allowing future actions by the brand.

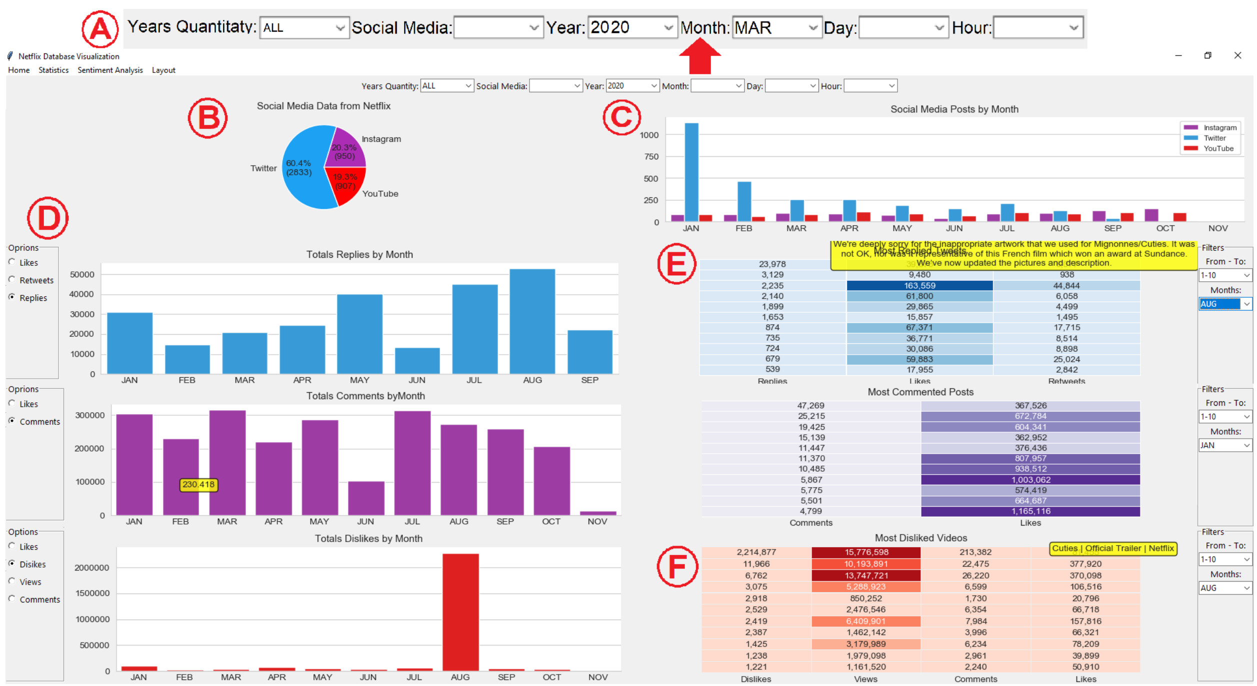

Figure 2. The main dashboard of our visual analysis approach: (A) interactive filters; (B) pie chart for global post number; (C) details of (B); (D) interactive bar charts; (E) interactive heatmaps.(F) YouTube dislike charts

The heatmap charts present posts with the highest or lowest number of interactions for each social media. The same radio button that controls the bar chart controls the heatmap sorting and changes their left column. So if the dislike option is selected, the heatmap will be sorted by the most disliked posts, and the dislike column will appear on the left of the chart (Figure 2F). Each chart shows 10 posts at a time, except when the post number is lower than 10. This chart shows detailed information about the selected filter. For instance, if a year is selected on the filter, the heatmap chart will display the posts by month. This option can be changed by combo box on the right. The heatmap charts also enable the visualization of the post's text by pointing the mouse to its cell. By clicking the cell, the respective social media is open, and the original post showed. The approach also provides the possibility to substitute the bar charts to line charts using the "Lines" option on the menu. Different from the bar chart, the line charts will show a comparison between the totals of interactions.

There are two other possible interactions for the dashboard: changing the layout orientation and selecting which social media will be displayed. The first one allows the user to change the dashboard orientation from horizontal to vertical. The second one allows the user to show or hide charts regarding specific social media data. For instance, the user may hide the charts that belong to Twitter.

Another dashboard to examine posts individually is also provided (Figure 3). This dashboard shows the list of collected posts for Instagram or the list of videos for YouTube (Figure 3A). A pie chart (Figure 3B) shows the percentage of positive, negative, and neutral comments received by one selected post or video. The bar chart (Figure 3C) presents the 10 comments with the highest number of likes. The bars' colors represent how the comment was classified (green for positives, red for negatives, and blue for neutrals).

The bar chart in Figure 3D shows the comments "evolution", i.e., each bar contains the total number of positive, negative, and neutral comments received each day over time. To facilitate the analysis, the videos and posts that reverberated for long periods have their comments grouped by week (one or two) as needed.

At the bottom of the dashboard, two word clouds are presented: one for the most frequent words in the positive comments (Figure 3E); and the other for the most frequent words in the negative comments (Figure 3F).

Besides the two dashboards to examine posts, the proposed approach also provides a statistical board regarding the collected data. This board is divided into two parts: one is fixed and shows statistics data considering the database as a whole; the other shows the statistics according to the filters selected on the dashboard. For example, if 2020 is set as a filter on

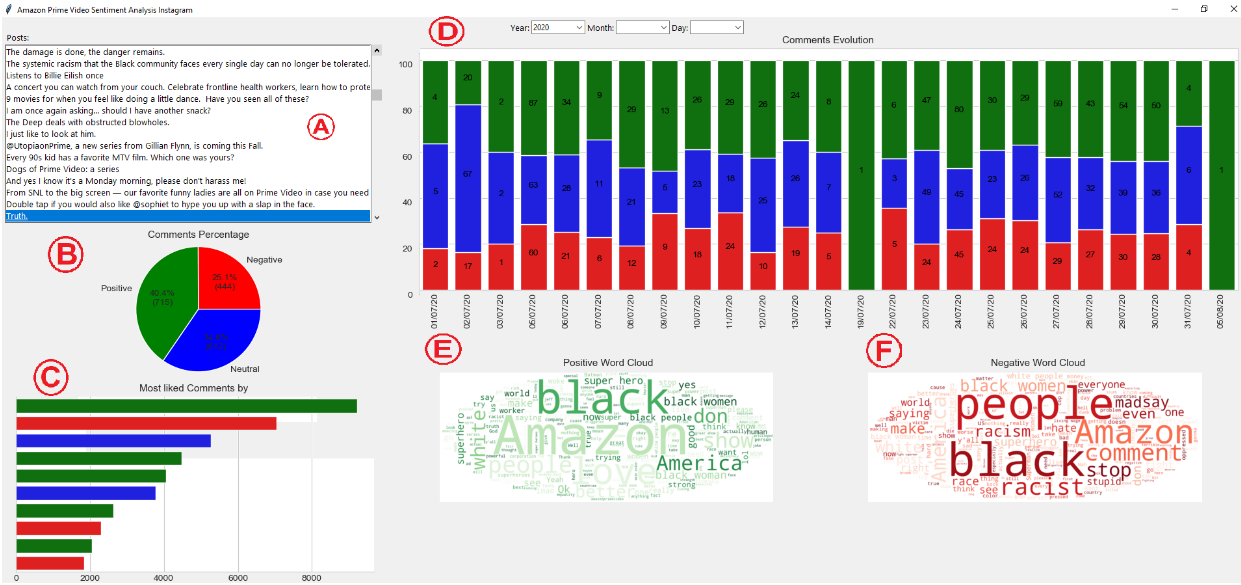

Figure 3. The sentiment analysis dashboard of our visual analysis approach's: (A) post list; (B) the total number for positive, neutral, and negatives comments for the post selected in (A); (C) the rank of the most liked comments; (D) the number of positive, neutral, and negatives by day over time; (E) word cloud of the positive comments; and, (F) word cloud of the negative comments.

the dashboard, it will show statistics just for 2020. The provided statistics differ according to social media, but they generally consider: the number of followers; the number of following; the total number and mean of posts, likes, replies, and comments (Figure 4).

4. Case Study

To analyze our approach, we applied it in two case studies, with Netflix, and with Amazon Prime Video. These two companies were selected because both are very active on social media and belong to the same segment, enabling comparisons between them.

4.1. Netflix

Netflix is known to have close contact with its users through social media, confirmed by the high number of user interactions. Netflix also has more than 50 million subscribers[6]. Netflix has joined Twitter in October 2008, Instagram in August 2012, and YouTube in July 2012.

To the analysis presented, the Netflix data were collected as follow: **Twitter** from 03/10/2008 to 15/09/2020; **Instagram** from 13/08/2012 to 03/11/2020; **YouTube** from 17/06/2012 to 03/11/2020.

The main dashboard from our approach allows the identification of points of interest. For instance, with the collected Netflix data, comparing the bar charts for the number of tweets with the ones for the number of likes is possible to notice that the number of likes for tweets got a huge increase from 2016 to 2017, but the number of tweets decreased in the same period. Relating this information with the statistics data is possible to notice that in 2016 Netflix tweeted 3097 times but only got an average of 344 likes, 10 replies, and 105 retweets. In contrast, in 2017, the company tweeted 2,051 and received an average of 2,940 likes, 65 replies, and 958 retweets. Verifying the statistics for the following years, we can observe the same phenomenon for 2018, 2019, and 2020. From these years, 2018 had the higher number of tweets but the lowest number of interactions. Considering this information, we can conclude that a high number of posts does not necessarily translate into a high number of interactions for Twitter.

Although this is true for Twitter, the dashboard did not indicate that this aspect would be replicated for other social media. Verifying the statistical data for Instagram and YouTube is possible to observe that, generally, the higher number of posts, the higher the number of interactions. There was only one exception between 2016 and 2017 for Instagram when the number of posts decreased (from 257 to 138), but the interaction number average increased (from 22,898 to 86,337 likes and from 658 to 2,225 comments).

Since our dashboard allows comparisons among social media is possible to notice that the largest number

[6]Data collected in December 2020 considering the three social media combined.

Twitter Filters Year 2019		Instagram Filters Year 2019		YouTube Filters Year 2019	
Tweets Posts Total:	2.616	Instagram Posts Totals:	744	YouTube Total Posted Videos	1.010
Tweets Likes Total:	9.734.737	Instagram Likes Total:	278.693.077	YouTube Likes Total:	21.028.438
Tweets Likes Mean:	3.721	Instagram Likes Mean :	374.587	YouTube Likes Mean:	20.820
Tweets Replies Total:	264.269	Instagram Comments Total:	2.490.155	YouTube Comments Total:	1.478.641
Tweets Replies Mean:	101	Instagram Comments Mean:	3.347	YouTube Comments Mean:	1.464
Retweets Total:	1.839.751			YouTube Views Total:	962.319.253
Retweets Mean:	703			YouTube Views Total Mean:	952.791
				YouTube Dislikes Total:	714.360
				YouTube Dislikes Mean:	707
Twitter Filters Year 2020		**Instagram Filters Year 2020**		**YouTube Filters Year 2020**	
Tweets Posts Total:	2.833	Instagram Posts Totals:	950	YouTube Total Posted Videos	907
Tweets Likes Total:	9.049.035	Instagram Likes Total:	371.780.403	YouTube Likes Total:	20.055.984
Tweets Likes Mean:	3.194	Instagram Likes Mean :	391.348	YouTube Likes Mean:	22.112
Tweets Replies Total:	265.056	Instagram Comments Total:	2.527.621	YouTube Comments Total:	1.532.612
Tweets Replies Mean:	94	Instagram Comments Mean:	2.661	YouTube Comments Mean:	1.690
Retweets Total:	1.679.112			YouTube Views Total:	831.663.985
Retweets Mean:	593			YouTube Views Total Mean:	916.939
				YouTube Dislikes Total:	2.772.950
				YouTube Dislikes Mean:	3.057

Figure 4. Netflix's statistics data for 2019 and 2020.

of posts occurs on Twitter, but Instagram and YouTube generate more interactions. Even if we compare the values by their average, Instagram and YouTube have much higher values for likes and comments than Twitter.

The Twitter heatmap chart shows that the two most-liked and retweeted posts of 2020 were about the "Black lives matter" movement and support for the black community. Even if we compare with the other years' data, they remain in the top positions. The only exception is a post from 2017 about "Net neutrality"[7] that had slightly more likes than the one about black community support.

On the other hand, the high likes and retweets number did not reflect a huge number of replies. Although these two posts occupy the second and third positions at the top liked posts, the post with the highest number of replies was regarding the movie "Cuties" (Figure 2E). This post had more than twice the number of replies as the second one.

Since the heatmap chart showed a high number of responses for a tweet about Cuties', it quickly became a point of interest. By glancing over the respective tweet, we could notice that the tweet was posting an apology. Thus, we looked for associated mentions of the film in other charts, from which we found out that the YouTube video with the trailer garnered the highest number of dislikes in our data. Factoring the year 2020, this single video tripled the average number of dislikes from less than 1,000 in the previous years to 3,057. As such, the data associated with the Cuties' film tell a story of controversy, which seems rooted in marketing choices on Netflix's part. The choice of visuals for the film poster in Netflix's release depicted a more provocative tone when compared to the original. This poster, in particular, had been targeted by netizens, as they accused Netflix of sexualizing young girls and promoting pedophilia[8]. An apology was issued, and the poster was changed in response to the backlash, but the measures did not seem to successfully contain the negative reaction (Figure 5).

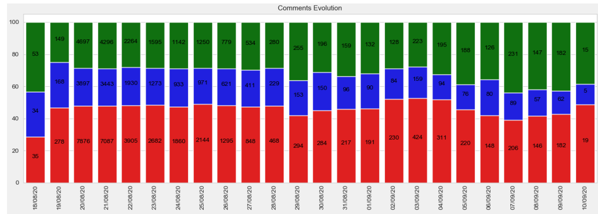

Figure 5. "Cuties" comments sentiment analysis over time.

The same phenomenon can also be observed in 2017 when the announcement of the series "Dear White People" was responsible for raising the average number of dislikes from less than 1,000 to 1,464. This video caused controversy due to its racial themes, being accused of reversed racism. Thus, we can perceive that this brand needs to be aware of posts that involve, for example, racial themes.

4.2. Amazon Prime

Amazon Prime is not as present in social media like Netflix, but it also has many followers and tries to build a close relationship with the public. Amazon Prime debuted on Twitter in October 2008, Instagram in November 2016, and YouTube in February 2014. To the analysis here presented, the Amazon Prime data were collected as follows: **Twitter** from 07/11/2008 to 29/08/2020; **Instagram** from 16/11/2016 to 06/09/2020; **YouTube** from 06/02/2014 to 08/09/2020.

Amazon Prime data presents some of the same aspects as Netflix. For instance, analyzing the dashboard's bar charts for posts number and the number of likes is possible to notice that, regarding Twitter, a high number of posts does not translate in a high number of interactions. The statistical visualization enhances this finding, showing, for example, that in 2017, Amazon Prime tweeted 38,829 times, but the average of likes and replies was inferior to two. The same aspect is also observed in the following years. Examing the Instagram and YouTube charts on the dashboard is also possible to notice that they present similar characteristics to Netflix. On those social media, the number of posts influences the number of interactions. So the higher the number of posts, the

[7] Net neutrality is a movement that believes that "owners of the networks that compose and provide access to the internet should not control how consumers lawfully use that network, and they should not be able to discriminate against content provider access to that network" according to Gilroy[26]

[8] https://tinyurl.com/mh7fxjv6

higher the number of interactions. The only exception was between 2017 and 2018 for YouTube, when the number of posts decreased (from 57 to 45), but the average number of likes increased from 1,186 to 2,970.

The dashboard showed that 2017 was a year of interest for Amazon Prime due to the much higher number of posts it presented. Enhancing the data for 2017 showed that in October 2017, Amazon Prime did a pizza promotion on Twitter. This promotion is responsible for the majority of Twitter's interaction for this year. If we check the replies, it is possible to notice that several of them are from Amazon Prime itself, congratulating the winner customers or consoling losers (Figure 6).

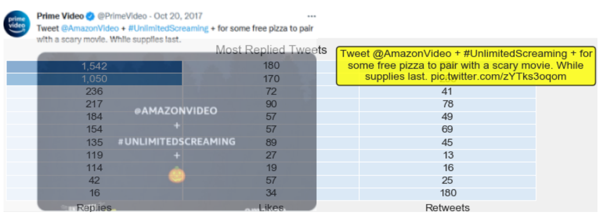

Figure 6. Amazon Prime Pizza Promotion tweets.

Examining the dashboard showed that the high number of interactions manifested on Twitter for 2017 was an exception. Typically, the high number of interactions happens on YouTube and Instagram, even though Twitter possesses the highest number of posts.

Verifying the heatmaps charts is possible to perceive that, for the year 2020, one post on Twitter and another on Instagram had much more likes than the others. Exploring these posts more deeply, it is possible to observe that both deal with racial issues, supporting the black community. So this can be considered an engaging topic. But once more, those posts did not have high numbers of comments.

Since Netflix presented outliers disliked videos, we decided to investigate if the same phenomenon occurs for Amazon Prime. For Amazon Prime, it was not possible to find clear outliers like the ones presented for Netflix. The video with the higher number of dislikes was the second trailer released for the "Tom Clancy's Jack Ryan" TV series. But this video was removed from the platform, so we could not analyze the motives for the removal. The first trailer for the series did not generate a high number of dislikes.

5. Discussion

By analyzing the case studies, we can spot several similarities between Netflix and Amazon Prime data. The first is that the interactions are highly concentrated among individual high-performing posts rather than

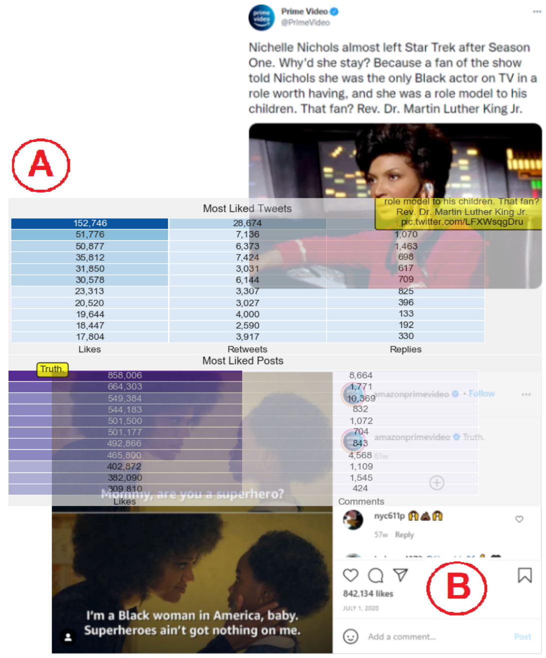

Figure 7. Amazon Prime most liked tweets (A) and Instagram posts (B) for 2020.

spread evenly across the data. This can be inferred from the scenarios in which a few select posts garner many likes and comments beyond the average number of interactions for each post. This phenomenon is replicated between both brands across different social platforms. For future work, we foresee the possibility of extending the analysis of posts to discover common factors between them and predict which ones will maximize the attention and engagement of those posts.

Among all the analyzed social media, Instagram has consistently had the highest number of interactions. By observing Instagram data, we can easily spot a much higher interaction growth when compared with other social media. One possible explanation is the different dynamics in the way these platforms operate. As opposed to Twitter's real-time updated stream of tweets, posts on Instagram remain on users' feeds for a much longer time, possibly enabling further interactions in scenarios where tweets would likely be missed by users. Moreover, when compared to YouTube, Instagram is still a more fast-paced environment. While on YouTube, each post is necessarily a video, which requires a higher time commitment for the user, given a video may take several minutes to be seen, on Instagram, each post comprises one or more pictures, which are visualized and read instantly by its users. We speculate that these differences have come to benefit Instagram regarding interaction numbers, but a more detailed analysis on exactly which factors enable this phenomenon is warranted for future works.

Lastly, another valid observation is that posts with high likes did not always translate into many comments and replies. This fact can also be observed in posts with a high number of replies but a relatively low number of likes. Our analysis verified that posts with a high number of likes generally meant public support regarding the posts and their contents. On the other hand, controversial topics generated many responses but a much lower number of likes. Therefore, we can infer that likes on social media are generally used to express support, but a higher number of replies can indicate controversy. In fact, the communities lingering in online spaces have noticed the dynamics associated with the balance of likes and replies, giving birth to the term "ratioed". Minot et al. [27] describe the ratio value in Twitter as the balance of replies to likes and retweets, meaning posts mustering a high number of replies, yet low counts of likes will generally work as an indicator of controversy. Hence, the "ratio". While the controversial tweets in our data verify the poor balance between replies and likes, we can also note that the ratio is not a phenomenon exclusive to Twitter, given the contentious YouTube videos in our data also sported an exceptionally high number of comments.

As the proposed visual analysis approach focuses on engagement, we argue that each of the evaluated metrics is able to show insights and pinpoint specific situations about how customers engage with and perceive brands in these online environments. For example, likes are strongly tied to notions of positive perception, such as brand loyalty. Algharabat [28] argues that the act of liking brands in social media is an indicator of brand love, in which by engaging with the brand, consumers then associate the brand as part of their online self-expression. On the other hand, the dynamics expressed by the fine balance between likes, shares, and replies can be an indicator of controversy and backlash. The proposed visual analysis approach enables close monitoring of those metrics over time while enabling quick browsing of the contents in the data, making for quick identification of posts perceived positively and negatively by the wider audience across multiple social platforms. We associate notions of customer engagement with these readings that can be driving points towards brand perception. Moreover, the analysis of the brands of our case studies, both from the same area, demonstrates that our approach allows us to obtain others perceptions, such as the content of a post is more important than the quantity of the posts; Instagram posts generate more interactions; it is important to avoid publishing controversial topics.

Therefore, we developed a visual analysis approach to collect, explore, and analyze data from three different Social Media. The provided interactive visualization techniques might be simple, but they allow a direct analysis by any users. Furthermore, the presented case studies show that it is possible to quickly note when there are a number of interactions bigger than normal and, mainly if negative comments predominate. The visual approach also provided ways to analyze the repercussions of a given situation on different social media, as in the trailer for the movie "Cuties" on Netflix's YouTube, which led to an apology on Twitter. This analysis was only possible because our approach allows you to easily do a visual analysis on multiple networks simultaneously. Consequently, it provides a tool to be used by brands to improve their social media communication, generating more engagement and avoiding backlash.

6. Conclusion

Social media has become an important information source. Brands can benefit from using social media data to obtain quick feedback for a product or service, especially nowadays, when the world is very dynamic, and perceptions can change in a matter of seconds.

Therefore, we developed an interactive visual analysis approach that provides a well-defined pipeline and simple but powerful visualization techniques that enable a direct analysis by any user without programming knowledge. The value of our approach is exemplified by two case studies, which provided valuable insights from the analyzed brands. To the best of our knowledge, this is the first work available at GitHub to analyze three social media together and provide comparisons between them. We also contribute with an easy-to-use script for YouTube data collection.

Our approach can be a rich data source and analysis tool for marketing and communications researchers, who usually do not have programming skills. Regarding computer science research, it raises interesting research problems as predicting algorithms for post engagement and big data streaming processing. Moreover, we aim to use this study as a start point for researches that include demographic and regional characteristics in the analysis.

To improve our approach, we aim to develop a classification method that shows the mutual characteristics of the posts with higher interaction numbers and, therefore, predicts the ones that generate more engagement. Also, we aim to provide a real-time system analysis that can be configured, by the user, to issue alerts in specific scenarios (e.g., a high number of negative comments). Finally, we intend to interview brand managers to get feedback about other features we should consider for extending our approach.

References

[1] A. K. Kirtiş and F. Karahan, "To be or not to be in social media arena as the most cost-efficient marketing strategy after the global recession," *Procedia - Social and Behavioral Sciences*, vol. 24, pp. 260–268, 2011.

[2] E. Constantinides, "Foundations of social media marketing," *Procedia - Social and Behavioral Sciences*, vol. 148, pp. 40–57, 2014.

[3] A. Arvanitidis, A. Serafi, A. Vakali, and G. Tsoumakas, "Branty: A social media ranking tool for brands," in *Proceedings of Machine Learning and Knowledge Discovery in Databases*, pp. 432–435, Springer Berlin Heidelberg, 2014.

[4] X. Liu, A. Xu, L. Gou, H. Liu, R. Akkiraju, and H. Shen, "Socialbrands: Visual analysis of public perceptions of brands on social media," in *Proceedings of 2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 71–80, IEEE, 2016.

[5] S. Singh, A. Chauhan, and S. Dhir, "Analyzing the startup ecosystem of india: a twitter analytics perspective," *Journal of Advances in Management Research*, vol. 17, pp. 262–281, 2019.

[6] H. Luan, J. Li, M. Sun, and T.-S. Chua, "The design of a live social observatory system," in *Proceedings of the 23rd International Conference on World Wide Web*, p. 1025–1030, ACM, 2014.

[7] Y. Gao, Y. Zhen, H. Li, and T. Chua, "Filtering of Brand-Related Microblogs Using Social-Smooth Multiview Embedding," *IEEE Transactions on Multimedia*, vol. 18, pp. 2115–2126, 2016.

[8] S. Qi, F. Wang, X. Wang, J. Wei, and H. Zhao, "Live multimedia brand-related data identification in microblog," *Neurocomputing*, vol. 158, pp. 225–233, 2015.

[9] A. Aggarwal and A. Singh, "Geo-localized public perception visualization using glopp for social media," in *Proceedings of 2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 439–445, IEEE, 2017.

[10] E. Fernández-Gómez and J. Martín-Quevedo, "La estrategia de engagement de netflix españa en twitter," *El profesional de la información (EPI)*, vol. 27, pp. 1292–1302, 2018.

[11] A. Lakhiwal and A. K. Kar, "Insights from twitter analytics: Modeling social media personality dimensions and impact of breakthrough events," in *Proceedings of Conference on e-Business, e-Services and e-Society*, pp. 533–544, Springer, 2016.

[12] L. Matosas-López and A. Romero-Ania, "The efficiency of social network services management in organizations. an in-depth analysis applying machine learning algorithms and multiple linear regressions," *Applied Sciences*, vol. 10, pp. 1–16, 2020.

[13] H. Lakkaraju and J. Ajmera, "Attention prediction on social media brand pages," in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pp. 2157–2160, ACM, 2011.

[14] M. Mazloom, R. Rietveld, S. Rudinac, M. Worring, and W. van Dolen, "Multimodal popularity prediction of brand-related social media posts," in *Proceedings of the 24th ACM International Conference on Multimedia*, pp. 197–201, ACM, 2016.

[15] N. Kumar, G. Ande, J. S. Kumar, and M. Singh, "Toward maximizing the visibility of content in social media brand pages: a temporal analysis," *Social Network Analysis and Mining*, vol. 8, pp. 1–14, 2018.

[16] G. Overgoor, M. Mazloom, M. Worring, R. Rietveld, and W. van Dolen, "A spatio-temporal category representation for brand popularity prediction," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pp. 233–241, ACM, 2017.

[17] A. S. de Campos Filho, F. Freitas, A. S. Gomes, and J. Vitorino, "Brandmap: An information visualization platform for brand association in blogosphere," in *Proceedings of 2012 16th International Conference on Information Visualisation*, pp. 316 – 320, IEEE, 2012.

[18] K. Kucher, R. M. Martins, C. Paradis, and A. Kerren, "Stancevis prime: visual analysis of sentiment and stance in social media texts," *Journal of Visualization*, vol. 23, pp. 1015–1034, 2020.

[19] J. P. Scherer, "Visual analysis approach for brands perception on social media," Master's thesis, Pontifical Catholic University Of Rio Grande Do Sul, 2021.

[20] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, pp. 216–225, 2014.

[21] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *The craft of information visualization*, pp. 364–371, Morgan Kaufmann, 2003.

[22] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky, "Guidelines for using multiple views in information visualization," in *Proceedings of the Working Conference on Advanced Visual Interfaces*, p. 110 – 119, ACM, 2000.

[23] J. Gray, L. Bounegru, S. Milan, and P. Ciuccarelli, *Ways of Seeing Data: Toward a Critical Literacy for Data Visualizations as Research Objects and Research Devices*, pp. 227–251. Springer International Publishing, 2016.

[24] C. D'Ignazio, "Creative data literacy: Bridging the gap between the data-haves and data-have nots," *Information Design Journal*, vol. 23, pp. 6–18, 2017.

[25] M. Galesic and R. Garcia-Retamero, "Graph literacy: A cross-cultural comparison," *Medical Decision Making*, vol. 31, pp. 444–457, 2011.

[26] A. A. Gilroy, "Net neutrality debate: Access to broadband networks," 2017. `https://digital.library.unt.edu/ark:/67531/metadc1042344/`. Last accessed in November 2020.

[27] J. R. Minot, M. V. Arnold, T. Alshaabi, C. M. Danforth, and P. S. Dodds, "Ratioing the president: An exploration of public engagement with obama and trump on twitter," *PLOS ONE*, vol. 16, pp. 1–22, 2021.

[28] R. S. Algharabat, "Linking social media marketing activities with brand love," *Kybernetes*, vol. 46, pp. 1801–1819, 2017.