Political Polarization and Movie Ratings: Web Scraping The Brazilian Contemporary Scenario

Ana Paula Moritz[1], Bruno Chagas[2]

[1]Department of Social Sciences, Pontifical Catholic University of Rio de Janeiro, Brazil,
[2]Department of Computer Science, Federal University of Minas Gerais, Brazil.

Abstract

Our work aims to analyze the impact of political polarization on movie ratings at the IMDb platform. For that we explore the concepts of Word of Mouth and Buzz marking the important role they play on polarized opinions in movie ratings.We develop a code on python to perform web-scrapping on the sample scope of Brazilian movies and interpret the data collected using a controversiality index based on standard deviation. The outcome sheds some light into the relation between Buzz and Controversiality within the framework of Brazil's current political scenario.

Keywords: *Political Polarization; Movie Ratings; Web-Scrapping; Brazil; Controversiality; Buzz.*

1. Introduction

The idea for this paper came with the urge to explore data related to polarization in public opinion, specifically when it comes to cultural products. In order to do so, we have chosen to web-scrap the International Movie Database (IMDb) platform in search for the most controversial movies in Brazil. Our idea is to point out how political opinions are associated with a controversiality index.

Some papers in this area have already been written, but none of them refer to the political aspects of polarized opinions in movie ratings. In Fischoff, Antonio, Lewis (1998), the authors explore favorite films and film genres as a function of race, age and gender; in Wühr, Lange, Schwarz (2017) the authors compare gender stereotypes with actual movie preferences; Oliver et al. (1998) observe the impact of sex and gender role self-perception in reactions to different types of movies; Koh, Hu, Clemons (2010) investigate movie reviews online and mark the cultural aspects of quality perception; Otterbacher (2012) analyzes writing styles and metadata features on movie review forums focusing on gendered comments. All referred papers examine a variety of identity markers to explain partiality in movie preferences, but the political aspects of polarization is something we believe should also be taken into consideration. It's still unclear to us whether this phenomenon happens only in Brazil, or if it was overlooked by previous research.

In Brazil, polarization has been a big part of society's everyday life since 2013, approximately. The country went through a powerful but somewhat diffuse street movement that claimed to be apolitical. The outcomes of this movement prove otherwise. There was a fracture in society that had just been exposed and, since then, the country was never the same. After an impeachment in 2016 -considered by some as a coup d'état- and electing a far-right, ultra-conservative president in 2018. Brazil currently has its most conservative and religion-oriented government. In case readers want a deeper analysis on the Brazilian political situation, we suggest Chagas-Bastos (2019) and Pinheiro-Machado (2014).

We observed the controversy generated by some movies in social media, in order to place and compare them with the data we've collected. Our methodological approach to this issue was to develop a code for web-scraping the IMDb page, we searched for movies with the highest numbers of votes, then analyzed the normalized standard deviation, correlated the results with our hypotheis and conceptualized it with applicable literature. In section 2 we describe the concepts of eWoM and Buzz and their association with polarized opinions as well as deepen our explanation of the web scraping process and controversiality index. In section 3 we present our results and analyze them further, connecting data, theory and scenario. In the last section we present our conclusions and remarks and also make suggestions on how to broaden this research.

2. Theoretical Background

In order to explain our hypothesis, we explored the concepts of Word of Mouth (WoM) and Buzz, as well as their influence on polarized opinions and movie ratings. In the first part of this section we demonstrate how they work and are instrumented in order to mobilize voters - by doing so we give life and meaning to the statistical facts. As for the second part of this section, we present our method for constructing the code on python, and how we applied it in our analysis. We also present the concept of hard controversiality that verses on love-hate movies and how it frames polarization in movie ratings.

2.1. eWoM & Buzz

The concept of WoM is defined by Richins (1984) and Sundaram (1998) as a way people influence each other through communication. Liu (2006) and Mohr (2007) apply this concept to the movie industry addressing box office revenue and marketing strategies for movie premieres. We apply the concept as an informal influence, exerted by opinion leaders that is based on interpersonal relations in an online environment, it is named eWoM.

As stated by Moon (2010), eWoM can act positively or negatively, depending on the movie and its public, the author enumerates five factors that should be observed when evaluating movie ratings and recommendations such as (1) number of ratings, (2) average rating, (3) rating standard deviation, (4) percentage of highest rating, (5) percentage of lowest rating. First, the author explains that the accumulated number of ratings would indicate how many people have seen the movie, as we will demonstrate further in our analysis, this behavior doesn't occur in the data we have collected. Second, the average rating means how viewers rate good movies high and bad movies low, but we can also analyze ratings by a evaluation of disagreement towards the same movie given by the standard deviation.

Hennig-Thurau et al. (2004) categorized what motivates people to articulate their opinion online, the main features of eWoM presented in this paper are: focus-related utility, consumption utility and approval utility. The first, and most important to our hypothesis is related to adding value to community, it could be interpreted as helping a political party, a politician or an ideology. The ones that are motivated by this, expect social and economic benefits out of their actions. The second consists of using the websites to give opinions and gather information about products, but it does not apply to our case, since IMDb doesn`t carry this interactive feature for its users. The third one is about self-enhancement and economic reward; users expect to be recognized and praised for their opinions and in some cases, users receive compensations as tokens of appreciation by the reward giver. Moreover, the authors also indicate some discursive practices such as vengeance, boycott, altruism, exertion of power and frustration.

The definition of Buzz in Mohr (2007) consists on "the practice of gathering volunteers formally either by actively recruiting individuals who naturally set cultural trends, or informally by drawing 'connectors': people who have lots of contacts in different circles, who can talk up their experiences with folks they meet in their daily lives". Some of the profiles the author refers to are: experts, members of the press, politicians, celebrities, or well-connected customers others rely on for information. In the Brazilian scenario we can observe different cultural patterns such as evangelical leaders influencing entire congregations to give bad reviews on a movie they believe is threatening to their faith or political militants who are often hired to spread Buzz and count with the help of bots on social media platforms in order to reach more people faster.

We noticed that previous literature doesn't cover the political component present on eWoM and Buzz, we argue that these concepts per se do not spawn polarization, there is an exterior factor that when injected into the debate sparks a fire. Kostakos (2009) evidences a strong bias when it comes to the crowd's wisdom, and also in the design of social media platforms, as demonstrated by Balasubramanian and Mahajan (2001) in virtual communities. An important feature of the Brazilian scenario is the extent of social media usage for debates and comments, and how strongly polarized opinions dispute control over a narrative that might influence the public's opinion.

2.2. Web-Scrap and Controversiality

We constructed a code on python language that uses the proper packages for web-scrap. Our purpose is to evaluate and analyze some relevant statistical components over the IMDB platform, which allowed us to study controversiality and engagement properties. We considered movies with at least 900 ratings, only brazilian productions, spoken in portuguese. Our sample performed a total of 214 feature movies with around 1,8 million individual ratings. The data was collected by the end of February 2020.

IMDB platform displays just the mean value of a specific movie. Although, as seen in Hoßfeld, Schatz and Egger (2011), solely the mean value causes a "loss of important information on the user rating diversity". In order to retain more information, we used a statistical component called normalized standard deviation or hard controversiality index proposed by Amendola, Marra and Quartin (2015) which allowed us to seek love-hate movies. We will call it by C throughout our text. Therefore, we can write the controversiality index C in the form:

$$C = \frac{1}{C_{max}} \left[\sum_{i=1}^{10} p_i (r_i - \bar{r})^2 \right]^{1/2}$$

where the index i varies over the number of ratings, r_i is the rating i, p_i the percentage of ratings r_i and $\overline{r}$ is the mean value and Cmax is the maximum value that C can be achieved. The maximum number is 4,5, and we can see this effect when the polarization is maximum: 50% of ratings in 1 and 50% in 10 and this is the maximum love-hate for a movie.

For our next section we will analyze some aspects of C and its relation to the number of individual ratings on a movie. We argue that there is an engagement factor, in which high values of C and ratings indicate some kind of mobilization. Furthermore, they shed some light on how WoM and Buzz are correlated to C and the amount of ratings.

3. Results and Analysis

We start our analysis illustrating the phenomenon of controversiality and polarization on movie ratings with the help of figure 1(a) and figure 1(b). As we can see, the first one depicts a highly controversial movie with C≅0.84, while the second has a normal (or expected) behavior with C≅0.35. We stand that all movies with a political component, internal or external, in the current Brazilian political scenario, are subjected to biased public scrutiny and we may say that this is a Buzz-oriented phenomenon. It is important to point out that this pattern repeats itself .

We highlight that figure 1(a) is our key example, and its characteristics permeate the other movies we have analyzed. A peculiar trait of this movie is that its premiere in 2019 was censored in Brazil and it is expected to be released in May 2020. Here we outline the main question of our paper: how come a movie that wasn't even released had already the 5th largest number of votes and the 4th highest C? Therefore, we assert that the political component in movie ratings, as described by us, is often overlooked throughout the existing literature.
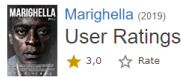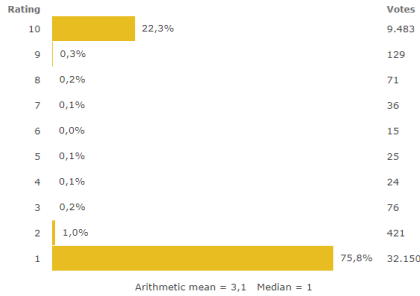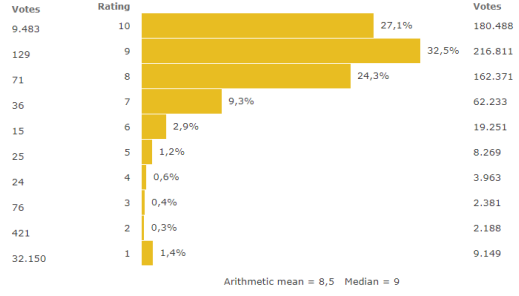
(a)Marighella (2019)   *(b) Cidade de Deus (2002)*

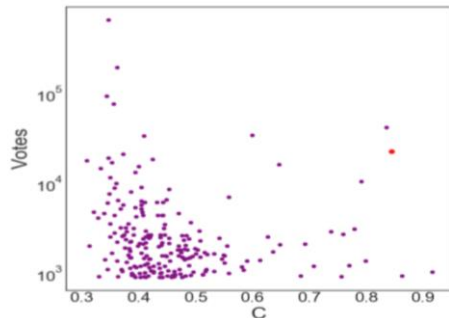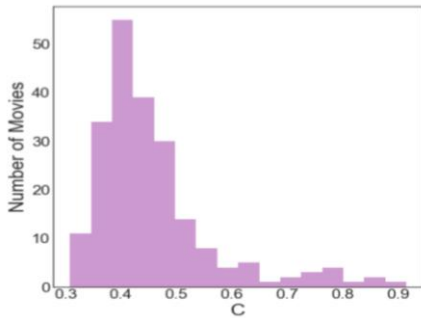*Figure 1.Statistical data from two brazilian movies. Source:IMDb (2020).*

Figure 2. Number of movies and votes as function of C on our web-scrap.

Figure 2 on the left depicts that C the number of movies do not converge to zero while C is increasing as normal distribution. Moreover, we can notice a bump around 0,8 and 0,6 along the C axis, pointing out some external cause to it. Figure 2 on the right shows us that there are some movies that have a higher C and number of votes, we will analyze as organized in table 1.

Table 1. Most controversial movies with higher number os votes.

Movie (Year)	Votes	C	Movie (Year)	Votes	C
Marighella (2019)	*42.395*	*0,84*	*The First Temptation of Christ (2019)*	*2.755*	*0,76*
Nothing to Lose (2018)	*22.777*	*0,84*	*Lula, O Filho do Brasil (2009)*	*2.140*	*0,70*
Aquarius (2016)	*16.412*	*0,65*	*Nosso Lar (2010)*	*2.568*	*0.63*
The Edge of Democracy (2019)	*10.582*	*0,80*	*Marighella (2012)*	*2.954*	*0,74*
O Mecanismo (2018)	*34.753*	*0,60*	*Polícia Federal: A Lei é para Todos (2017)*	*3.154*	*0,70*

We have organized Table 1 according to the highest values of C and the highest number of votes in order to evidence where the cluster of polarization occurs. It is noticeable that the movies that generate the highest C and mobilize a larger number of votes are quite recent. Our hypothesis is that the political turmoil that Brazil went through in 2013 led the country into a generalized social and political instability which affected people's perspectives regarding cultural products in general, and more specifically as we show, movies. We sustain our hypothesis with Aquarius, the movie was nominated for a Palme d'Or in Cannes, and at the award ceremony the cast protested against the 2016 coup d'état - this action is what sparked the buzz - then the movie and cast were fiercely criticized which granted the movie its controversial status.

Returning to our analysis on Hennig-Thurau et al. (2004), we would like to recodify and broaden the utility types for engaging. When the authors categorize a motive as "adding value to community" and the person is motivated by the idea of helping the company, we can interpret that as a portion of society that wants to help their country, ideology or political party, in short, their motivation is partisanship. People who identify with the right-wing, envision the country as a corrupted institution and want to save it by restoring its traditional moral values. People who identify with the left-wing are trying to restore what the country once was during the leftist governments. To that end, what Hennig-Thurau et al. (2004) call altruism, here we can describe as an ideological component.

This ideological component is the answer to our analysis of the correlation between eWoM and Buzz with C. The movies listed below are in some way associated with either the right-wing (O Mecanismo, Polícia Federal: A Lei é para Todos and Nothing to Lose) or the left-

wing (The Edge of Democracy, Lula, O Filho do Brasil, Marighella, Aquarius and The First Temptation of Christ). The only outsider is Nosso Lar, a movie based on spiritism, a minoritarian religion in Brazil that suffers with persecution and intolerance.

4. Conclusion and Remarks

In summary, we have presented evidence on the correlation between the concepts of eWoM and Buzz with the controversiality index. The proportion and motivation of public engagement is boosted by the influence of the Buzz in social media platforms. Our results indicate that the political and ideological components affect and bias public scrutiny when it comes to movies and ratings. We believe that this paper is relevant for it opens possibilities for deeper analysis of a political and ideological component in movie preferences, something that hasn't been done before. As in present time political disputes tend to become more vicious and intense, we insist on the observance of democratic principles, and affirm the importance of caution with eWoM and Buzz spreaders.

Results so far have been encouraging, we intend on expanding this research in order to fill some gaps, and apply this method to other cultural products and countries where political polarization is on the rise. We also consider expandidng the research to other platforms, following Kostakos (2009) hypothesis that the design of platforms favours biased opinions. We would also like to explore the action of bots and botnets, as it was an element that we weren't expecting on encountering and it prompted some consistent questions as to: who orchestrated their actions and why, if there was any financial gains involved and what's the purpose behind that. Another perspective for this analysis that didn't fit our research scope is text-mining via sentiment analysis and summarization techniques, through that we expect to find the most important or commonly used words, and if they fit a political and ideological context or agenda.

References

Amendola, L., Marra, V., & Quartin, M. (2015). The evolving perception of controversial movies. *Palgrave Commun, 1*.

Balasubramanian, S., Mahajan, V. (2001). The economic leverage of the virtual community. *International Journal of Electronic Commerce, 5(3).*

Chagas-Bastos, F. (2019). Political realignment in Brazil: Jair Bolsonaro and the right turn. Revista de Estudos Sociales [Online], 69.

Dent, A., Pinheiro-Machado, R. (2014). Introduction: the cellularity and continuity of protest in Brazil. Anthropological Quaterly, Vol. 84, No. 3, pp. 883-885.

Fischoff, S., Antonio, J., Lewis, D. (1998). Favorite films genres as a function of race, age, and gender. *Journal of Media Psychology*, Volume 3, Number 1.

Hennig-Thurau, T., Gwinner, K., Walsh, G., Gremler, D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?. *Journal of Interactive Marketing. 18. 38 - 52.*

Hoßfeld, T., Schatz, R., & Egger, S. (2011). SOS: The MOS is not enough! *QoMEX 2011*, September, Belgium.

Koh, N., Hu, N., & Clemons, E. (2010). Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications, 9*.

Kostakos, V. (2009). Is the crowd's wisdom biased? A quantitative analysis of three online communities. 2009 International Conference on Computational Science and Engineering, Vancouver, BC, 2009, pp. 251-255.

Liu, W. (2013). Word of mouth for movies: its dynamics and impact on box office revenue. Journal of Marketing, Vol. 70, No. 3, pp. 74-89.

Mohr, I. (2007). Buzz marketing for movies. *Business Horizons*, 50, 395-403.

Moon, S., Bergey, P., Iacobucci, D. (2010). Journal of Marketing, Vol. 74, pp. 108-121.

Oliver, M., Sargent, S., Weaver, J. (1998). The impact of sex and gender role self-perception on affective reactions to different types of film. Sex Roles, Vol. 38, No. 112.

Otterbacher, J. (2012). Gender, writing and ranking in review forums: a case study of the IMDb. Knowl Inf Syst 35, 645–664.

Sundaram, D., Mitra, K., Webster, C. (1998). Word-of-mouth communications: a motivational analysis. Advances in Consumer Research, Vol. 25, pp. 527-531.

Wanderer, J. (1970). In defense of popular taste: film ratings among professionals and lay audiences. American Journal of Sociology, Vol. 76, No. 2, pp. 262-272.

Wühr, P., Lange, B., & Schwarz, S. (2017). Tears or Fears? Comparing gender stereotypes about movie preferences to actual preferences. *Frontiers in psychology, 8, 428*.