A Systematic Analysis of Depression-Related Discourse Within Facebook: A Comparison Between Brazilian and American Communities

Silas Lima Filho[1,2][a], Mônica Ferreira da Silva[2][b]
and Jonice Oliveira[1,2][c]

[1]*Instituto de Computação (IC), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*

[2]*Programa de Pós-Graduação em Informática (PPGI), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*

fi

Keywords: Depression, Social Media, Social Networks, Textual Analysis.

Abstract: Identifying the symptoms of a depressive disorder can help potential sufferers seek professional help, increasing their chances of recovery. This article presents the operationalization of systems and tools to systematize the analysis process using data from depression-related communities within Facebook. We discuss how we can utilize the data to understand details about depression and the discourse surrounding the disorder through textual analysis using LIWC. The results show a low correlation between textual analysis and the features of social media interaction. This study, through a systematic use of data collection and analysis tools, aims to provide explanatory insights into messages discussing the topic of depression.

1 INTRODUCTION

Computer science has studied textual language, specifically in the natural language processing domain. However, this concern is also shared in the field of health studies. In Castro's work, Lacan's contributions are discussed and compared with other approaches regarding the importance of language and its structuring in the unconscious. As mentioned in the paragraph above, the author discusses various language interpretations. Language is recognized as a manifestation of already formed thoughts or the upshot of an unconscious process (De Castro, 2009).

Social media are potential tools for monitoring populations in epidemic control, information dissemination, and combating misinformation about certain diseases (Skaik and Inkpen, 2021). Public health entities can understand patterns in specific groups or populations through systems that integrate social media information to identify public opinion on services that may not be as good as they should be, identify individuals at risk, and communicate potential urgent diseases (Horvitz and Mulligan, 2015). The use of technology directly supports institutions and professionals, helping to raise awareness of certain diseases. During the COVID-19 pandemic, many patients sought information about the disease on social media (Chen and Wang, 2021).

According to the World Health Organization (WHO), approximately 300 million people of different ages suffer from some level of depression[1]. Depressive disorders are the fourth leading cause of disability and have progressed steadily over the years (Brody et al., 2018) and (James et al., 2018). Major depressive disorder is classified as such when the patient presents a set of predefined symptoms (Association. and Association., 2013), e.g., daily depressed mood, loss of interest in regular activities, weight loss, and insomnia.

Several factors can complicate the diagnosis of depressive disorders, intrinsic to the clinical approach, e.g., costs, longer patient follow-up time by the professional, and the number of patients each professional can assist (Li et al., 2020). In Brazil, there is unequal access to healthcare professionals and facilities, which can make identifying mental disorders challenging. Populations with lower income do not always have easy access to social media [2].

[a] https://orcid.org/0000-0002-5611-7297
[b] https://orcid.org/0000-0003-0951-6612
[c] https://orcid.org/0000-0002-2495-1463

[1] www.who.int/en/news-room/fact-sheets/detail/mental-disorders [accessed 01-03-2023]
[2] http://revistapesquisa.fapesp.br/tempos-de-incerteza/

Filho, S., Ferreira da Silva, M. and Oliveira, J.
A Systematic Analysis of Depression-Related Discourse Within Facebook: A Comparison Between Brazilian and American Communities.
DOI: 10.5220/0012381100003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 2*, pages 466-473
ISBN: 978-989-758-688-0; ISSN: 2184-4305

Given the described scenario, identifying symptoms related to depressive disorders and promptly and discreetly assisting someone who may be a potentially depressive patient can be very useful, both for the patient and the professional. This identification process involves challenges such as correlating the same signs and symptoms of depressive disorders from the clinical environment with the often abundant data from social media. We present a study on the operationalization of tools for the problem of identifying depressive symptoms and an analysis of the correlation between different obtained features. We apply textual analysis through psychological variables to understand and identify possible patterns within the text used by communities discussing depression phenomena. The results show a low correlation between textual analysis and the features of social media interaction.

Next, Section 2 presents related works. Section 3 defines the concepts and supporting tools. Section 4 presents the methodology and research method. Section 5 presents the results obtained. Section 6 discusses the results. Section 7 concludes the work.

2 RELATED WORK

Studies that address the identification of depressive users on social media employ algorithms and pattern recognition techniques, leveraging Natural Language Processing (NLP) to perform a systemic analysis of text in social media posts. They delve into technical aspects of computational pragmatics, incorporating information and knowledge from the Health domain, such as Psychology and Medicine, and utilize psychometric questionnaires (Giuntini et al., 2020).

(De Choudhury et al., 2013) collect data from social media, among the many articles and research on identifying depression in the population using social media information. They employ psychometric questionnaires representing the theory and technique of measuring mental processes as applied in Psychology and Education. In this work, they extracted data from Twitter from individuals with a clinical diagnosis of depression through crowdsourcing. They create a corpus and develop a probabilistic model to detect whether a post indicates depression. (Tsugawa et al., 2015) apply the same methodology but on a group of Japanese users, analyzing the replicability of results from previous studies.

Using a different social media platform, (Park et al., 2015) demonstrates how activities on Facebook are associated with users' depressive states. Observing an increase in the suicide rate among students, they aim to raise awareness of depression issues at the university where the study was conducted. (Andalibi et al., 2017) explore self-disclosure posts on Instagram marked with the #depression tag to understand which disclosures are highly sensitive in this social media. (Li et al., 2016) take a qualitative approach to understand the behavior and describe the Chinese population's understanding of depression. It differs from the previous studies in that it explores post-disclosure without creating a classification model and does not rely on a quantitative epistemology.

An observational study was conducted to understand the interactions between clinically depressed users and their network connections compared to a group of users without depression (Vedula and Parthasarathy, 2017). The authors identify relevant linguistic and emotional signals in social media exchanges to detect symptomatic signs of depression.

3 CONCEPTS

This work adopts a textual analysis approach, consistent with the literature, to find helpful information that can aid in identifying symptoms of depressive disorders. Tools for data capture from social media, text analysis, and statistical analysis were utilized. The Crowdtangle tool was used for data capture, preparation, and preprocessing, which were carried out using Python routines, and the captured message data was selected for textual analysis.

Crowdtangle is a research tool that collects data from public profiles and groups for link verification and post monitoring. It monitors over 7 million verified Facebook public pages, groups, and profiles, 2 million public Instagram accounts, and 20k subreddits. Two post-monitoring approaches are available: defining sources or searching for popular posts. Crowdtangle returns the number of reactions to a post (likes, reactions, and shares) and calculates the interaction rate of a post compared to the interaction history of a page or community [3].

The Linguistic Inquiry and Word Count (LIWC) system is used in related works. It allows for text processing and analysis using a lexicon predefined by dictionaries, where words are categorized into different domains. It was developed for exploratory text analysis, with the premise that daily words carry psychological characteristics, e.g., emotions, beliefs, and habits (Boyd et al., 2022). It provides processing and analysis modules for various purposes. Its primary

[accessed 01-03-2023]

[3] www.crowdtangle.com/resources [accessed 01-03-2023]

analysis module uses dictionaries where each analyzed word is compared with those already defined in the dictionaries, with their pre-established values in different domains. One of the advantages is the enhancement of the original dictionary by including new terms. The Brazilian Portuguese version is limited to an older version of the original dictionary, dating back to 2015.

The primary analysis module allows for the quantification of four aspects. *Analytical thinking* captures the extent to which people use words that suggest formal, logical, and hierarchical thinking patterns. People with low levels of analytical thinking tend to write and think using more intuitive and personal language. High language scores in Analytical Thinking are rewarded in academic environments and correlate with grades and reasoning skills. Language with low scores in analytical thinking is often seen as less formal and rigid and more friendly and personal.

The second aspect is *Clout*, which refers to the relative social status, trust, or leadership that people display through their writing or speaking. The third aspect, *Authenticity*, deals with when people reveal themselves as ”authentic” or honest. They tend to speak more spontaneously and do not self-regulate or filter their words. Examples of texts with low authenticity scores include prepared texts (i.e., pre-written speeches) and texts in which a person is socially cautious. Examples of texts with high authenticity scores tend to be spontaneous conversations between close friends or political leaders with little or no social inhibition.

Finally, although LIWC-22 includes positive and negative tone dimensions, *Emotional Tone* combines these two dimensions into a single summary variable. The algorithm is constructed so that the higher the number, the more positive the tone. Numbers below 50 suggest a more negative emotional tone.

4 METHODOLOGY

This work presents an empirical post-positivist approach (Wohlin et al., 2012), an explanatory perspective on depressive disorders. Data analysis extracts information and patterns that explain the phenomenon and problem addressed without exhausting all possibilities and subject to improvement (Creswell and Creswell, 2018). While the primary goal of related works is the development of artifacts and constructs to identify depressive users on social media, there is room for those that aim to investigate and analyze these phenomena and behaviors (Recker, 2012).

It is challenging to develop computational systems considering the sociotechnical aspect as a relevant factor, not just the technical factor (Boscarioli et al., 2017), considering subjective and humanistic aspects. The sociotechnical factor contributes to the construction of solutions in society (Cafezeiro et al., 2017). Previous research was conducted among healthcare professionals to explore which information is most relevant to them in therapy or a similar process for identifying someone with depression. They assess the relevance of a computational and technological artifact in patient care and how such an artifact could be helpful as support for verifying patients' social media data (Lima Filho et al., 2022).

This research uses the concepts of the Design Science Research (DSR) approach (Wieringa, 2014) as a reference. Given that the literature defines that an artifact should be relevant to domain experts who deal directly with the problem at hand (Pimentel et al., 2020). In this research context, the artifact created aims to be useful for healthcare professionals, such as psychologists or psychiatrists.

Grounded in the Design Science Research (DSR) paradigm with a focus on supporting healthcare professionals, this research adopts a mixed-method approach, aligning with the post-positivist epistemological perspective described by Creswell (Creswell and Creswell, 2018). This paradigm acknowledges the traditional reductionist approach in scientific research, where data acquisition and analysis aim to comprehend the complexity of ”reality,” emphasizing the absence of absolute truth and the inherent susceptibility of evidence to imperfections. In adhering to this methodological stance, the study employs correlation analyses that integrate qualitative and quantitative elements, exploring metrics and variables from analysis tools and scrutinizing specificities in online communities. The analyses conducted herein lay the groundwork for future research focused on addressing the overarching question: ”How to identify symptoms of psychological diseases through social media?”

4.1 Method

The research method is divided into two stages: the data collection stage and the analysis stage. Both stages and their implementation details are described below.

4.1.1 Data Collection

The data were collected from open Facebook communities that directly discuss depression or related topics, obtained through Crowdtangle. Communities discussing depression-related issues in both English and Brazilian Portuguese were sought in September 2022, a period when media actions usually promote mental health care.

Crowdtangle allows the filtering of the type of posts to be selected, such as video posts, images, and text. Textual content posts of the "status" type, where the user generates textual content, were selected. The search was limited to open groups only. In this type of search, the tool excludes paid advertisements, verified profiles, and Facebook pages, which may represent companies.

Initially, depression-related terms were defined based on works that used the same approach after an extensive systematic literature review. The terms defined in Brazilian Portuguese and English were, respectively, in Brazilian Portuguese: "quero morrer (*I want to die*)", depressão (*depression*), deprimid (*depress*), depressiv, angustia (*anguish*); in English: depression, depressed, depressive, anguish, distress.

To refine the search, the focus was on groups' status that directly discussed depression. There are a total of four English communities that objectively discuss depression, totaling 835 posts. Seven Portuguese-language communities, including 1945 posts, are divided into four communities that directly discuss depression, two on psychological treatment, and one on mourning.

4.1.2 Data Analysis

Relevant attributes from the data collection were selected as **continuous**: "Post Views," "Total Views," "Post Created," "Post Created Date", "Post Created Time", "Total Interactions", "Likes", "Comments", "Shares", "Love", "Wow", "Haha", "Sad", "Angry", "Care", and "OverperformingScore"; or as **categorical**: "Type", "Group Name", "User Name", "Page Category", "Facebook Id", "Message", and "URL".

The categorical attribute that identifies the post author is null; therefore, it does not allow system users to access such sensitive data. Only the "Message" attribute was selected to analyze the content of these posts. An approach adopted in this method was the translation of community messages from Portuguese to English. Thus, the exact version of the lexical was applied to both communities. Given that the most recent version of the Portuguese dictionary is from 2015. This would make using the metrics mentioned in Section 3 impossible. After proper selection and export to a .csv file, the data set with only the index and message text was analyzed in the LIWC tool. Classifications of the most commonly used words in each community were generated, and the attributes of the main analysis module of the tool were also analyzed. The step for processing textual data, i.e., preprocessing, was performed in the LIWC system.

With the values of the attributes described in Section 4.1.2, Pearson and Spearman correlation measures were applied between the continuous variables obtained by Crowdtangle and the attributes obtained by LIWC. The Pearson correlation measure is a parametric measure, while Spearman is non-parametric. Both generate values between -1 and +1. A value of -1 indicates a negative correlation, meaning that when one variable increases, the correlated variable decreases. A positive correlation, or a correlation value of +1, indicates that changes in one variable affect the behavior of the other. A value closer to 0 implies a weak or nonexistent correlation. Pearson's correlation is used for linear data, while Spearman is used for non-linear data.

4.2 Ethical Aspects

The Certificate of Presentation and Ethical Appreciation (CAAE) assigned by the Research Ethics Committee (CEP) is 54865821.5.0000.5263. The data do not identify users since the collection system provides data sets without user identification, and no additional sensitive user data was collected.

5 RESULTS

The first data analysis involved quantifying the frequency of word usage to discover the most commonly used words in each community. Figure 1 presents the most repeated words in Brazilian Portuguese communities. The highlighted words in Figure 1 are irrelevant since no preprocessing was applied to the Portuguese words, as the LIWC system does not have data processing for this language.

Figure 2 represents the same community but with an automatic translation approach using the Python library *googleTrans*[4]. Meanwhile, Figure 3 shows the most used words in English-language communities.

Table 1 provides detailed frequency occurrences, showing the most used words in English and Brazilian Portuguese communities. It is worth noting that the top four words in both tables are the same, only differing in their positions. Therefore, the words "anxiety, depression, people," and "life" are the most commonly used in both English and Portuguese communities. Some words further down the ranking, even though their positions may vary, are repeated, such as "pain, day," and "time."

Metrics values for "authenticity," "influence," "emotional tone," and "analytical thinking" are also

[4]github.com/ssut/py-googletrans [accessed on 29-05-2023]

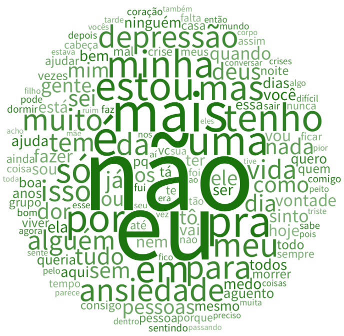

Figure 1: Wordcloud of communities in Portuguese.

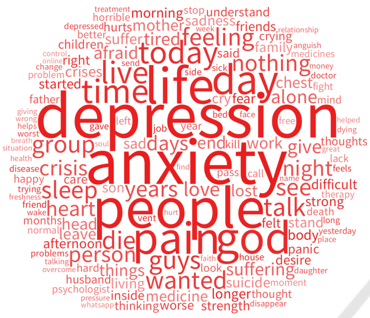

Figure 2: Wordcloud of words translated into English from communities in Portuguese.

obtained for both data sets. Figure 4 shows the distribution of these values for messages in English and Brazilian Portuguese communities. The metrics measured for these attributes range from 0 to 100. Despite differences between the communities, the nuances in data distribution have some similarities. It's important to remember that the number of messages differs for each community, with 835 messages for English and 1945 for Brazilian Portuguese communities.

In the “Analytical Thinking” attribute, the highest concentration of messages is found with low scores (0 to 30), which applies to both communities. There is a more significant distribution of higher scores in the Portuguese community. This dynamic also applies to the ”Influence” attribute, with a higher concentration for low scores (between 0 and 10). However, for both communities, there is a slight increase for a score of

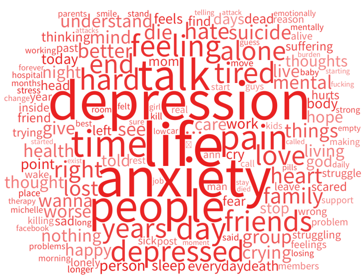

Figure 3: Wordcloud of communities in English.

40 and between 90 and 100. The ”Authenticity” attribute has a higher concentration of messages with scores between 90 and 100, but it also has a concentration of messages with low scores between 0 and 20. The score for the ”Emotional Tone” attribute has three main concentrations. The most significant concentration is for lower scores, between 0 and 10. The second-largest concentration is around 20. The last concentration is for higher values, between 90 and 100.

Table 3 provides values representing statistical measures for reactions, likes, shares, and the number of comments on posts from each community analyzed. In English communities, specific reactions are more commonly used than others. The most frequently used reactions fall into the categories of “Love,” “Sad,” and “Care,” while responses like “Wow,” “Haha,” and “Angry” are less commonly used. The reactions most widely used are those of “Sad,” followed by “Care,” and then “Love.” These reactions also have a higher dispersion in their sum for each post, while reactions with lower averages have less dispersion and, therefore, more uniformity among the posts. The average number of likes and shares corresponds to the amount of “Sad” reactions.

In Brazilian Portuguese communities, the sequence of average reactions is similar to the previous case. There is a more minor difference in the average between “Sad” and “Care” reactions and a more significant difference between “Care” and “Love” reactions. Therefore, the first two reactions are used more frequently in Portuguese communities. The standard deviation for the most commonly used reactions indicates greater dispersion in the most used reactions, similar to the English community, and less dispersion in the less used reactions. However, the standard deviation value for the “Sad” reaction (the most used in both communities) is considerably lower in the Portuguese language groups.

Table 1: Ranking of the most used words in Portuguese-speaking communities.

Word	Language Com.	Freq.	N. Posts	Percent.
anxiety	Pt-Br	373	317	23.105
depression	Pt-Br	371	297	21.6472
people	Pt-Br	328	240	17.4927
life	Pt-Br	292	217	15.8163
pain	Pt-Br	250	180	13.1195
god	Pt-Br	250	167	12.172
day	Pt-Br	228	177	12.9009
time	Pt-Br	177	135	9.8397
today	Pt-Br	164	138	10.0583
live	Pt-Br	162	129	9.4023

Figure 5 shows the heat map of the non-parametric Spearman measure for the English-speaking commu-

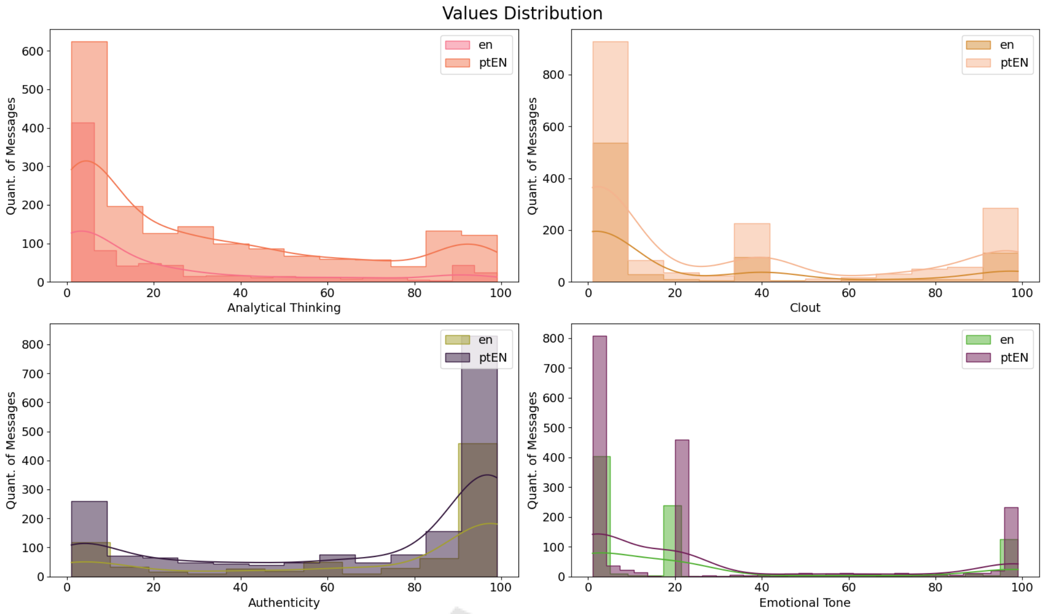

Figure 4: Distribution of values for both sets of communities.

Table 2: Ranking of the most used words in English-speaking communities.

Word	Language Com.	Freq.	N. Posts	Percent.
life	Eng	111	88	15.5203
anxiety	Eng	99	83	14.6384
depression	Eng	97	91	16.0494
people	Eng	77	61	10.7584
talk	Eng	75	64	11.2875
time	Eng	64	55	9.7002
pain	Eng	58	51	8.9947
day	Eng	51	44	7.7601
feeling	Eng	49	39	6.8783
hard	Eng	48	40	7.0547

nity, displaying relevant scores only between the total number of interactions and the number of likes. In this representation, the goal is to discover which attributes are related. With these correlation measures, we aim to find out if there is an association between the emotional measures obtained by LIWC and how the post was received by the community it was posted.

6 DISCUSSION

Not all analyzed messages relate to mental health or directly discuss the disorder. Many texts have a religious theme, contain motivational messages, or share personal experiences. This dynamic occurred for searches with terms in both English and Portuguese.

While applications and translations are available for the LIWC dictionaries in Portuguese, dated 2015, the current version of the dictionary, dated 2022, still needs a Portuguese version. Some recent works, such as (Carvalho et al., 2019) and (Nascimento et al., 2020), use the Brazilian Portuguese dictionary. In the LIWC 22 manual, it is suggested to translate the text using automated methods to English rather than translating the system's base dictionary (Boyd et al., 2022), i.e., translating the text to be analyzed and not the dictionary itself.

Analytical thinking and authenticity metrics reflect how spontaneous and informal a user's discourse is (Boyd et al., 2022). A low value for analytical thinking reflects that the person is not using formal words and is being more intuitive. The highest values for analytical thinking are found in the Brazilian communities, which include therapy discussion groups. As for Authenticity, the higher the metric, the greater the spontaneity. This is reflected in the image. However, messages with low authenticity also have considerable value. The emotional tone metric reflects the type of discussion in the communities. Both for Brazilian and non-Brazilian communities, most messages have a negative tone, with values below fifty.

Using correlation measures, strong correlations between the continuous variables obtained by Crowdtangle and LIWC were expected to be found. However, neither correlation measure obtained relevant correlation values, whether positive or negative. The values for correlations between post-reactions and textual analysis metrics do not show a relevant degree of correlation. The highest correlation values were concentrated in attributes from the same tool. For example, the highest Spearman correlation value was

Table 3: Statistical Description of Community Message Reactions.

	Language Com.	Love	Wow	Haha	Sad	Angry	Care	Likes	Comments	Shares
Avg.	Eng.	2.617	0.041	0.062	8.380	0.013	3.076	8.058683	15.549701	8.427545
Std.Dev.	Eng.	4.957	0.276	0.392	12.288	0.133	4.191	7.362661	16.863349	19.510323
min	Eng.	0	0	0	0	0	0	0	0	0
25%	Eng.	0	0	0	0	0	0.0	4	4	0
50%	Eng.	1	0	0	4	0	1	6	12	3
75%	Eng.	3	0	0	11	0	4	10	22.5	8
max	Eng.	60	6	7	102	2	27	69	194	220
Avg.	Pt-Br	0.676	0.097	0.099	3.945	0.035	2.560	10.513	28.335	0.762
Std.Dev.	Pt-Br	4.488	0.469	0.461	5.466	0.294	3.638	9.409	20.160	2.181
min	Pt-Br	0	0	0	0	0	0	0	0	0
25%	Pt-Br	0	0	0	0	0	0	5	16	0
50%	Pt-Br	0	0	0	3	0	2	8	25	0
75%	Pt-Br	0	0	0	6	0	4	13	36	1
max	Pt-Br	164	13	8	111	6	76	131	213	42

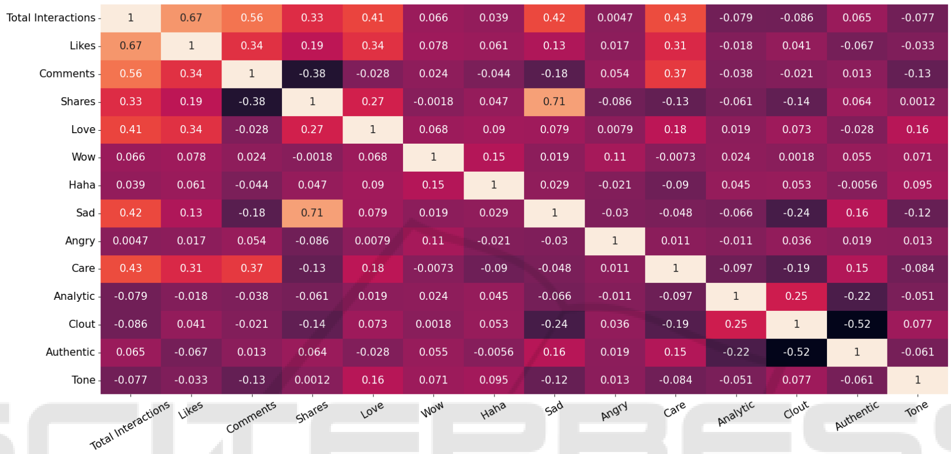

Figure 5: Heatmap for attributes referring to communities in English.

between the number of shares and the number of sad reactions. The second-highest positive value was between Crowdtangle attributes, total interactions, and the number of likes. The most relevant negative correlation was between the attributes Authenticity and Influence, both obtained in the LIWC tool.

7 CONCLUSION

The present work introduced an approach to social media analysis focused on communities discussing the topic of depression. It is worth noting that this study, through a systematic use of data collection and analysis tools, aims to provide explanatory insights into messages discussing depressive disorder. Despite the results presented and discussed here, there is room for further investigation of the correlations between the values obtained from these tools.

Identifying more communities within Facebook for data extraction and obtaining data from other social media platforms are areas to explore in the future. Creating a dataset with greater variety and diversity would be possible with data from various sources. This would allow for the development of more robust and reliable classification models, which healthcare professionals could use in potential analysis tools.

ACKNOWLEDGEMENTS

This work was carried out with the support of CAPES - Notice nº 09/2020 - Prevention and Combating Outbreaks, Endemics, Epidemics, and Pandemics. Proc. #223038.014313/2020-19, and partially supported by the *Oracle for Research* program (award number CPQ-2160239).

REFERENCES

Andalibi, N., Ozturk, P., and Forte, A. (2017). Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, pages 1485–1500, Portland, Oregon, USA. ACM Press.

Association., A. P. and Association., A. P. (2013). *Diagnos-*

tic and statistical manual of mental disorders : DSM-5. American Psychiatric Association Arlington, VA, American Psychological Association. 750 First Street NE, Washington, DC 20002-4242, 5th ed. edition.

Boscarioli, C., Araujo, R., and S. Maciel, R. (2017). *I GranDSI-BR Grand Research Challenges in Information Systems in Brazil 2016-2026*. SBC, Porto Alegre.

Boyd, R. L., Ashokkumar, A., Seraj, S., and Pennebaker, J. W. (2022). The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, pages 1–47.

Brody, D. J., Pratt, L. A., and Hughes, J. P. (2018). *Prevalence of depression among adults aged 20 and over: United States, 2013-2016*. US Department of Health and Human Services, Centers for Disease Control and . . . , 1600 Clifton Rd., Atlanta, GA 30329.

Cafezeiro, I., Viterbo, J., Costa, L., Salgado, L., Rocha, M., and Monteiro, R. (2017). *Strengthening of the Sociotechnical Approach in Information Systems Research*, pages 133–147. SBC-Sociedade Brasileira de Computação, Av. Bento Gonçalves, 9500, Bairro Agronomia, 91501-970.

Carvalho, F., Rodrigues, R., Santos, G., Cruz, P., Ferrari, L., and Guedes, G. (2019). Avaliação da versão em português do liwc lexicon 2015 com análise de sentimentos em redes sociais. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 24–34, Porto Alegre, RS, Brasil. SBC.

Chen, J. and Wang, Y. (2021). Social media use for health purposes: Systematic review. *J Med Internet Res*, 23(5):e17917.

Creswell, J. and Creswell, J. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 1 Oliver's Yard, London, EC1Y 1SP, England.

De Castro, J. C. L. (2009). O inconsciente como linguagem: de Freud a Lacan. *CASA: Cadernos de Semiótica Aplicada*, 7(1):12–42.

De Choudhury, M., Counts, S., and Horvitz, E. (2013). Social Media As a Measurement Tool of Depression in Populations. In *Proceedings of the 5th Annual ACM Web Science Conference*, WebSci '13, pages 47–56, New York, NY, USA. ACM.

Giuntini, F. T., Cazzolato, M. T., de Jesus Dutra dos Reis, M., Campbell, A. T., Traina, A. J. M., and Ueyama, J. (2020). A review on recognizing depression in social networks: challenges and opportunities. *Journal of Ambient Intelligence and Humanized Computing*, 11:1–17.

Horvitz, E. and Mulligan, D. (2015). Data, privacy, and the greater good. *Science*, 349(6245):253–255.

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., and Acharya, D. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):1789–1858.

Li, G., Li, B., Huang, L., Hou, S., et al. (2020). Automatic construction of a depression-domain lexicon based on microblogs: text mining study. *JMIR medical informatics*, 8(6):e17650.

Li, G., Zhou, X., Lu, T., Yang, J., and Gu, N. (2016). SunForum: Understanding Depression in a Chinese Online Community. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW 16*, pages 514–525, New York, New York, USA. ACM Press.

Lima Filho, S. P., Ferreira da Silva, M., Oliveira Sampaio, J., and Couto Ruback Rodrigues, L. (2022). A study about gathering features in depression detection' problem with health professionals community. *iSys - Brazilian Journal of Information Systems*, 15(1):10:1–10:25.

Nascimento, R., Nascimento, G., Carvalho, F., and Guedes, G. (2020). Mineração de opiniões com liwc: abordagem prática sobre sistemas judiciais eletrônicos brasileiros. In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 132–141, Porto Alegre, RS, Brasil. SBC.

Park, S., Kim, I., Lee, S. W., Yoo, J., Jeong, B., and Cha, M. (2015). Manifestation of Depression and Loneliness on Social Networks: A Case Study of Young Adults on Facebook. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 557–570, New York, NY, USA. ACM.

Pimentel, M., Filippo, D., and Santos Marcondes, T. (2020). Design Science Research: pesquisa científica atrelada ao design de artefatos. *RE@D - Revista de Educação a Distância e eLearning*, 3(1):37–61.

Recker, J. (2012). *Scientific Research in Information Systems: A Beginner's Guide*. Springer Publishing Company, Incorporated, Springer Publishing Company, Incorporated.

Skaik, R. and Inkpen, D. (2021). Using Social Media for Mental Health Surveillance. *ACM Computing Surveys*, 53(6):1–31.

Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., and Ohsaki, H. (2015). Recognizing depression from twitter activity. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 3187–3196, New York, NY, USA. Association for Computing Machinery.

Vedula, N. and Parthasarathy, S. (2017). Emotional and linguistic cues of depression from social media. In *Proceedings of the 2017 International Conference on Digital Health*, DH '17, page 127–136, New York, NY, USA. Association for Computing Machinery.

Wieringa, R. J. (2014). *Design Science Methodology for Information Systems and Software Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in software engineering*, volume 9783642290. Springer Berlin Heidelberg, Heidelberger Platz 3, 14197 - Berlin.