

**CARACTERIZAÇÃO E PREVISÃO DO TOM
EMOCIONAL DE USUÁRIOS DAS
COMUNIDADES ONLINE DE TRANSTORNOS
MENTAIS**

BÁRBARA SILVEIRA FRAGA

CARACTERIZAÇÃO E PREVISÃO DO TOM
EMOCIONAL DE USUÁRIOS DAS
COMUNIDADES ONLINE DE TRANSTORNOS
MENTAIS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: ANA PAULA COUTO DA SILVA
COORIENTADOR: FABRICIO MURAI FERREIRA

Belo Horizonte
Novembro de 2019

© 2019, Bárbara Silveira Fraga.
Todos os direitos reservados.

Fraga, Bárbara Silveira

F811c Caracterização e previsão do tom emocional de
usuários das comunidades online de transtornos mentais
/ Bárbara Silveira Fraga. — Belo Horizonte, 2019
xxii, 86 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de
Minas Gerais

Orientador: Ana Paula Couto da Silva

Coorientador: Fabricio Murai Ferreira

1. Computação - Teses. 2. Redes sociais on-line -
Teses. 3. Análise de sentimento - Teses. 4. Saúde
Mental - Teses. 5. Aprendizado do Computador - Teses.
I. Título.

CDU 519.6*04(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO


Caracterização e Previsão do Tom Emocional de Usuários das Comunidades
Online de Transtornos Mentais

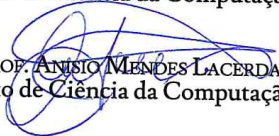
BÁRBARA SILVEIRA FRAGA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:


PROFA. ANA PAULA COUTO DA SILVA - Orientadora
Departamento de Ciência da Computação - UFMG


PROF. FABRÍCIO MURAI FERREIRA - Coorientador
Departamento de Ciência da Computação - UFMG


PROFA. JUSSARA MARQUES DE ALMEIDA GONÇALVES
Departamento de Ciência da Computação - UFMG


PROF. ANÍSIO MENDES LACERDA
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 8 de Novembro de 2019.

Dedico este trabalho a Deus e à minha família.

Agradecimentos

Primeiramente agradeço a Deus por ter me ajudado em cada segundo deste trabalho. Por me dar forças de chegar até o final e não desistir. Também agradeço a minha família, mãe, pai e irmã por entenderem os momentos de ausência e de estresse. Agradeço imensamente ao meu esposo Arthur por ter me apoiado em todas as situações e estar sempre disponível para me escutar, mesmo quando eu não sabia o que queria falar. Agradeço muito por terem acreditado em mim família! Mesmo quando eu estava descrente que chegaria até o final.

Agradeço a todos amigos que torceram para minha chegada até aqui.

Agraço aos professores da PUC por me incentivarem a realizar o mestrado, em especial o Marcelo Werneck e a Maria Augusta (Guta) por terem feito a carta de recomendação. E a Poliana por ter sido a primeira professora a me incentivar a seguir nesta jornada.

Agradeço aos meus orientadores Ana Paula e Fabrício por sempre estarem presentes e terem construído este trabalho comigo.

Agradeço a todos os professores por compartilharem tanto conhecimento.

Agradeço aos amigos que a UFMG me trouxe e que tanto me ajudaram a chegar até aqui. Anderson, muito obrigada pelas ajudas em PAA. Gustavo e Rafael muito obrigada pelas conversas reconfortantes sobre que tudo daria certo.

Agradeço as pessoas com quem trabalhei no projeto do ministério da saúde e contribuíram para meu aprendizado.

Agradeço ao pessoal dos laboratórios LBD e do Locus pela convivência e troca de conhecimento.

Agradeço ao pessoal da A3Data, pessoas com quem trabalho e que me fazem aprender e crescer diariamente.

Agradeço a Fapemig pelas bolsas recebidas.

Por fim, agradeço a todos que cruzaram meu caminho ao longo desta jornada e que me fizeram acreditar que daria certo.

“You can’t connect the dots looking forward; you can only connect them looking backwards. So you have to trust that the dots will somehow connect in your future.”

(Steve Jobs)

Resumo

O aumento alarmante do número de pessoas atingidas por distúrbios de saúde mental tornou-se um dos principais problemas de saúde pública enfrentados pelos governos em todo o mundo. As intervenções clínicas tradicionais são caras, podendo deixar de fora um número considerável de pessoas que estão lutando para melhorar suas condições de saúde mental. Assim, diversos estudos investigam alternativas de intervenção que têm um alcance maior e possibilitam interação contínua com redução de custos. Uma destas alternativas é o uso de redes sociais online (RSOs), que nos últimos anos passaram a conectar pessoas dispostas a trocar experiências relacionadas a problemas de saúde (por exemplo, obesidade e depressão). O objetivo principal desta dissertação é analisar como comunidades com foco na discussão de problemas relacionados à saúde mental auxiliam na melhoria das condições de saúde de seus usuários. Através da caracterização dos usuários de quatro comunidades do Reddit (Depression, SuicideWatch, Anxiety e Bipolar), analisamos como as interações através de posts e comentários influenciam no tom (estado) emocional destes usuários. Em outras palavras, analisamos se a busca por auxílio nestas redes são eficazes e resultam em mudanças de como os usuários exprimem seus sentimentos ao longo do tempo. Nossos resultados mostram que o estado dos usuários muda ao longo do tempo, e que em 68% dos casos a mudança é positiva, indícios que o apoio e incentivo entre usuários destas comunidades são eficazes para melhoria das condições de saúde mental de seus participantes. Ademais, propomos modelos preditivos para capturar a variação do tom emocional dos usuários. Esta variação assume valores em $[-2,2]$. Nossos modelos capturam com boa acurácia a variação do estado emocional dos usuários: no pior caso, o valor do erro médio quadrático foi igual a 1,062 e no melhor, igual a 0,5747. Uma possível aplicação dos modelos é auxiliar intervenções promovidas por profissionais da área de saúde em redes sociais criadas para dar suporte a pessoas com problemas relacionados à saúde mental.

Palavras-chave: Saúde Mental, Redes Sociais Online, Análise de Sentimento, Aprendizado de Máquina.

Abstract

The alarming increase of the number of people affected by mental health disorders has become one of the major public health problems faced by governments around the world. Traditional clinical interventions require high budgets and might not include a considerable number of people struggling to improve their mental health conditions. Thus, several studies have investigated alternative interventions approaches capable of reach further the target population and enable continuous interaction at lower costs. One such alternative is using online social networks (OSN), which in recent years have connected people willing to exchange experiences related to health problems (e.g., obesity and depression). The main objective of this dissertation is to analyze how communities focused on the discussion of mental health disorders are able to help and improve their user's health conditions . Through characterizing users from four Red-dit communities (Depression, SuicideWatch, Anxiety, and Bipolar), we analyzed how interactions through posts and comments might influence the emotional tone (state) of these users. In other words, we analyzed whether seeking help on these networks is effective and if it's followed by behavioral changes in how users express their feelings over time. Our results show that the user's emotional tone changes over time, which 68% of cases are considered a positive change, providing evidence that support and encouragement among users in these communities are effective to improve their patients' mental health. participants. Additionally, we propose predictive models to capture the variation of users' emotional tone. This variation takes on values in $[-2,2]$. Our models accurately capture the variation of users' emotional state: in the worst case, the mean square error value was 1.062 and in the best case 0.5747. The models presented might assist interventions promoted by health professionals in social networks in order to support people with mental health disorders.

Keywords: Mental Health, Online Social Networking, Sentiment Analysis, Machine Learning.

Lista de Figuras

3.1	Resumo da Metodologia Utilizada	15
3.2	Evolução do total de <i>posts</i> e comentários em cada subreddit.	16
3.3	Metodologia da Análise Textual	22
3.4	Comparação do LIWC e Vader.	26
4.1	Evolução do total de <i>posts</i> e comentários em cada subreddit no ano de 2017.	30
4.2	Função de Distribuição Cumulativa Complementar (CCDF) dos graus de entrada e saída do usuário.	30
4.3	Histograma de números de pares de usuários quantizados por volume de interação.	31
4.4	Atividades dos usuários (total de posts e comentários) em cada hora do dia.	33
4.5	Atividades dos usuários em cada dia da semana.	34
4.6	Atividades dos usuários em cada mês do ano.	35
4.7	Tempo, em dias,entre duas atividades.	35
4.8	Número de <i>Posts</i> por usuário.	36
4.9	Número de Comentários por usuário.	36
4.10	Lifespan em dias por usuário.	36
4.11	Grupos em cada Subreddit Por Característica.	39
4.12	Histograma da profundidade das árvores de discussão.	40
4.13	Largura média por nível das árvores de discussão (condicionado a profundidade $>$ nível). Nós de mesma profundidade estão no mesmo nível da árvore.	41
4.14	Probabilidade de cada post ou comentário ser associado a um tópico.	44
4.15	Tamanho dos <i>Posts</i> x Tom Emocional.	46
4.16	Tamanho dos Comentários x Tom Emocional.	46
4.17	Tamanho dos Comentários dos Usuários que realizaram o post x Tom Emocional.	47
4.18	Tom Emocional x Clusters.	48

4.19	Apenas as árvores de discussão circuladas foram consideradas. O usuário A iniciou as <i>threads</i> 1, 2, 4 e 5. Não consideramos a sequência 4, pois o usuário ainda tinha interação na <i>thread</i> 2. A variável p corresponde a um <i>post</i> , As variáveis c são os comentários de quem realizou o <i>post</i> . Os pontilhados são os comentários de outros usuários.	49
4.20	Tom Emocional.	50
4.21	Tom Emocional dos Posts x Tom Emocional Médio dos Comentários (exceto dono do Post).	50
4.22	Tom Emocional dos Posts x Tom Emocional do Último Comentário.	51
4.23	Média do Tom Emocional dos Comentários x Tom Emocional do último comentário	51
4.24	TE último comentário x Média dos comentários nas árvores de discussão	52
4.25	Histograma da Diferença do Tom Emocional entre comentário e post	52
4.26	Média do Tom Emocional dos Comentários x Diferença do tom emocional entre último comentário e post.	53
5.1	Resumo da Tarefa Desenvolvida	55
5.2	Arquitetura RNN	60
5.3	Unidade LSTM	62
6.1	Multilayer Perceptron	68
6.2	Multilayer Perceptron	68

Lista de Tabelas

3.1	Estatísticas básicas de cada comunidade.	17
3.2	Interseção dos Usuários. Cada célula é a razão entre cardinalidade de interseção e a cardinalidade do subreddit indicado na linha.	18
3.3	Análise das Ferramentas de Sentimento.	26
4.1	Score e Tamanho das Palavras nas Comunidades.	28
4.2	Caracterização da Faixa Etária dos Usuários de 2017.	29
4.3	Métricas do grafo de interação do usuário por subreddit.	32
4.4	Percentual de usuários em cada grupo por comunidade.	37
4.5	Dois exemplos de itens na pesquisa sobre a intenção do post e respectivos resultados.	40
4.6	Descritores e seus termos principais.	42
4.7	Contexto de alguns descritores.	42
6.1	Coefficientes da Regressão Linear. Cada ‘*’ equivale a significância estatística. Quanto mais ‘*’, maior é a significância, considerando o p-valor.	64
6.2	MSE’s dos Baselines para cada comunidade.	65
6.3	Resultados (MSE) do modelo RLM com regularização das quatro abordagens para o Bipolar (λ é a constante de regularização).	66
6.4	Resultados (MSE) do modelo RLM com regularização da abordagem 4 para todas comunidades (λ é a constante de regularização).	67
6.5	Resultados MSE das variações dos Modelos Redes Neurais Multilayer’s Perceptrons para a comunidade Bipolar. Baseado em 5 folds em termos de média e desvio padrão.	69
6.6	Resultados do MSE Modelo Rede Neural Multilayer Perceptron. Baseado em 5 folds em termos de média e desvio padrão.	70
6.7	Resultados dos modelos executados para o subreddits Bipolar (Abordagem 1 e 2) com 5 folds.	72

6.8 Resultados dos modelos LSTM padrão de todas comunidades abordagem 2 com 5 folds.	73
--	----

Sumário

Agradecimentos	ix
Resumo	xiii
Abstract	xv
Lista de Figuras	xvii
Lista de Tabelas	xix
1 Introdução	1
1.1 Objetivos	4
1.2 Contribuições	5
1.3 Organização da Dissertação	6
2 Trabalhos Relacionados	7
2.1 Análise de comportamento de usuários em redes sociais online	7
2.2 Estudos sobre transtornos de saúde mental baseados em métodos de computação e estatística	9
2.3 Transtornos de Saúde Mental no Reddit	11
2.4 Diferencial do Trabalho	12
3 Metodologia	15
3.1 Conjunto de Dados	16
3.2 Análise das Interações	19
3.3 Análise das Threads	20
3.4 Análise Textual	21
3.5 Análise de Perfis dos Usuários	23
3.6 Modelagem da Variação do Tom Emocional	24

4	Caracterização das Comunidades	27
4.1	Caracterização Geral dos Usuários	27
4.2	Análise das Interações	29
4.3	Análise das Atividades dos Usuários	32
4.4	Análise dos Perfis dos Usuários	37
4.5	Análise dos Posts e Comentários	38
4.5.1	Análise das Threads	38
4.5.2	Análise Textual: Descritores	41
4.6	Análise do Tom Emocional dos Usuários	45
4.6.1	Tom Emocional e Clusters	47
4.6.2	Variação do Tom Emocional nas Árvores de Discussão	47
5	Modelos para Previsão do Tom Emocional	55
5.1	Tarefa	55
5.2	Atributos	56
5.2.1	Análise Textual com Atributos Psicolinguísticas	56
5.2.2	Análise baseada em Word Embeddings	57
5.3	Modelos de Previsão	58
5.3.1	Regressão com Atributos do LIWC	58
5.3.2	Redes Neurais	59
6	Análise Experimental	63
6.1	Baselines	63
6.2	RLM com regularização	65
6.3	Modelo Rede Neural - Multilayer Perceptron	67
6.4	Modelo Long Short Term Memory (LSTM)	70
6.5	Discussão dos Resultados	73
7	Conclusão e Trabalhos Futuros	75
	Referências Bibliográficas	79

Capítulo 1

Introdução

Nos últimos anos, dados da Organização Mundial de Saúde (OMS) alertam para o aumento do total de pessoas no mundo que sofrem de algum tipo de transtorno de saúde mental; 1 em cada 4 pessoas são afetadas por esses problemas em algum ponto de sua vida [Organization et al., 2013]. No momento da elaboração desta dissertação, o último relatório da OMS [Organization et al., 2017] aponta que o número de casos de depressão aumentou 18% entre 2005 e 2015: são 322 milhões de pessoas em todo o mundo, a maioria mulheres. Até 2030, a depressão será a principal causa mundial das doenças. Outro transtorno que atinge mundialmente 264 milhões indivíduos é a de ansiedade, caracterizada por sentimentos de medo, de pânico, de ansiedade social, etc. O transtorno bipolar¹ afeta cerca de 60 milhões de pessoas e, muitas vezes, este problema pode ser confundido com um caso de depressão ou de ansiedade. No seu estado mais grave, esses transtornos podem levar ao suicídio. Aproximadamente 800 mil pessoas morrem devido ao suicídio todo ano, o que corresponde a 1(uma) morte a cada 40 segundos no mundo. Este problema é a segunda principal causa de morte de jovens entre 15 a 29 anos².

Apesar desses números preocupantes, muitas das pessoas que sofrem de distúrbios mentais não recebem tratamento. A OMS informa que a cada 4 pessoas, 3 não recebem qualquer tipo de tratamento e que 45% da população mundial vive em um país com menos de um psiquiatra para 100 mil pessoas³. Outro fator agravante é que os sistemas de saúde não atuam adequadamente no tratamento dos transtornos mentais. A OMS⁴

¹<https://www.who.int/en/news-room/fact-sheets/detail/mental-disorders>

²<https://www.who.int/docs/default-source/mental-health/suicide/infographics-suicide.pdf>

³https://www.who.int/mental_health/evidence/atlas/interactive_infographic_2015.pdf

⁴<https://www.who.int/news-room/fact-sheets/detail/mental-disorders>

mostra que nos países de baixa e média renda, entre 76% e 85% das pessoas com transtornos mentais não recebem tratamento para seu transtorno. Nos países de alta renda, entre 35% e 50% das pessoas com transtornos mentais estão na mesma situação.

A combinação entre recursos escassos (principalmente em países de economia mais instável), o estigma social associado aos transtornos mentais [Barney et al., 2006] e a resistência em pedir ajuda, faz com que muitas pessoas que sofrem destes transtornos não sejam ajudadas da melhor maneira possível, levando a situações drásticas e irremediáveis, como o suicídio.

Devido ao aumento de casos destes tipos de transtornos, políticas de saúde pública eficientes precisam ser implementadas. No caso específico da depressão, por exemplo, a OMS é responsável pelo programa *Mental Health Gap Action Programme*⁵, que visa ajudar os países a aumentar o volume dos serviços prestados às pessoas com transtornos mentais, neurológicos e de uso de substâncias, por meio de cuidados providos por profissionais de saúde que não são especialistas em saúde mental. A iniciativa defende que, com cuidados adequados, assistência psicossocial e medicação, dezenas de milhões de pessoas com transtornos mentais, incluindo depressão, poderiam começar a levar uma vida normal – mesmo quando os recursos são escassos.

Este quadro geral faz com que novos recursos em busca de compreender e auxiliar os indivíduos afetados por estes transtornos sejam explorados. Em particular, podemos ressaltar o papel das redes sociais online. Inicialmente focadas em fomentar amizades, trocar imagens ou vídeos, elas passaram a conectar pessoas dispostas a compartilhar experiências relacionadas a problemas de saúde, como por exemplo, obesidade [Pappa et al., 2017] e depressão [De Choudhury & De, 2014; Kavuluru et al., 2016]. O trabalho realizado por De Choudhury [2013] mostra que o compartilhamento online de pensamentos, sentimentos e experiências com outras pessoas contribui com a melhoria do bem-estar dos indivíduos. Os autores em [Lopes et al., 2014] utilizaram uma rede social para identificar a percepção da promoção da saúde por grupo de profissionais da saúde. Neste contexto, a presente dissertação pretende analisar comunidades online com foco na discussão de transtornos de saúde mental. Para investigar como os usuários destas comunidades trocam experiências e exprimem seus sentimentos, utilizamos a rede social Reddit⁶.

O Reddit é um site de fóruns com características de redes sociais: é composta por comunidades (subreddits), onde os usuários compartilham suas experiências e dúvidas sobre os mais diversos assuntos. Esta rede permite aos usuários criarem e compartilharem conteúdos em torno de um assunto. Em uma espécie de fórum, um usuário

⁵https://www.who.int/mental_health/mhgap/en/

⁶<https://www.reddit.com/>

inicia uma *thread* ao publicar um *post*. Este *post* pode ser respondido por comentários de outros usuários ou mesmo daquele que escreveu o *post*. Comentários, por sua vez, podem ser respondidos por outros comentários. Outra característica importante é que os usuários podem permanecer anônimos, definindo identidades temporárias, encorajando os mesmos a discutir sobre assuntos mais delicados e compartilhar pensamentos e sentimentos que muitas vezes não são aceitos facilmente pela sociedade [Caplan & Turner, 2007].

Dados de 2018 mostram que o Reddit é constituído por 330 milhões de usuários ativos por mês. Esta métrica é contabilizada através dos usuários que fazem *login*, ao menos uma vez por mês. Pessoas que acessam a rede sem efetuar login somente para visualizar alguma mensagem ou assistir um vídeo não são consideradas como usuários ativos. Ao todo são 153 milhões de *posts*, 1,2 bilhões de comentários. O Reddit é composto por, aproximadamente, 150 mil comunidades ativas e 14 bilhões de visualizações por mês⁷. É indiscutível o uso massivo desta rede social.

No Reddit existem diversos subreddits relacionados a transtornos da saúde mental [Gkotsis et al., 2016, 2017], como por exemplo, Anxiety (/r/anxiety), Bipolar (/r/bipolar), Depression (/r/depression), SuicideWatch (/r/SuicideWatch), MentalHealth (/r/mentalhealth), Selfharm (/r/selfharm), entre outros. Podemos citar 4 comunidades que se destacam pelo número de usuários e alto nível de troca de mensagens. O subreddit Depression se define como um espaço de apoio para quem luta contra a depressão. O SuicideWatch é utilizado por usuários que estão com pensamentos suicidas ou que tentaram cometer efetivamente suicídio. O subreddit Anxiety é direcionado às discussões e apoio a pacientes que sofrem de transtorno de ansiedade e seus entes queridos. A definição apresentada na página da comunidade Bipolar é a seguinte: “um porto seguro para questões bipolares. Aqui somos uma comunidade, não apenas uma página de ajuda. Faça parte de algo que se importa com quem você é.”

Possíveis características que incentivam diferentes pessoas a buscarem apoio nestas comunidades são listadas a seguir. Primeiro, os usuários dos fóruns podem permanecer anônimos. Segundo, além da garantia do anonimato, existem diretrizes aplicadas para cada comunidade. Na comunidade Depression, por exemplo, existem regras como: “Não poste ou responda com: Amor difícil de qualquer descrição; Mensagens de melhoria geral ou “melhora”. O incentivo não ajuda, a menos que integre uma compreensão real e pessoal dos sentimentos e da situação; qualquer coisa explícita ou incitante relacionada a suicídio ou dano próprio.”. Terceiro, essas comunidades disponibilizam canais diretos para que pessoas em crise possam entrar em contato mais rapidamente⁸

⁷<https://redditblog.com/2018/12/04/reddit-year-in-review-2018/>

⁸<https://www.reddit.com/r/SuicideWatch/wiki/hotlines>

com algum mecanismo de ajuda. Por fim, usuários com comportamento inadequado são removidos, garantindo que as discussões nessas comunidades tenham sempre como objetivo a melhoria da saúde mental dos seus participantes.

O objetivo principal desta dissertação é analisar como comunidades com foco na discussão de problemas relacionados à saúde mental auxiliam na melhoria das condições de saúde de seus usuários. Através da caracterização dos usuários de quatro comunidades do Reddit (Depression, SuicideWatch, Anxiety e Bipolar), analisamos como as interações através de *posts* e comentários influenciam no tom emocional (estado) destes usuários. Em outras palavras, analisamos se a busca por auxílio nestas redes é eficaz e resulta em mudanças de como os usuários exprimem seus sentimentos ao longo do tempo. Além disso, propomos modelos preditivos para capturar a variação do tom emocional dos usuários. Uma possível aplicação dos modelos é auxiliar intervenções promovidas por profissionais da área de saúde em redes sociais criadas para dar suporte a pessoas com problemas relacionados à saúde mental.

1.1 Objetivos

Os objetivos principais desta dissertação são os seguintes: (1) compreender e medir como redes sociais online são utilizadas para apoiar na melhoria do tom emocional de usuários com problemas de saúde mental e (2) definir modelos que permitam acompanhar a evolução do tom emocional do usuário. Com base nesses objetivos, organizamos esta dissertação em torno de três questões de pesquisa (QP) que serão apresentadas a seguir.

- **QP1 - Como são caracterizadas as interações dos usuários com outros usuários e com o conteúdo dentro do Reddit?**

Caracterizamos a maneira com que os usuários interagem em um subreddit, a fim de identificar o objetivo de tal interação. Além disso, analisamos interações entre os usuários e o conteúdo para entender sobre o que é essa comunicação e como a mesma varia ao longo de uma *thread*.

- **QP2 - Existem grupos de usuários que desempenham papéis semelhantes dentro do Reddit?**

Identificamos os grupos de usuários com base em suas atividades, ou seja, total de *posts* e comentários realizados dentro das *threads* e a partir da análise textual de *posts* e comentários, identificamos o tipo de conteúdo que os usuários compartilham nestas comunidades.

- **QP3 - Como mensurar o estado emocional de um usuário e como modelar as suas mudanças a partir das interações com os demais usuários de uma comunidade?**

Sabendo que existem usuários que procuram estas comunidades em busca de ajuda e outros que estão dispostos a ajudar, analisamos a variação do estado emocional de um usuário dentro de uma *thread*, verificando se os outros usuários influenciam no tom emocional do usuário que iniciou a discussão. A partir destas análises, propomos modelos para prever a evolução do tom emocional dos participantes destas comunidades.

1.2 Contribuições

As contribuições deste trabalho perpassam por diversas análises com dados de comunidades voltadas a transtornos mentais no Reddit. A seguir descrevemos as principais contribuições desta dissertação.

Modelagem e caracterização das interações. A partir de grafos, modelamos as interações dos usuários em cada comunidade. Uma das conclusões desta análise é que os usuários nessas comunidades são centrados em torno de tópicos de discussão e não em formar amizades, como acontece em outras redes sociais.

Modelagem e caracterização das discussões em threads. Modelamos as discussões entre usuários dentro de uma mesma *thread* através de uma árvore, onde o *post* é a raiz da árvore. Assim, verificamos que as árvores mais profundas retratam conversas entre somente dois usuários: um deles buscando ajuda e o outro oferecendo apoio e atenção.

Extração de principais tópicos abordados nas quatro comunidades. Analisamos os principais tópicos discutidos nas quatro comunidades. Para isto aplicamos o RMN (*Relationship Modeling Network*), proposto por Iyyer et al. [2016], buscando encontrar semelhanças na forma com que as pessoas se expressam quando discutem sobre transtornos de saúde mental. Assim identificamos alguns papéis desempenhados nas comunidades, entre eles, o papel de conselheiro.

Análise de perfis dos usuários. Identificamos perfis de usuários de acordo com o nível de atividades dentro da comunidade. Diante dos grupos com perfis semelhantes, analisamos as características de cada grupo e, conseqüentemente, como eles respondem os estímulos da comunidade, assim como quais estímulos eles proveêm. Com isso, pretendemos identificar qual é a melhor maneira de ajudar um usuário, ou seja, verificar

qual abordagem é mais efetiva e quais são os indícios de apoio social oferecidos por essas comunidades.

Modelagem do tom emocional dos usuários. Para mapear o estado emocional do usuário em uma publicação, definimos uma forma de mensurar o tom emocional (TE) de um usuário. Adicionalmente, propomos modelos para prever a variação do TE do usuário que iniciou a *thread*. Este modelo pode auxiliar intervenções guiadas por profissionais de saúde em comunidades online focadas na discussão e apoio de pessoas com transtornos mentais.

1.3 Organização da Dissertação

O restante deste trabalho está organizado da seguinte maneira. O Capítulo 2 apresenta os trabalhos relacionados a esta dissertação, envolvendo estudos sobre: Análise de Comportamento de Usuários em Redes Sociais Online; Estudos sobre Transtornos de Saúde Mental Baseados em Métodos de Computação e Estatística e Transtornos de Saúde Mental no Reddit. O Capítulo 3 descreve detalhadamente a metodologia aplicada em cada etapa. O Capítulo 4 apresenta a caracterização das comunidades. O Capítulo 5 apresenta modelos para a previsão do tom emocional do autor da *thread*. O Capítulo 6 exhibe as análises experimentais dos modelos propostos. Por fim, o Capítulo 7 apresenta as conclusões e contribuições desta dissertação, além de trabalhos futuros.

Capítulo 2

Trabalhos Relacionados

Esta dissertação trata da análise do comportamento de usuários em comunidades do Reddit voltadas para a discussão de transtornos de saúde mental. Portanto, é englobada dentro de estudos que analisam o comportamento de usuários em redes sociais online, e dentro de estudos sobre transtornos de saúde mental baseados em métodos de computação e estatística. O capítulo é dividido em quatro seções: a Seção 2.1 apresenta trabalhos relacionados à análise de comportamento de usuários em redes sociais online; a Seção 2.2 discorre sobre estudos com métodos computacionais e estatísticos sobre transtornos de saúde mental; a Seção 2.3 apresenta trabalhos sobre transtornos mentais no Reddit. Finalmente, a Seção 2.4 expõe o diferencial deste trabalho com relação aos apresentados.

2.1 Análise de comportamento de usuários em redes sociais online

A análise de redes sociais online é uma ferramenta importantíssima para diversas áreas do conhecimento. Isso ocorre porque, essas mídias podem influenciar o comportamento de seus usuários a partir das interações ou de trocas de informações com seus relacionamentos.

O estudo feito por Althoff et al. [2017] mostra como as redes sociais influenciam o comportamento do usuário em um aplicativo de rastreamento de atividades físicas. Outro trabalho que abordou a influência dos usuários nas redes sociais foi feito por Bakshy et al. [2011], onde os autores investigaram a relação entre as características e a influência relativa dos usuários do Twitter. O trabalho desenvolvido por Sigurdsson et al. [2019] estuda a relação de redes sociais de empresas de aviação com o envolvimento

dos usuários nas interações nas mídias sociais.

Alguns trabalhos têm como foco troca de informações entre os usuários: Souza et al. [2017]; Wang et al. [2016] abordam a caracterização de informações, através de extração de tópicos a fim de descrever o contexto analisado. Outra técnica que pode ser utilizada para esta análise é o modelo RMN (*Relationship Modeling Network*), proposto por Iyyer et al. [2016]. O RMN tem como objetivo modelar a relação entre um par de entidades que aparece múltiplas vezes em diversos documentos ou contextos. A vantagem do RMN em relação ao LDA (*Latent Dirichlet Allocation*), método clássico introduzido por Blei et al. [2003], para extração de tópicos latentes a partir de um corpus, é que o RMN é capaz de modelar relacionamentos entre entidades associadas aos textos.

As redes sociais tiveram seu papel diversificado ao longo dos últimos anos. Existem atualmente, por exemplo, comunidades para auxiliar no controle de peso. O estudo realizado por Cunha et al. [2017] propõe um *framework* para estudar a causalidade da perda de peso no subreddit loseit. Os resultados mostram que os usuários que receberam um determinado número de comentários no primeiro *post* são mais prováveis de retornar para outras atividades na comunidade e relatar posteriormente uma alta perda de peso em comparação com aqueles que não tiveram. Também existem trabalhos que abordam transtornos alimentares nas redes sociais, como o estudo realizado por Ramírez-Cifuentes et al. [2018] que propõe um modelo para identificar precocemente o usuário que sofre com problemas de anorexia. Desta maneira usuários identificados previamente poderão ser tratados mais rapidamente e terão uma maior probabilidade de melhora.

Existem ainda trabalhos que estudam relações do aspecto offline (vida real) e online (redes sociais) dos usuários. O estudo de Falavarjani et al. [2019] é motivado pela teoria da influência social normativa, a qual argumenta que os usuários podem adaptar seu comportamento ou suas expressões para serem aceitos na comunidade. Assim, o objetivo do trabalho é entender se as atividades offlines diárias do usuário têm algum impacto no comportamento do mesmo nas redes sociais online. Um dos resultados do estudo mostra que existem alterações comportamentais dos usuários em situações sociais.

O estudo realizado por Stawarz et al. [2019] tem como objetivo investigar as tecnologias que apoiam os usuários em sua saúde mental. Entre as tecnologias utilizadas estão aplicativos para *smartphones*, fóruns de discussão, mídias sociais, sites e programas baseados na Web. Para atingir o objetivo, os autores realizaram uma pesquisa online para coletar respostas de indivíduos que utilizam tecnologia para apoiar seu bem-estar/saúde mental. Um dos resultados encontrados é que os usuários da pesquisa

utilizam as mídias sociais e os fóruns de discussão para aprender com as experiências de outras pessoas e usar esse conhecimento para entender sua própria situação.

2.2 Estudos sobre transtornos de saúde mental baseados em métodos de computação e estatística

Os problemas relacionados a saúde mental são diversos. O uso de ferramentas para auxiliar pessoas que sofrem desses transtornos, além de consultas presenciais, não é um paradigma novo. Por exemplo, a utilização de atendimento telefônico, onde voluntários prestam suporte é uma maneira de auxílio muito difundida em vários países, como Argentina¹, Brasil², Estados Unidos³, França⁴, entre outros.

No entanto, novas formas de uso dos meios de comunicação, sejam telefones celulares ou a Internet, vem revolucionando o suporte oferecido às pessoas que sofrem de transtornos como depressão e ansiedade. Assim, alguns trabalhos na literatura buscam entender como novas tecnologias podem ser eficazes em minimizar os sintomas destas doenças.

Neste cenário, os autores em Althoff et al. [2016] analisam conversas, via SMS, de indivíduos auxiliados por conselheiros ligados à organização *Crisis Trends*⁵. O estudo busca avaliar o comportamento dos conselheiros, uma vez que estes desempenham um papel crucial no apoio dos indivíduos. Uma das principais conclusões é que os conselheiros bem sucedidos em ajudar os indivíduos são mais sensíveis à trajetória da conversa, respondem às mensagens de forma mais criativa, sem usar frases genéricas e rapidamente identificam o foco do problema no qual o indivíduo se encontra, colaborando para a solução do mesmo.

O estudo de problemas de saúde mental é investigado com base nos dados de várias redes sociais. No Instagram, uma rede online de compartilhamento de fotos e vídeos entre os usuários, o estudo feito por Blair & Abdullah [2018] analisa os desafios enfrentados pelos usuários que divulgam seus problemas de saúde mental nesta rede. O trabalho mostra que a mídia social pode ter um impacto positivo para esses usuários, uma vez que permite que eles construam suas próprias comunidades. O trabalho feito por Islam et al. [2018] utiliza o Facebook como objeto de estudo para detecção de

¹<https://www.casbuenosaires.com.ar/ayuda>

²<https://www.cvv.org.br/>

³<https://www.crisistextline.org/texting-in>

⁴<http://www.suicide-ecoute.fr/>

⁵<https://crisistrends.org/>

depressão através de técnicas de aprendizado de máquina. Os autores mostram que o método proposto, que utiliza características psicolinguísticas, é eficiente para realizar a classificação dos usuários com depressão. Em outro trabalho, os autores Weerasinghe et al. [2019] analisaram tópicos no Twitter para descobrir padrões de linguagem que diferenciam indivíduos com doenças mentais de um grupo de controle. Como resultado, os pesquisadores confirmaram certos padrões e descobriram outros novos.

Os padrões estéticos estabelecidos pela sociedade também contribuem para os transtornos mentais. Isso pode ser visto no trabalho realizado por Ramseyer Winter et al. [2017], que explora correlações entre apreciação corporal e saúde mental em uma amostra diversificada racialmente e etnicamente de mulheres adultas. Os resultados da regressão linear indicam que a valorização do corpo foi significativamente e inversamente associada à depressão.

Outro trabalho interessante, conduzido por Seabrook et al. [2018] se baseia na atualização do status do Facebook e Twitter para identificar relações entre o nível da depressão, a variabilidade e a instabilidade na expressão de palavras de emoção. As interações dos usuários em suas redes sociais podem refletir um transtorno mental. Os autores Dutta et al. [2018] observam que o aumento dos níveis de ansiedade de um indivíduo resulta em maior interação futura com laços fracos, o que indica uma tendência para buscar apoio em comunidades online.

O trabalho feito por Coppersmith et al. [2018] utilizou técnicas de processamento de linguagem natural para detectar sinais quantificáveis em torno de tentativas de suicídio. Além disso, os autores descreveram como seria um sistema utilizado por especialistas, por exemplo: médicos e psicólogos, que poderiam intervir em uma situação de risco. É interessante que neste trabalho, os autores também realizaram uma discussão ética sobre o uso de tal tecnologia na privacidade dos indivíduos.

Trabalhos mais recentes abordam a detecção da ansiedade, bipolaridade de depressão e suicídio em redes sociais [Shen & Rudzicz, 2017; Yusof et al., 2018; Wongkoblap et al., 2018; Murrieta et al., 2018; Lee et al., 2018; Chen et al., 2018; Leis et al., 2019; Wolohan et al., 2018; Wongkoblap et al., 2019; Gruda & Hasan, 2019; Sahota & Sankar, 2019; Baba et al., 2019]. Infelizmente, o número de casos de transtornos mentais aumenta a cada ano e atingem as pessoas das mais diversas idades. Esses trabalhos procuram entender o comportamento dos usuários presentes nessas redes sociais como forma de propor políticas que contribuam para a diminuição do volume de pessoas afetadas por estes problemas. O diferencial do nosso trabalho é que pretendemos modelar o usuário durante o período em que ele permanece em uma comunidade, a fim de monitorar sua melhoria ao longo da conversa, viabilizando métodos de intervenção mais efetivos para auxiliar esses usuários rapidamente.

O uso das redes sociais online pode ser efetivo para auxiliar pessoas que estão passando por algum problema. Neste contexto, os autores Saha et al. [2018] estudam o efeito de recomendações de conselhos após as mortes de estudantes em um campus universitário, seja por suicídio ou por um incidente incontrolável. Os resultados mostram que os indivíduos expostos ao luto apresentam maior tristeza e que as recomendações surtem efeitos psicológicos positivos nesses indivíduos.

Os autores Costa et al. [2018] apresentam um estudo da rede de apoio Mood-Panda, para mostrar como o humor do usuário evolui à medida que a rede de contatos do usuário aumenta. Os autores mostram que a evolução do humor do usuário depende dos tipos de conexões que são criadas. Adicionalmente, o humor dos usuários melhora quando o apoio emocional é fornecido pelo sexo oposto, e pessoas com instabilidade contínua de humor não devem receber apoio emocional de outras pessoas com o mesmo transtorno, pois isso não é efetivo.

2.3 Transtornos de Saúde Mental no Reddit

O Reddit é uma mídia que permite anonimato, fazendo com que as pessoas se sintam mais livres para falar de assuntos por que seriam criticadas em seu meio social físico. Assim sendo, existem diversos trabalhos que estudam problemas mentais nas comunidades do Reddit.

O estudo de Gkotsis et al. [2016] investiga a linguagem dos *posts* do Reddit relacionados à saúde mental e identifica várias das características linguísticas marcantes. Em um outro trabalho os mesmos autores analisam as postagens do Reddit para desenvolver classificadores que reconheçam e classifiquem postagens relacionadas à doença mental, através de técnicas de *deep learning* [Gkotsis et al., 2017].

Considerando o uso do Reddit e comunidades direcionadas à saúde mental, o trabalho de De Choudhury & De [2014] apresenta uma análise do discurso feito por usuários de comunidades ligadas à saúde mental. Por exemplo, os autores investigam como o grau de desinibição nos comentários e *posts* feitos por usuários anônimos se difere daqueles feitos pelos usuários que se identificam. O anonimato, por meio de contas descartáveis, permite um maior auto discurso em torno do tópico estigmático da saúde mental, uma vez que não existe a preocupação em ser identificado. Além disso, postagens com menos inibição normalmente discutem problemas de relacionamentos e de saúde e são as que atraem mais atenção. Essas postagens reúnem maior apoio comunitário, através de votos e comentários.

Os autores Kavuluru et al. [2016] analisam o subreddit SuicideWatch, com o

objetivo de classificar, de forma automática, comentários que possam impactar positivamente o comportamento de indivíduos com pensamentos suicidas. Para isso, selecionaram 3000 comentários aleatórios do subreddit escolhido, para que três avaliadores classificassem cada um dos comentários como: útil genérico, útil específico e não útil. Em seguida, usaram dados rotulados para treinar modelos de aprendizado de máquina. O estudo de Alambo et al. [2019] também analisou este subreddit, com o objetivo de encontrar uma estrutura automática para identificar riscos de suicídio através de respostas às perguntas. Para isto, aplicaram agrupamento semântico e modelos de sequência. Com o auxílio de psiquiatras clínicos, os autores identificaram quatro grupos: indicador de suicídio, ideação suicida, comportamento suicida e tentativa de suicídio.

Os autores De Choudhury et al. [2016] caracterizam os participantes das comunidades de saúde mental do Reddit que postam no subreddit de suporte a pessoas com ideações suicidas usando uma série de métricas baseadas na interação linguística e social as quais caracterizam o estado psicológico e comportamental do indivíduo. Adicionalmente, os autores apresentam uma nova aplicação do *propensity score matching* para explorar como os usuários podem compartilhar o conteúdo de ideação suicida no futuro. Por fim, examinam se é possível prever automaticamente a tendência de indivíduos que discutem questões de saúde mental a se engajarem na comunidade de apoio ao suicídio.

Há ainda os autores Park & Conway [2018], que estudam a estrutura da escrita dos indivíduos dos subreddits que tratam de depressão, transtorno bipolar e esquizofrenia. As análises realizadas mostram que os usuários que sofrem dessas doenças têm dificuldade em expressar suas ideias. O trabalho realizado pelos autores Yates et al. [2017] apresenta um *framework* neural para apoiar e estudar usuários nas comunidades do Twitter e do Reddit. Eles propuseram métodos para identificar o risco de um usuário cometer uma autoagressão, o que está relacionado com usuários que sofrem de depressão.

2.4 Diferencial do Trabalho

Conforme apresentado neste capítulo, estudos sobre transtornos mentais em redes sociais online têm sido abordados em diversos trabalhos. A maior parte deles tem um objetivo em comum: oferecer informações sobre o comportamento das pessoas que sofrem destas doenças para assistir a elaboração de políticas de saúde pública, desenvolver aplicações que auxiliem estas pessoas, entre outras medidas.

A maioria dos estudos abordam apenas um distúrbio mental. Nosso trabalho utiliza quatro comunidades relacionadas a esses problemas. Com isso, identificamos semelhanças e diferenças entre os tópicos discutidos nas comunidades.

A principal contribuição deste trabalho, com relação aos demais estudos que utilizaram o Reddit, é que iremos trabalhar com maior granularidade na análise do texto. Nosso estudo analisa o usuário desde o momento em que publica um *post* até o momento em que realiza seu último comentário dentro de cada *thread*. Com isso, é possível analisar a evolução de seu humor e suas interações dentro da comunidade.

Com base na análise desta granularidade mais fina das interações dos usuários nos subreddits analisados, propomos modelos que representam variações do humor dos usuários. Estes modelos podem ser utilizados para identificar as intervenções que são efetivas na melhora do humor do usuário. Além disso, podem auxiliar na identificação das pessoas que possuem o perfil de conselheiros nestas comunidades.

Em muitos trabalhos, verificamos que os autores necessitam de especialistas para rotular os dados. Após a rotulação, realizam um modelo de classificação para prever se um *post* ou um comentário tem, por exemplo, conteúdo depressivo. Assim sendo, outro diferencial deste trabalho é que estudamos um valor quantitativo para o humor do usuário, denominado tom emocional. Nós construímos modelos baseados na variação desta medida dentro de uma *thread*. Desta forma, não precisamos de rotulações e, através de modelos, verificamos se o usuário com um humor mais negativo apresentou alguma melhora após as interações realizadas dentro da comunidade.

Capítulo 3

Metodologia

Neste capítulo, apresentamos a metodologia proposta para a análise e modelagem de interações entre usuários em comunidades relacionadas a problemas de saúde mental. Esta análise pode contribuir para identificar os usuários que mais precisam de suporte. Assim, medidas de intervenção, como conselhos, trocas de experiências, conversas, inclusive com especialistas, podem ser utilizadas para contribuir com a melhora destes usuários.

A Figura 3.1 apresenta uma visão geral dos métodos utilizados para atingir o objetivo descrito anteriormente. Nas análises das interações, das *threads* e dos textos, o objetivo é caracterizar as interações entre usuários e destes com o conteúdo do Reddit (QP1). Na análise dos perfis, identificamos grupos de usuários similares no que se refere aos padrões de postar e comentar no Reddit (QP2). Na abordagem dos modelos apresentamos como mensurar o estado de um usuário e a modelagem da relação entre mudanças neste estado e suas interações dentro de um subreddit (QP3).

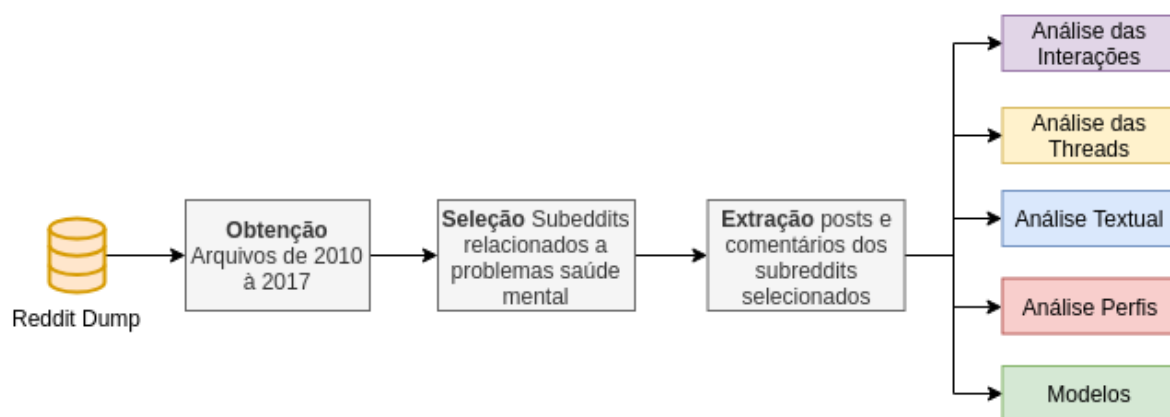


Figura 3.1: Resumo da Metodologia Utilizada

As principais etapas da metodologia descritas neste capítulo são: a coleta de

dados do Reddit e a extração do conjunto de dados (Seção 3.1); a modelagem dos dados referentes às interações dos usuários em um grafo e as métricas calculadas (Seção 3.2); a modelagem das *threads* para investigar os *posts* e comentários feitos pelos usuários (Seção 3.3); uso do RMN para a extração dos assuntos nas comunidades (Seção 3.4); o método para encontrar grupos de usuários similares em cada comunidades (Seção 3.5); e a modelagem do tom emocional dos usuários ao longo das *threads* (Seção 3.6).

3.1 Conjunto de Dados

Para as análises apresentadas nesta dissertação, recuperamos *online*¹ todos os dados de *posts* e comentários de todas as comunidades (também chamadas de subreddits) do Reddit entre o período de Janeiro de 2011 à Dezembro de 2017. Como nossas análises foram realizadas a partir de comunidades relacionadas a problemas de saúde mental, limitamos a extração aos dados de *posts* e comentários a essas comunidades.

Existem mais de 25 subreddits que focam na discussão de transtornos mentais. Os quatro com o maior número de *posts* e comentários Gkotsis et al. [2016, 2017] são²: Depression (/r/depression), SuicideWatch (/r/suicide), Anxiety (/r/Anxiety) e Bipolar (/r/bipolar). O subreddit opiates não foi considerado na análise, por não ser um tipo de problema de saúde mental³.

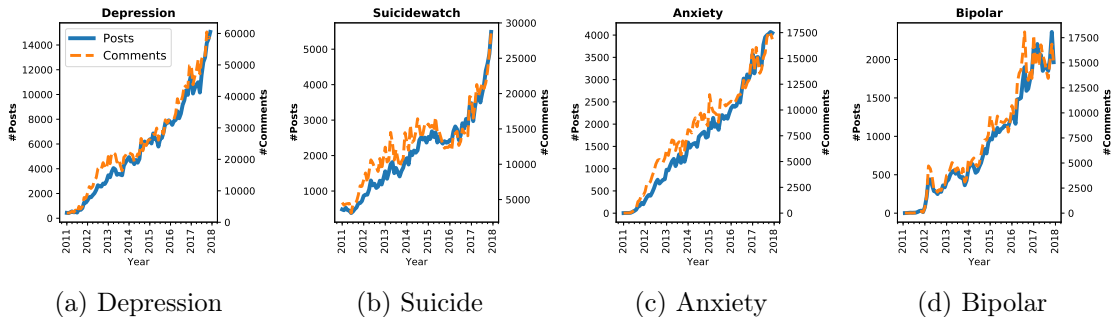


Figura 3.2: Evolução do total de *posts* e comentários em cada subreddit.

A Figura 3.2 apresenta o volume mensal de *posts* e comentários em cada um dos subreddits, no período de 2011 e 2017. Ao analisar a quantidade de *posts* e comentários em cada ano, observa-se que em 2011 foram realizadas menos de 1000 *posts* nas comunidades Anxiety e Bipolar. No entanto, em 2012, este número cresceu 7 e 30 vezes nas

¹<http://files.pushshift.io/reddit/>

²Não consideramos o subreddit Opiates (/r/opiates) por estar mais ligado à dependência de re-médios.

³<https://www.mind.org.uk/information-support/types-of-mental-health-problems/>

respectivas comunidades. Diante disso, iremos considerar apenas o período entre 2012 e 2017 nas nossas análises, uma vez que 2011 as comunidades ainda estavam surgindo.

Ao analisar detalhadamente a quantidade de *posts* e comentários, entre 2012 e 2017, esses números cresceram substancialmente em todos os subreddits: o subreddit Bipolar teve o maior aumento no número de *posts* (6,7 vezes), seguido por Anxiety (6,5 vezes), Depression (5,3 vezes) e SuicideWatch (3,6 vezes). Em termos de número de comentários, o subreddit com maior aumento é Bipolar (5,7 vezes), seguido por Anxiety (3,8 vezes), Depression (3,6 vezes) e SuicideWatch (2,3 vezes).

	Depression	SuicideWatch	Anxiety	Bipolar	Total
Usuários Únicos	333.624	162.363	121.945	35.081	569.122
Posts	468.507	169.541	142.168	73.254	853.470
Comentários	2.128.991	1.024.171	685.867	607.262	4.446.291
Posts/usuário	1,40	1,04	1,17	2,09	1,51
Comentários/usuário	6,38	6,31	5,62	17,31	7,81
Comentários/post	4,54	6,04	4,82	8,29	5,21

Tabela 3.1: Estatísticas básicas de cada comunidade.

A Tabela 3.1 apresenta estatísticas básicas agregadas sobre o período selecionado. Consideramos todos *posts* e comentários, inclusive aqueles que foram removidos, deletados ou estavam vazios. Neste momento, o objetivo é quantificar o volume de publicações realizadas nas comunidades, independentemente de como elas estão agrupadas nas *threads*. A única restrição desta tabela, é que desconsideramos os *posts* e comentários realizados por usuários deletados (i.e., usuários que foram excluídos do Reddit), porque não conseguiríamos distinguir os autores destas publicações.

Ainda na Tabela 3.1, note que a comunidade Bipolar é a que realiza, proporcionalmente, mais *posts* e comentários. Além disso, a média dos comentários por usuário nesta comunidade é cerca de três vezes maior que nas outras três. Isto é um indício que os usuários nesta comunidade são mais engajados durante as discussões.

O subreddit Depression contém 333.624 usuários únicos, 162.363 no SuicideWatch, 121.945 no Anxiety e 35.081 no Bipolar. Ao verificar os usuários deletados em cada comunidade, este número é grande. Correspondem a 518.644 no Depression, 304.540 no SuicideWatch, 111.652 no Anxiety e 68.591 no Bipolar. Como pode ser observado, a quantidade de ocorrências de usuários deletados é grande em todos subreddits, exceto na comunidade Anxiety eles não superam a quantidade de usuários únicos identificados.

É interessante notar que o subreddit Depression é a maior comunidade em número de usuários, seguida pela SuicideWatch, Anxiety e Bipolar. Ela é cerca de 2 vezes maior

que o SuicideWatch, 2,7 vezes maior que o Anxiety e 9,5 vezes maior que o Bipolar, em número de usuários. É natural que haja interseção entre os conjuntos de usuários dessas comunidades, uma vez que tratam de transtornos mentais. A fim de caracterizar estas interseções, apresentamos a Tabela 3.2, que mostra a porcentagem de usuários em comum entre cada par de comunidades. Por exemplo, a célula da linha SuicideWatch e coluna Depression, mostra que 23,64% dos usuários da comunidade SuicideWatch que também participam da comunidade Depression.

Muitos dos usuários das comunidades SuicideWatch, Anxiety e Bipolar também estão na comunidade Depression (em torno de 21% a 26%). Por outro lado, uma fração bem menor dos usuários da comunidade Depression também estão nas outras comunidades. Parte desta assimetria pode ser explicada pelo fato da comunidade Depression ser a maior dentre as comunidades aqui estudadas.

	Depression	SuicideWatch	Anxiety	Bipolar
Depression	-	11,5%	9,45%	2,22%
SuicideWatch	23,64%	-	6,23%	2,06%
Anxiety	25,85%	8,29%	-	3,25%
Bipolar	21,12%	9,54%	11,29%	-

Tabela 3.2: Interseção dos Usuários. Cada célula é a razão entre cardinalidade de interseção e a cardinalidade do subreddit indicado na linha.

Segundo o número de *posts* por usuário, as comunidades são ordenadas (do maior para o menor) da seguinte forma: Bipolar, Depression, Anxiety e SuicideWatch. Já segundo o número de comentários por usuário, temos: Bipolar, Depression, SuicideWatch e Anxiety. Note que embora o número de *posts* por usuário seja um pouco maior em Anxiety que em SuicideWatch, a ordem se inverte ao considerarmos o número de comentários por usuário. Uma possível explicação é que no subreddit Anxiety há maior busca de ajuda através de postagens e no SuicideWatch o engajamento através dos comentários é superior.

Seguindo o número de comentários por *post*, as comunidades são ordenadas (do maior para o menor) da seguinte forma: Bipolar, SuicideWatch, Anxiety e Depression. É interessante notar que Bipolar, que apresenta o maior número de comentários por *post*, também apresenta o maior número de *posts* por usuário, indicando um alto grau de engajamento dos usuários.

Além disso, a maior parte dos acessos no Reddit foram realizados nos Estados Unidos⁴.

⁴<https://www.alexa.com/siteinfo/reddit.com>

3.2 Análise das Interações

Analizamos as interações entre os usuários através do seguinte modelo matemático. Em cada comunidade, as interações foram representadas por um grafo direcionado $G_d = (V, E_d)$, onde o conjunto de vértices V representa usuários que realizaram um *post* ou um comentário na rede e o conjunto de arestas E_d são interações entre pares de usuários em V . Essas arestas são direcionadas: se o vértice v respondeu a um *post* ou comentário de um vértice u , v aponta para u . O grafo modelado possui pesos nas arestas, onde o peso é igual ao número de interações (respostas a *post* ou comentários) que ocorreram entre cada par de vértices em uma determinada direção. Para o cálculo de algumas métricas, utilizamos uma versão não-direcionada $G = (V, E)$ do grafo, em que cada aresta $(u, v) \in E$ indica a existência de $(u, v) \in E_d$ ou $(v, u) \in E_d$. O peso associado a $(u, v) \in E$ é igual à soma dos pesos das arestas entre u e v , em ambas as direções, no grafo direcionado.

Para caracterizar o grafo de interações entre os usuários calculamos as seguintes métricas clássicas de redes complexas [Newman, 2003]:

- Diâmetro da rede: é a maior distância entre qualquer par de vértices.

$$\text{diam}(G) = \max_{v,u:\text{existe caminho entre } u \text{ e } v} \{d(v, u)\}.$$

- Número de Triângulos: triângulos são triplas não-ordenadas de vértices que possuem conexões entre si.

$$\text{Tri}(G) = |\{\{a, b, c\} : (a, b) \in E \wedge (b, c) \in E \wedge (c, a) \in E\}|.$$

- Transitividade: é a fração de tríades (triplas conectadas em G) que formam triângulos.

$$T(G) = 3 \frac{\text{Tri}(G)}{\text{número de tríades}}.$$

- Excentricidade: é a distância máxima de um vértice v para cada um dos outros vértices da rede. Redes com estruturas de comunidades tendem a apresentar valores mais baixos de excentricidade, uma vez que os vértices estabelecem muitas conexões entre subgrupos diferentes.

$$\text{exc}(v) = \max_{u \in \mathcal{N}(v)} \{d(v, u)\}.$$

- *Closeness*: é o recíproco da distância média entre um vértice v e todos os outros

vértices do grafo. Quanto mais central é o vértice, menor é a sua distância para todos os outros vértices.

$$\text{clo}(v) = \left(\frac{\sum_{u \in V - \{v\}} d(v, u)}{n - 1} \right)^{-1},$$

onde $d(v, u)$ é a distância entre os vértices v e u .

- Coeficiente de clusterização: esta métrica quantifica a tendência que os vizinhos de um nó tem de formar triângulos. Usuários com um valor elevado para o coeficiente de clusterização são aqueles cujos vizinhos tem alta probabilidade de interagir.

$$c(v) = \frac{1}{|\mathcal{N}(v)|(|\mathcal{N}(v)| - 1)} \sum_{u \in \mathcal{N}(v), z \in \mathcal{N}(v) \setminus \{u\}} (\hat{w}_{vu} \hat{w}_{vz} \hat{w}_{uz})^{1/3},$$

onde $\mathcal{N}(v)$ é o conjunto de vizinhos de v em G , \hat{w}_{ij} representa o peso da aresta entre os vértices i e j normalizado pelo maior peso encontrado no grafo $\hat{w}_{ij} = w_{ij} / \max_{(k,l) \in E} \{w_{kl}\}$.

- Grau de Entrada: é calculado pela soma dos pesos das arestas que chegam no vértice v . Na rede, representa o número de vezes em que os usuários responderam algum *posts* ou comentário publicado pelo usuário v .

$$d_{\text{in}}(v) = \sum_{(u,v) \in E_d} w_{u,v}.$$

- Grau de Saída: é calculado pela soma dos pesos das arestas que partem do vértice v . Na rede, representa o número de vezes em que o usuário v escreveu *posts* ou comentários.

$$d_{\text{out}}(v) = \sum_{(v,u) \in E_d} w_{v,u}.$$

3.3 Análise das Threads

Além da análise das interações entre os usuários, estudamos também as interações realizadas no contexto dos tópicos criados pelos usuários. Assim, consideramos as *threads*, que são tópicos criados por um usuário sobre um tema específico, das quais outros usuários podem participar através da publicação de comentários. Então, a *thread* é o conjunto de *post* e seus comentários.

As *threads* são modeladas a partir de uma estrutura de árvore [Laniado et al., 2011], na qual a troca de informação se inicia por um *post* (raiz da árvore), seguido de comentários feitos pelo próprio usuário que gerou o *post* ou por outros usuários.

A modelagem proposta permite analisar, por exemplo, a largura média e a profundidade (máxima) das discussões. A largura média corresponde a quantidade de *posts* e/ou comentários em determinado nível dividido pela quantidade de árvores com profundidade de pelo menos o nível que está sendo considerado. Ou seja, para cada profundidade ℓ , calculamos a largura média no nível ℓ sobre todas as árvores com profundidade ao menos ℓ .

Através da análise da largura e profundidade destas árvores é possível entender quais são os tópicos que geram maior repercussão entre os usuários, bem como se existem usuários-chave que atraem a atenção dos demais membros da comunidade. Entender características das árvores de maior profundidade/largura pode ajudar a entender como os usuários procuram ajuda e como, a mesma, é provida pelos usuários.

Além disso, é importante identificar como os diferentes tipos de publicação – pedidos de ajuda e ofertas de suporte – estão distribuídos na rede. Encontrar os diferentes perfis de usuário em um subreddit é importante para projetar intervenções direcionadas que possam auxiliar na melhoria da saúde mental dos participantes.

3.4 Análise Textual

A fim de explorar os usuários e os conteúdos abordados dentro do Reddit (QP1), realizamos uma análise textual dos *posts* e comentários para caracterizar as comunidades Depression, SuicideWatch, Anxiety e Bipolar. Um dos objetivos da nossa análise é investigar os tópicos mais centrais que os usuários discutem nas quatro comunidades, buscando encontrar semelhanças da forma que os usuários se expressam nestas comunidades. Além disso, extraímos os papéis desempenhados pelos usuários que se destacam nas comunidades (QP2) através do modelo conhecido como Relationship Modeling Network (RMN). Para esta análise, utilizamos apenas os dados do ano de 2017, uma vez que pretendemos analisar os assuntos/tópicos mais recentes da nossa base.

O RMN é uma rede neural recorrente projetada para modelar relacionamentos entre pares de entidades a partir de texto. Um relacionamento em um dado instante é representado por um vetor de pesos sobre k descritores. As entidades não precisam ser da mesma classe: por exemplo, nesta dissertação, as entidades são os usuários e as comunidades que eles interagem. Cada *post* ou comentário corresponde a um momento diferente. Testamos valores de $k = \{15, 30\}$, o melhor resultado foi com $k = 30$.

As palavras são representadas por *word embeddings* de dimensão P , ou seja, cada palavra w de um vocabulário \mathcal{V} é um vetor em \mathbb{R}^P . Usuários e comunidades são representados por *embeddings* de dimensão U e C respectivamente. Assim como nos trabalhos anteriores, iremos utilizar *embeddings* pré-treinados pela técnica GloVe [Pennington et al., 2014]. Os descritores obtidos pelo RMN são vetores em \mathbb{R}^P , permitindo que encontremos as palavras mais próximas a cada um desses vetores. A representação de um *post* ou comentário é denotada por $v_{post} \in \mathbb{R}^P$. Este vetor é igual a média dos *embeddings* das palavras que o *post* contém. As representações dos usuários e das comunidades são denotadas por v_{user} e v_{comm} , respectivamente.

Para cada *post* ou comentário, o RMN é alimentado com um vetor $v \in \mathbb{R}^{P+U+C}$ obtido a partir da concatenação de v_{post} , v_{user} e v_{comm} . Estes vetores são combinados através dos pesos da rede neural para obter uma representação $d_t \in \mathbb{R}^K$ para a relação entre o usuário e a comunidade naquele instante específico. O RMN utiliza um parâmetro de suavização $\alpha \in (0, 1)$ para evitar mudanças bruscas na representação de uma mesma relação em instantes consecutivos, d_t e d_{t-1} . A matriz de descritores $R \in \mathbb{R}^{K \times P}$ é usada para tentar reconstruir o *post* v_{post} , fazendo-se $r_t = R^T d_t$. Os parâmetros do RMN (pesos e matriz dos descritores) são treinados de forma a maximizar uma função objetivo que visa aproximar r_t e v_{post} e distanciar r_t de outros *posts* amostrados aleatoriamente. Veja Iyyer et al. [2016] para mais detalhes sobre o RMN.

A Figura 3.3 apresenta o pré-processamento realizado para extração dos principais tópicos/assuntos dentro das quatro comunidades. Os dados de entrada para o RMN são pré-processados da seguinte forma. Primeiro, são removidos todos os *posts* e comentários marcados como *deleted* ou *removed*, as *stopwords* (listados na biblioteca NLTK) e as pontuações. A seguir, são considerados somente os *posts* e comentários dos usuários que realizaram no mínimo 50 atividades (*posts*/comentários) em cada subreddit, seguindo a metodologia apresentada por Wang et al. [2016]. O objetivo desta filtragem é identificar os principais assuntos de usuários ativos dentro das comunidades, e não assuntos de usuários que aparecem esporadicamente na comunidade. Por fim, são selecionadas apenas as palavras que aparecem em todos os quatro subreddits analisados, buscando encontrar possíveis semelhanças na forma com que as pessoas se expressam quando discutem sobre transtornos de saúde mental.

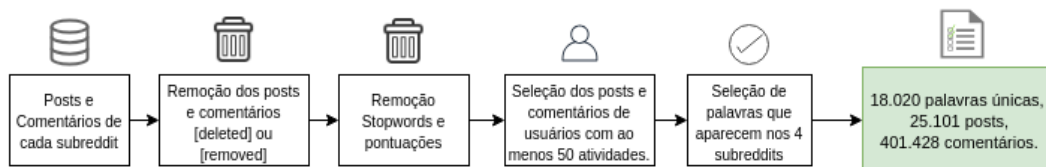


Figura 3.3: Metodologia da Análise Textual

3.5 Análise de Perfis dos Usuários

Identificar grupos de usuários é importante para entender diferentes perfis desempenhados no Reddit (QP2). Assim, é possível analisar as características de cada grupo e, conseqüentemente, como eles respondem às interações (*posts* e comentários) dentro da comunidade. Com esses perfis mapeados, é possível aplicar técnicas de intervenção mais apropriadas para diferentes usuários. Por exemplo, esta análise permitiria concluir que um usuário que possui determinado perfil é um usuário que precisa de mais ajuda.

Nas comunidades abordadas nesta dissertação existem usuários que possuem diferentes perfis de atividades dentro da árvore de discussão. Assim, pretendemos identificar grupos semelhantes neste universo de usuários heterogêneos. Esta análise de perfis dos usuários é centrada nas árvores de discussão, uma que vez que mostra o perfil do usuário, considerando o engajamento na comunidade e com os demais usuários.

Identificaremos diferentes grupos nas comunidades: usuários que participam quase que exclusivamente das suas próprias publicações; usuários que raramente iniciam uma *thread*, mas estão participando constantemente em outras; ou ainda usuários que participam frequentemente com *posts* ou comentários.

Para encontrar os perfis de usuários utilizamos a caracterização das atividades. Atividades correspondem aos *posts* e comentários realizados por um usuário. Assim, para mapear os usuários nos perfis, definimos os seguintes atributos:

1. **Frequência de Atividades:** Frequência de atividades por unidade de tempo (dias). Corresponde à razão entre a quantidade de atividades do usuário dividida e o seu *lifespan* (isto é, diferença, em dias, entre a última e a primeira atividade do usuário).
2. **Tipo de Atividades:** A fração de *threads* das quais o usuário participa que foram disparadas por ele. Equivale a razão entre a quantidade de árvores de discussão iniciadas pelo usuário e a quantidade de atividades que ele participa.
3. **Auto-Engajamento:** Fração dos comentários do usuário que foram realizados em árvores de discussão iniciadas por ele. É a razão entre a quantidade de comentários em árvores de discussão iniciadas pelo usuário e o número total de comentários do usuário.

O primeiro atributo, “frequência de atividades”, distingue os usuários que são mais ativos em uma comunidade durante os seus *lifespan*s. O atributo “tipo de atividades” apresenta os usuários que tem o hábito de realizar mais *posts* do que comentários.

Quando este valor é superior a 0,5 o usuário realiza mais postagens do que comentários. Definimos que o engajamento dos usuários está relacionado aos comentários realizados, quanto mais comentários, mais engajado é o usuário. O último atributo, “auto-engajamento”, distingue os usuários que são engajados em suas próprias publicações, ou seja, priorizam comentar seus próprios *posts*.

Nesta análise consideramos apenas os usuários que tinham pelo menos 10 atividades em cada comunidade. Assim, descartamos os usuários que tem pouca interação na comunidade. Além disso, removemos os usuários “[deleted]”, a fim de não prejudicar nossa análise com atividades cujos autores não são possíveis distingui-los.

Para identificar os grupos de usuários com comportamento semelhante, primeiramente calculamos os três atributos descritos acima para os usuários de cada comunidade. Em seguida, normalizamos os valores de cada atributo, subtraindo-se a média e dividindo-se pelo desvio padrão. Isto é necessário para que todas as variáveis estejam na mesma ordem de grandeza (possuam uma média igual a 0 e um desvio padrão igual a 1), e assim evitar que uma variável com um valor alto se sobressaia na criação dos grupos apenas por causa da sua ordem. O algoritmo de agrupamento *Kmeans* [Hartigan & Wong, 1979] e a técnica *Elbow* [Bholowalia & Kumar, 2014] foram utilizados para definir os grupos e a quantidade de grupos em cada comunidade, respectivamente.

O *Kmeans* é um método de clusterização cujo objetivo é dividir n observações em um número k pré-definido de grupos/centroides. Assim observações próximas estarão no mesmo grupo, mantendo a distância para os centroides tão pequena quanto possível. A técnica *Elbow* consiste em executar o algoritmo k-means várias vezes, com um número crescente de quantidade de grupos k e, em seguida, plotar e analisar o menor k que é capaz de distinguir bem os grupos (isto é, o cotovelo da curva de custo).

3.6 Modelagem da Variação do Tom Emocional

Pretendemos mensurar o estado emocional atual de um usuário e modelar a relação entre as mudanças deste estado e suas interações dentro de um subreddit (QP3). Para isso, avaliamos ferramentas de análise de sentimentos para extração de tom emocional a partir de texto. Tom emocional (estado) é o sentimento (combinação do positivo e negativo) expressado em um texto. Diversos trabalhos fizeram uso do tom emocional em diferentes contextos, como, por exemplo, política e depressão. [Jordan et al., 2018; Coppersmith et al., 2014; Holtzman et al., 2017; De Choudhury et al., 2013]. Em todos estes trabalhos foi utilizada a ferramenta LIWC (Linguistic Inquiry and Word Count) [Tausczik & Pennebaker, 2010] para extração do tom emocional. Entretanto, a forma

de calcular o tom emocional varia com a ferramenta escolhida. Uma vez que definimos utilizar o tom emocional nos textos das comunidades escolhidas, iremos avaliar duas técnicas: LIWC e Vader, assim escolher a mais apropriada.

O LIWC é uma ferramenta de análise textual que calcula um valor para diferentes categorias das palavras, por exemplo: afetividade, processos sociais, artigos, verbos, entre outros. Desta forma, para cada um dos *posts* e comentários, o LIWC retorna variáveis com valores que representam diversos aspectos do texto, como verbos, adjetivos, palavras positivas e negativas que estão no texto. Esses valores variam de 0 a 100. No nosso caso particular, estamos interessados nas seguintes variáveis relacionadas aos sentimentos dos textos: emoção positiva (“posemo”), emoção negativa (“negemo”) e “Tone” (“posemo” - “neg”). Além do LIWC, também analisamos o Vader [Hutto & Gilbert, 2014], que é uma ferramenta de análise de sentimentos baseada em regras e um dicionário léxico, construída em um contexto de mídias sociais. Esta ferramenta calcula 4 valores a partir de um texto, denominados: positivo, negativo, neutro e “compound”. Os três primeiros representam a proporção de texto que cai nestas categorias léxicas e sempre somam 1.0. Já o “compound” é a composição dos três valores léxicos normalizados entre -1 (extremo negativo) e 1 (extremo positivo).

Primeiramente, analisamos a relação do tom emocional do LIWC (“tone”) e do Vader (“compound”) com o tamanho dos *posts* e comentários. A Figura 3.4 apresenta esta análise para o subreddit Bipolar. Observe que no LIWC os textos mais longos (acima de 2500 caracteres) estão quase sempre com “tone” menor que 50. O Vader, por sua vez, mostra uma distribuição mais uniforme, pois localiza textos de comprimento maior que 1000 nos dois extremos do eixo “compound”. Assim, entendemos que o Vader tem um comportamento mais robusto na classificação de textos com diferentes tamanhos, ou seja, o tamanho do texto não enviesava o tom emocional do texto. Os autores Hutto & Gilbert [2014] mostram que o Vader classifica bem revisões de filmes, as quais possuem grandes variações de tamanhos. Além disso, também realizamos outras análises apresentadas na Tabela 3.3 para melhor direcionar a escolha da ferramenta de análise de sentimento. Note que as duas ferramentas utilizam o dicionário léxico e que combinam as emoções positivas e negativas. Entretanto, o Vader funciona melhor para dados de redes sociais, uma vez que o LIWC não considera os seguintes itens léxicos nos sentimentos: os emoticons, gírias e siglas [Hutto et al., 2013]. Os autores Davidov et al. [2010] mostram que esses itens são importantes para análise de sentimentos em um texto social. Além disso, os autores Hutto & Gilbert [2014] apresentaram as frases: “A comida aqui é excepcional” e “A comida aqui é ok.”. Eles afirmam que o LIWC é incapaz de explicar as diferenças na intensidade do sentimento das palavras, assim, os dois textos teriam a mesma pontuação, sendo que a palavra excepcional transmite

mais intensidade que ok.

Diante dessas análises, escolhemos a ferramenta Vader para analisarmos o tom emocional do usuário no texto. Assim, o “compound” do Vader será, em nosso trabalho, equivalente ao tom emocional (TE).

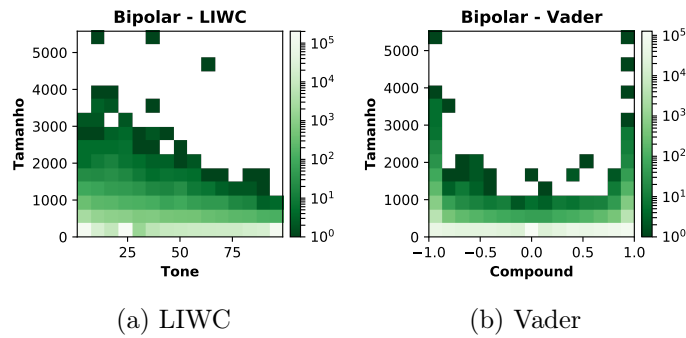


Figura 3.4: Comparação do LIWC e Vader.

Perguntas	LIWC	Vader
Utiliza dicionário Léxico?	X	X
Utiliza dados de redes sociais?		X
Funciona para textos grandes?		X
Combina o sentimento positivo e negativo?	X	X

Tabela 3.3: Análise das Ferramentas de Sentimento.

Um dos principais objetivos deste trabalho é analisar a variação do tom emocional de um usuário em função das suas interações na rede. Pretendemos entender se essas interações possuem algum impacto no comportamento do usuário do ponto de vista do tom emocional.

Finalmente estamos prontos para definir a tarefa de previsão. As entradas do problema são o *post* inicial p_i , com tom emocional t_{p_i} realizado por um usuário u_i e um conjunto de n comentários da mesma *thread* (c_1, c_2, \dots, c_n), onde o comentário c_n , com tom emocional t_{c_n} , é o último comentário realizado pelo usuário u_i . Queremos retornar uma previsão da variação do tom emocional, $d_{(t_{c_n}-t_{p_i})}$, do usuário u_i . Desta forma iremos verificar se houve alguma mudança do sentimento do usuário após as interações.

Capítulo 4

Caracterização das Comunidades

Este capítulo apresenta uma caracterização dos usuários participantes dos subreddits relacionados a transtornos de saúde mental. A Seção 4.1 apresenta uma caracterização geral dos usuários. Na Seção 4.2 analisamos como são as interações dos usuários dentro das comunidades. Na Seção 4.3 analisamos o nível de atividades dos usuários, considerando os *posts* e comentários. Na Seção 4.4 analisamos perfis de usuários considerando aspectos como frequência de atividades e engajamento nas discussões que ocorrem nos subreddits, ou seja, se os usuários concentram sua atenção em discussões iniciadas por eles mesmos ou em discussões de outros usuários. Na Seção 4.5 caracterizamos os *posts* e comentários no contexto das árvores de discussão, inclusive em relação aos tópicos abordados em cada uma das quatro comunidades. Por fim, na Seção 4.6 apresentamos uma caracterização detalhada do tom emocional das publicações feitas pelos usuários, que será a base para a tarefa de predição da dinâmica do tom emocional nestas comunidades, foco do Capítulo 5.

4.1 Caracterização Geral dos Usuários

Analisamos os *posts* e comentários escritos por usuários de cada comunidade em relação ao comprimento (medido em número de palavras) e ao seu “score”. O “score”, também é conhecido como “Karma”, corresponde a quantidade de votos recebidos (podendo ser negativos ou positivos) pelos usuários. É uma forma de avaliação dos *posts* e/ou dos comentários. O autor Van Mieghem [2011] mostra efeitos que este “score” causa nos usuários.

A Tabela 4.1 apresenta a média e a mediana calculadas para as métricas acima. Observamos que o score de um *post* é, em média, 3,9 vezes maior que o “score” de um comentário. Uma possível interpretação é que os autores dos *posts* recebem mais

suporte social do que os autores dos comentários. Outra interpretação é que os *posts* recebem mais votos positivos porque naturalmente atraem mais atenção que os comentários associados a eles. Além disso, note que os valores são, na média, sempre positivos, indicando que os usuários incentivam uns aos outros com votos positivos, ao invés de inibi-los com votos negativos. Apesar da média do “score” do *post* ser acima de 5, em todas comunidades, mais de 50% dos *posts* tem “score” bem menor que média. Isto mostra que existe grande variação entre os “scores” nos *posts* das comunidades.

Em relação ao tamanho médio do texto (em quantidade de palavras), observamos que os *posts* são aproximadamente 2 vezes maiores que os comentários. Além disso, 50% dos *posts* possuem pelo menos 99 palavras no Depression, 105 no SuicideWatch, 105 no Anxiety e 79 no Bipolar. Este número é 54,6% do tamanho médio dos *posts* no Depression, 55,6% no SuicideWatch, 64,6% Anxiety e 57,7% no Bipolar. A razão entre a média e a mediana do tamanho dos comentários para cada comunidade é bem próxima a razão calculada para os *posts*. Em suma, há grande variabilidade no tamanho dos *posts* e comentários em todas comunidades.

	Depression		SuicideWatch		Anxiety		Bipolar	
	Média	Mediana	Média	Mediana	Média	Mediana	Média	Mediana
Score Geral	3,45	1	2,02	1	3,53	2	3,04	2
Score Post	8,25	2	5,06	3	9,22	2	10,59	4
Score Comentário	2,39	1	1,52	1	2,35	1	2,13	2
Palavras Geral	77,21	33	74,32	30	82,34	44	64,47	34
Palavras Post	181,20	99	188,86	105	162,61	105	136,94	79
Palavras Comentário	54,33	29	55,36	27	65,71	40	55,73	32

Tabela 4.1: Score e Tamanho das Palavras nas Comunidades.

A fim de identificar a faixa etária dos usuários que participam dessas comunidades, aplicamos expressões regulares para encontrar a idade do usuário que a informa explicitamente. As expressões regulares utilizadas foram “i’m (d+) years old” e “i am(d+) years old”. Encontramos alguns outliers, como o valor 5000 encontrado na comunidade Depression. Este valor ocorreu em um cenário no qual o usuário usava uma figura de linguagem, como pode ser visto no trecho retirado: “... *I am just so tired emotionally like I am 5000 years old.*” Também encontramos idade igual a 2, retirada de: “*My therapists print off sheets about "happy time"and "worry time"like I'm 2 years old and don't know any better or what's going on...*”

A Tabela 4.2 apresenta os resultados para os usuários que informaram suas idades no ano de 2017. Escolhemos utilizar apenas o ano de 2017, para evitar duplicidades de usuários que aparecem em diferentes anos. Na tabela apresentamos a quantidade de usuários que conseguimos identificar as idades, a média, o desvio padrão e a mediana das idades encontradas. Note que na comunidade Depression, o desvio padrão é o mais alto, justificado pelo trecho que expomos anteriormente (idades como forma de

expressão). Observe que as idades apresentadas são aquelas que os usuários informaram em algum contexto, não sendo necessariamente a idade que possuem. Assim, a idade informada pode ter vários significados, como a idade real, a idade que sentem ter, idade percebida por outros, ou até mesmo uma idade fictícia.

	Depression	SuicideWatch	Anxiety	Bipolar
Usuários Únicos	1448	679	452	142
Média	26,05	22,77	23,04	25,68
Desvio Padrão	131,01	8,58	7,40	9,81
Mediana	21	21	22	24

Tabela 4.2: Caracterização da Faixa Etária dos Usuários de 2017.

4.2 Análise das Interações

Nosso conjunto de dados é composto por postagens e comentários dos usuários que estavam nas comunidades entre 2012 e 2017. Especificamente nesta análise, focamos nos dados que compreendem o período de Janeiro de 2017 a Dezembro de 2017, uma vez que o objetivo é entender a estrutura mais atual das comunidades deste estudo. No total, obtivemos 261.511 publicações e 1.256.669 comentários de 184.708 usuários únicos. O número total de comentários em cada comunidade é pelo menos 4,8 vezes maior que o número de postagens, o que indica a existência de uma rede de suporte ativa entre os usuários.

A Figura 4.1 apresenta o volume mensal de *posts* e comentários em cada um dos subreddits no ano de 2017. É interessante notar que, exceto no subreddit Bipolar, o número mensal de publicações manteve uma tendência geral de crescimento durante todo o ano. Contudo, observa-se uma pequena queda nos meses próximos ao verão no hemisfério norte. Outro ponto a ser ressaltado é que os picos de *posts* e comentários nestas comunidades podem estar correlacionados com eventos reais que geram grande comoção entre pessoas. Por exemplo, se considerarmos o subreddit SuicideWatch, observamos que, entre os meses de fevereiro e início de abril o total de *posts* aumentou em torno de 16% enquanto o total de comentários aumentou em torno de 30%. Este crescimento coincide com a divulgação e lançamento da série *13 Reasons Why*, que gerou forte discussão em torno do tema de suicídios cometidos por adolescentes¹.

A partir de um grafo direcionado com pesos que representam o número de interações do nó A com o nó B, calculamos as distribuições dos graus de entrada e de saída dos vértices. O grau de entrada de um vértice é a quantidade de comentários que o

¹<https://www.theatlantic.com/entertainment/archive/2017/08/13-reasons-why-demonstrates-cultures-power/535518/>

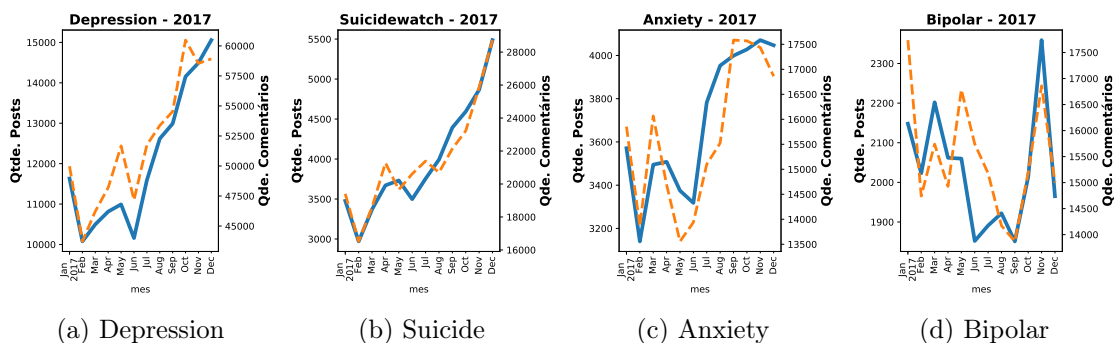


Figura 4.1: Evolução do total de *posts* e comentários em cada subreddit no ano de 2017.

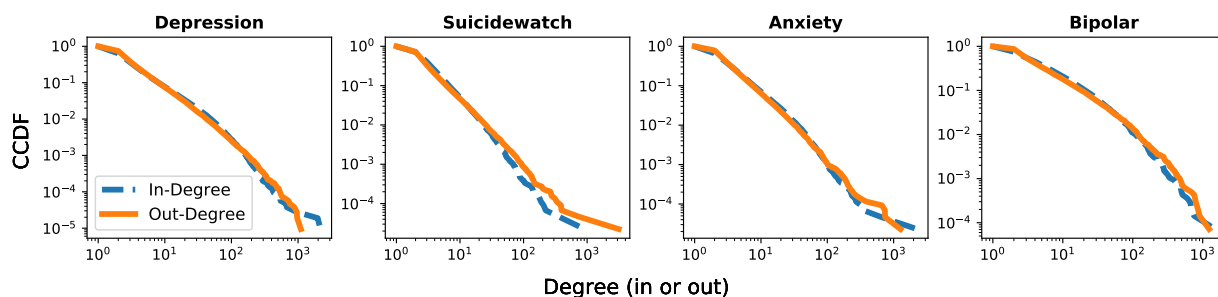


Figura 4.2: Função de Distribuição Cumulativa Complementar (CCDF) dos graus de entrada e saída do usuário.

usuário correspondente recebeu em suas publicações. O grau de saída de um vértice é quantidade de publicações feitas por um usuário.

A Figura 4.2 mostra a Função de Distribuição Cumulativa Complementar (CCDF) dos graus de entrada e saída. Ambas são distribuições de cauda pesada para todos os subreddits. Os maiores valores de grau de entrada e saída são semelhantes, exceto para Depression, em que o maior grau de entrada é cerca de duas vezes maior que o grau de saída. Uma inspeção minuciosa do grau de entrada revela que 63,8% (Depression), 71,5% (Suicide), 66,2% (Anxiety) e 75,2 % (Bipolar) dos usuários receberam pelo menos um comentário em suas postagens. Além disso, observamos que 11,9% (Depression), 8,9% (SuicideWatch), 12,0% (Anxiety) e 20,5% (Bipolar) de usuários comentaram 5 a 20 vezes, sugerindo altos níveis de troca de informações nessas comunidades. Ademais 2,8% (Depression), 1,4% (SuicideWatch), 2,0% (Anxiety) e 8,8% (Bipolar) de usuários contribuem com mais de 20 comentários.

Ambos os graus de entrada e saída são medidas de diversidade de interação. No entanto, também estamos interessados em sua intensidade. Nesse caso, analisamos os pesos das arestas do grafo, que representam o número de comentários feitos/recebidos

por um usuário. Cerca de 84,7% dos membros interagem um com o outro apenas uma vez no subreddit Depression, 73,6% no SuicideWatch, 85,7% no Anxiety e 81,5% no Bipolar. Um número não desprezível de membros interage entre si de 2 a 20 vezes: 15,2% no Depression, 26,3% no SuicideWatch, 14,3% no Anxiety e 18,5% no Bipolar.

Uma outra maneira de quantificar o volume de *posts* e comentários em um subreddit é mensurar a quantidade de interações entre pares de usuários. Dizemos que houve uma interação entre um par de usuários (i, j) quando i comenta em uma publicação de j ou vice-versa. A Figura 4.3 apresenta o histograma contendo o número de pares encontrados quantizados por números de interações. Podemos observar que a grande maioria dos usuários estabelece poucos diálogos entre si (primeira coluna dos histogramas apresentados). Este resultado corrobora a filosofia da rede social Reddit, onde o objetivo principal é o engajamento em torno de conteúdos em que os usuários possuem interesse, independentemente dos usuários que participam da discussão (dado que a identidade dos usuários sequer é revelada, em muitos casos).

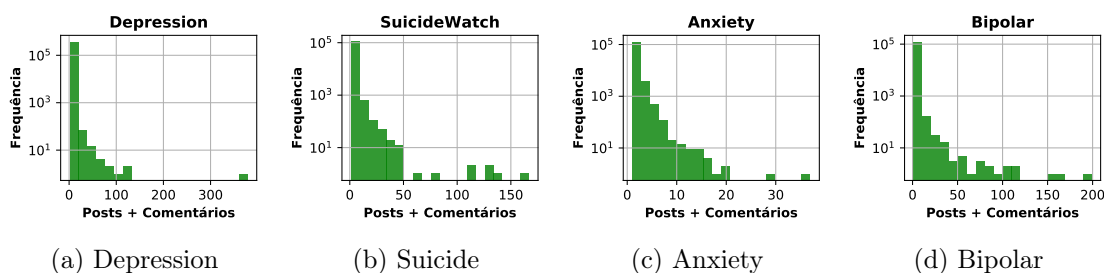


Figura 4.3: Histograma de números de pares de usuários quantizados por volume de interação.

Calculamos várias métricas estruturais a partir do grafo de interação de cada subreddit. Mais especificamente, calculamos o diâmetro da rede (maior distância entre nós), transitividade (fração de todos os triângulos possíveis que existem em G), além de estatísticas – média, mediana, coeficiente de variação (CV) – das métricas dos nós [Newman, 2003], especificamente, excentricidade, *closeness*, coeficiente de *clustering* e número de triângulos. A Tabela 4.3 mostra os resultados obtidos para cada métrica. No geral, observamos que os padrões de interação são muito semelhantes entre os subreddits. A baixa transitividade combinada com alto diâmetro e excentricidade do nó corrobora o fato de que as interações do usuário são baseadas principalmente no conteúdo das postagens e comentários, independentemente dos usuários que os geram. Este comportamento é encorajado pelo Reddit, que não tem nenhum filtro de conteúdo baseado em amizades e proíbe os “anéis de votação” em seu termo de serviços. Nós conjecturamos que esta é também a razão para os baixos valores de *closeness*. Também

Subreddit	Métrica Grafo		Métrica por nó											
			Excentricidade			Centralidade Closeness			Centralidade Clustering			Número de Triângulos		
	Diâmetro	Transitividade	Mediana	Média	CV	Mediana	Média	CV	Mediana	Média	CV	Mediana	Média	CV
Depression	12	0,007	8	8,47	0,0736	0,1262	0,0854	0,8361	0,00	0,029	4,495	0	2,300	18,332
SuicideWatch	15	0,001	10	10,17	0,0816	0,1028	0,0752	0,7380	0,00	0,022	5,505	0	0,315	19,105
Anxiety	12	0,004	8	8,27	0,0813	0,1262	0,0872	0,8039	0,00	0,033	4,303	0	0,889	16,234
Bipolar	8	0,039	6	5,65	0,1103	0,2339	0,1822	0,6021	0,00	0,127	1,731	0	58,007	8,335

Tabela 4.3: Métricas do grafo de interação do usuário por subreddit.

observamos uma grande variação (em termos de CV) na distribuição do número de triângulos, indicando que esse número varia substancialmente em diferentes nós. Essa é uma consequência direta da grande variação no número de *posts* e comentários realizados por diferentes usuários.

Em suma, nossa análise mostra que nesses quatro subreddits relacionados à discussão sobre saúde mental, o modelo de interação é centrado em torno do conteúdo das postagens e comentários, e não em usuários. O que é positivo, pois este modelo ajuda os novos usuários a iniciar sua participação na rede, uma vez que impede a formação de grupos muito unidos.

4.3 Análise das Atividades dos Usuários

Neste trabalho, definimos como atividades todos os *posts* e comentários feitos por cada usuário em cada subreddit. Esta análise das atividades dos usuários contempla todos os anos analisados (2012 a 2017).

A Figura 4.4 apresenta a variação do volume de atividades em cada comunidade, em diferentes horas do dia. Observamos que as comunidades Anxiety e Bipolar têm os maiores períodos de atividades, durante a madrugada e no período da noite, enquanto as comunidades Depression e SuicideWatch registram os picos somente na madrugada. É importante saber que o banco de dados Reddit usa o UTC, assim, todas as análises estão neste fuso. As quatro comunidades apresentam o menor volume de atividades durante o horário da manhã. Além disso, a razão entre o número máximo e o mínimo de *posts* varia entre 2,5 (Anxiety) e o 2,8 (SuicideWatch) ao longo do dia. Já a mesma razão calculada com base nos comentários varia entre 2,3 (Depression) e 2,9 (Bipolar).

A Figura 4.5 apresenta as atividades dos usuários em cada dia da semana. Os maiores níveis de atividades ocorrem às segundas-feiras, com exceção do subreddit Bipolar, cujo o pico acontece na quarta-feira. Em contrapartida, os menores níveis de atividades ocorrem aos sábados nos 4 subreddits. A diferença de postagens entre a segunda-feira para o sábado é 18,2% no Depression, 15,9% no SuicideWatch, 23,9% no Anxiety e 18,2% no Bipolar.

Com relação aos comentários, o pico de atividades representadas segue o mesmo

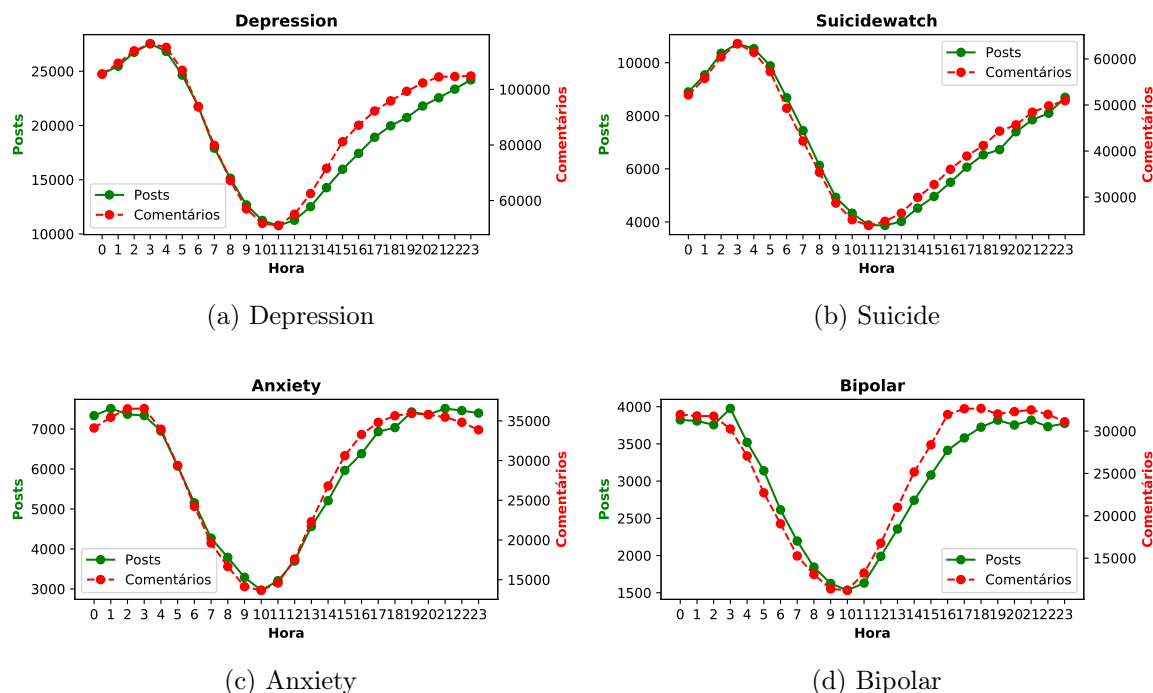


Figura 4.4: Atividades dos usuários (total de posts e comentários) em cada hora do dia.

comportamento dos *posts*. O decaimento, em relação da segunda-feira e o sábado, é 16,4% no Depression, 14,1% no SuicideWatch, 19,4% no Anxiety e 14,1% no Bipolar. Os valores, em termos percentuais, nas comunidades, são muito semelhantes. Também observamos que sempre a queda nos *posts* é maior que a dos comentários. A comunidade que mais se difere das outras é a Anxiety que tem maior queda, tanto nos *posts* quanto nos comentários.

É interessante notar que as ocorrências de picos de atividades às segundas-feiras seguem o efeito “Blue Monday” apresentado por Stone et al. [1985], onde as segundas-feiras são frequentemente associadas a restrições, como de lazer e felicidade. Por outro lado, a pesquisa apresentada por Cranford et al. [2006] mostra que, geralmente, o humor negativo diminui e a energia aumenta durante os finais de semana, o que pode justificar a queda acentuada de atividades nessas comunidades aos sábados e domingos.

A Figura 4.6 apresenta a quantidade de atividades em cada mês do ano. Em todas as comunidades os dois primeiros meses são marcados por menos atividades. Também confirmamos esse comportamento em uma análise ano a ano. Provavelmente, o número menor de mensagem se deve às promessas e esperanças que o ano novo traz às pessoas. Ao longo dos meses, as atividades aumentam, mostrando que a “euforia” típica de início de um novo ciclo se perde e as pessoas voltam a focar nos problemas e

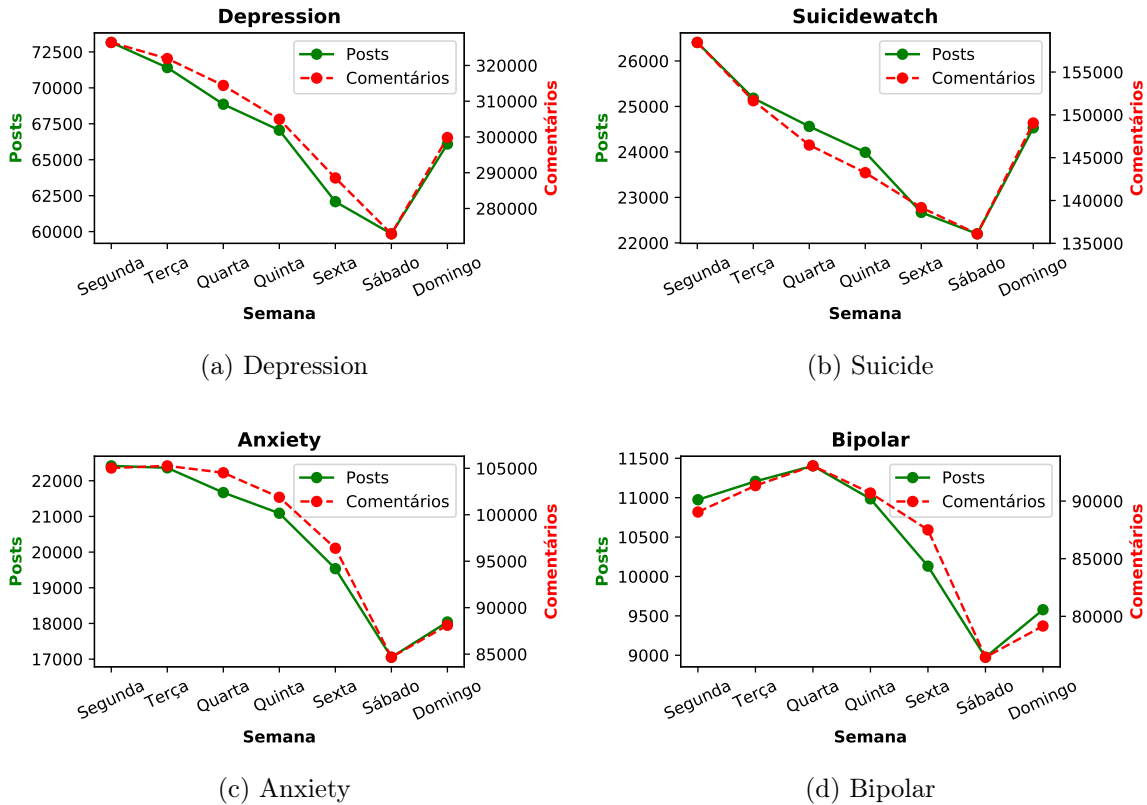


Figura 4.5: Atividades dos usuários em cada dia da semana.

frustrações encontrados no dia a dia. A comunidade Bipolar tem um comportamento diferente das demais: a partir do terceiro mês do ano ocorrem oscilações das atividades dos usuários. O crescimento do número de *posts* nas comunidades, considerando o intervalo entre os meses de Janeiro e Dezembro, foi de 34,4% no subreddit Depression, 31,7% no SuicideWatch, 22,7% no Anxiety, 22,6% Bipolar. Nos comentários o aumento foi de 28,6% no Depression, 25% no SuicideWatch, 15,6% no Anxiety, e 8,3% no Bipolar.

Analisamos os intervalos entre atividades de um mesmo usuário. A Figura 4.7 apresenta a CCDF (Função de Distribuição Cumulativa Complementar) da diferença, em dias, entre duas atividades consecutivas de um mesmo usuário. Observa-se que mais de 80% das atividades consecutivas ocorrem no mesmo dia. Em raríssimas ocasiões, os intervalos chegam a 2000 dias. Esta distribuição decai mais rapidamente que uma lei de potência.

Analisamos também o volume de publicações realizadas por usuário. As Figuras 4.8 e 4.9 apresentam, respectivamente, a distribuição da quantidade de *posts* e de comentários por usuário, em cada comunidade. Os usuários que foram deletados foram desconsiderados, pois eles representam uma grande quantidade na base e são

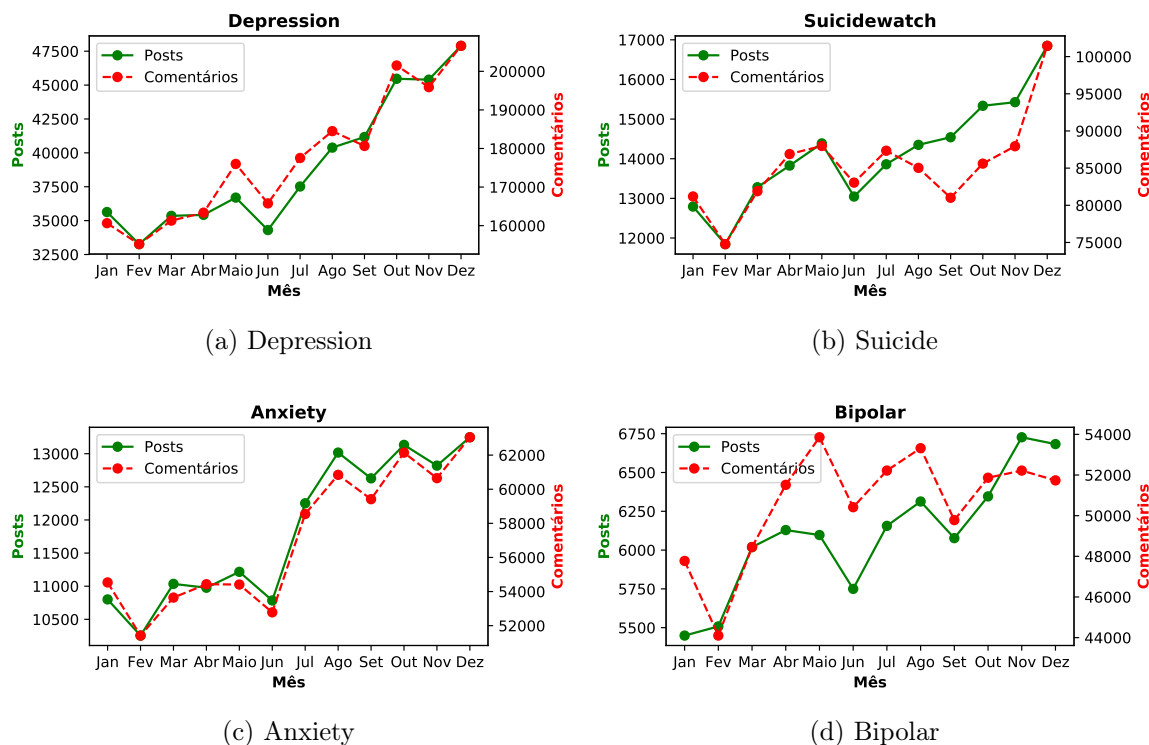


Figura 4.6: Atividades dos usuários em cada mês do ano.

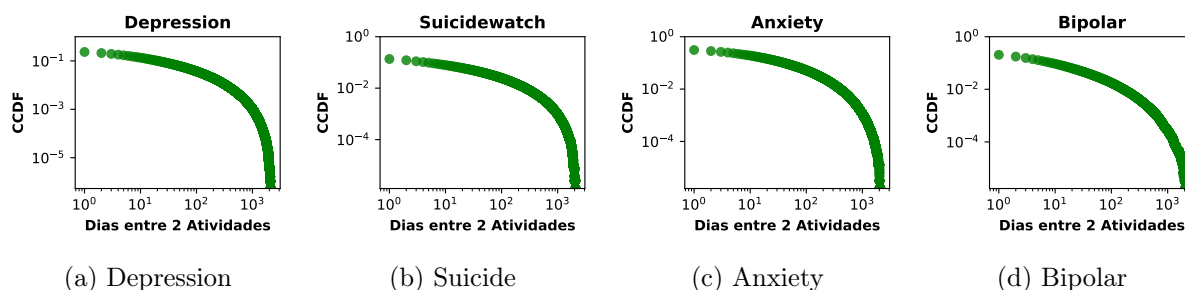


Figura 4.7: Tempo, em dias, entre duas atividades.

marcados como “[deleted]”, caso o considerássemos teríamos muitos “outliers” e eles não representariam o mesmo usuário.

A Figura 4.10 apresenta o *lifespan* de um usuário em cada comunidade. O *lifespan* foi calculado como a diferença, em dias, entre a primeira e a última atividade do usuário dentro da comunidade. Conseqüentemente, os usuários que tiveram apenas uma atividade na comunidade, não foram considerados. Tais usuários representam na comunidade Anxiety 40%, na Bipolar 31,2%, na Depression 42,4% e na SuicideWatch 42,3%. A distribuição do *lifespan* é muito semelhante entre as comunidades. Como

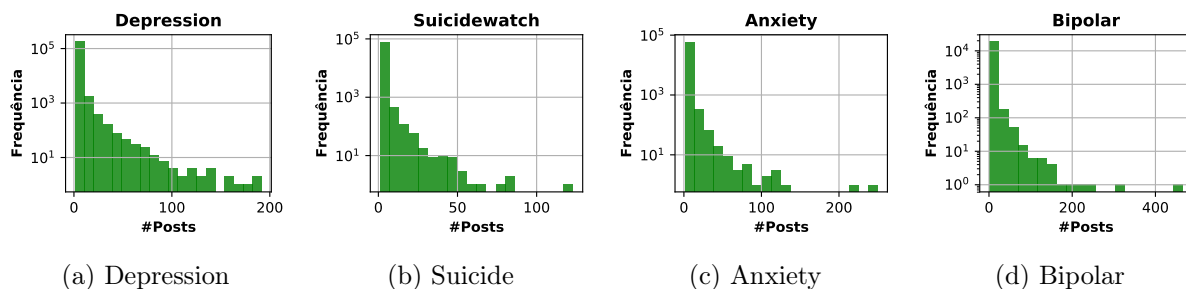
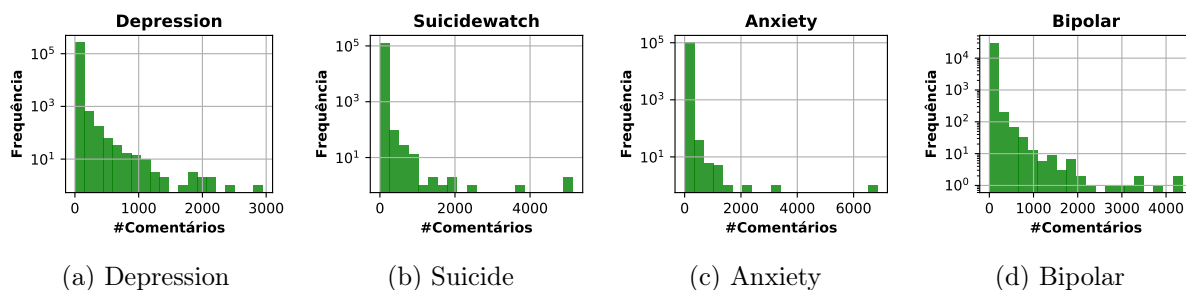
Figura 4.8: Número de *Posts* por usuário.

Figura 4.9: Número de Comentários por usuário.

os dados analisados correspondem ao intervalo de 2012 a 2017, o *lifespan* mais longo possível é 2190 dias. Observamos que há pouquíssimos usuários com *lifespan* maior que 2000 em cada uma das comunidades, e que a maior parte dos usuários permanece por menos de 100 dias. A mediana do *lifespan* em dias da comunidade Depression é 9,6922, SuicideWatch 0,8436, Anxiety 25,0902 e Bipolar 39,739.

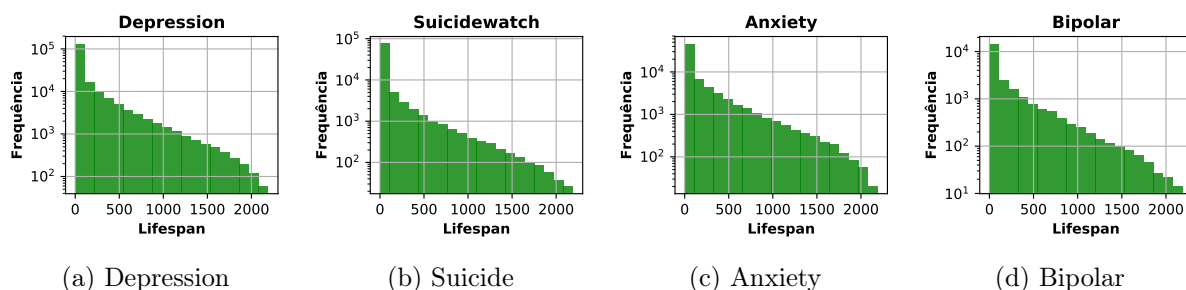


Figura 4.10: Lifespan em dias por usuário.

4.4 Análise dos Perfis dos Usuários

A partir do agrupamento dos usuários com base nos atributos que descrevem suas atividades (frequência de atividades, tipo de atividades e auto engajamento), encontramos quatro clusters em cada uma das comunidades. O tipo de atividades e o auto engajamento estão associados com a participação dos usuários nas árvores de discussões. Usuários que possuem valores altos de auto engajamento podem ser considerados “egocêntricos”, uma vez que sua participação se dá, na maior parte das vezes, em suas próprias árvores de discussão. Por outro lado, usuários que participam de muitas árvores de discussões criadas por usuários podem ser considerados “altruístas”. Os usuários com um valor alto no atributo frequências de atividades, são aqueles que são mais participativos na realização de *posts* e comentários, podendo ser considerados “ativos”.

Observe que estes atributos são como dimensões independentes: um usuário pode ser bem ativo e ter como foco principal suas próprias *threads*. Contudo, o algoritmo de clusterização irá agrupar os usuários segundo as características que destoam mais intensamente da média.

A Tabela 4.4 apresenta o percentual de usuários em cada grupo. Como os clusters foram calculados de maneira independente para cada comunidade, reindexamos os grupos encontrados de forma que grupos com mesmo índice representassem usuários com distribuição de atributos similares. Coincidentemente, a proporção entre os grupos ficou semelhante para as diferentes comunidades.

Grupo	Percentual de usuários em cada grupo			
	Depression	SuicideWatch	Anxiety	Bipolar
0	17,36%	28,15%	22,08%	17,86%
1	0,80%	3,38%	0,12%	0,04%
2	13,32%	8,39%	11,02%	14,98%
3	68,52%	60,07%	66,78%	67,13%

Tabela 4.4: Percentual de usuários em cada grupo por comunidade.

Para a visualização da distribuição de cada atributo, retiramos a padronização. A Figura 4.11 apresenta a distribuição das características em cada grupo dos subreddits Depression, SuicideWatch, Anxiety e Bipolar.

O grupo 0 contém os usuários com maior auto engajamento. Assim, o consideramos como o grupo mais egocêntrico. O grupo 1 contém os usuários com maior frequência de atividades por unidade de tempo. Este mesmo grupo possui, exceto pelos *outliers*, um auto-engajamento relativamente baixo. Além disso, a fração de atividades do tipo *post* é baixo em relação aos demais grupos. Assim, o Grupo 1 é predominantemente composto por usuários ativos, que comentam bem mais do que

postam e que são altruístas, uma vez que atuam mais comentando outros *posts*. O grupo 2 é marcado por usuários um pouco menos ativos que aqueles dos outros grupos. Além disso, são usuários com o segundo maior auto engajamento, ou seja, realizam muitos comentários em suas próprias publicações. Ademais, é um grupo com a razão mais elevada de *posts*/comentários. Em suma, os usuários menos ativos, relativamente egocêntricos e que mais realizam postagens, prevalecem no grupo 2. Os usuários do grupo 3 são os mais altruístas dentre todos os grupos, e se distinguem do grupo 1 por não realizar atividades com frequência acima da média.

Os clusters obtidos para os outros subreddits são semelhantes àqueles obtidos para o subreddit Depression. A única diferença notável é que, nas comunidades Anxiety e Bipolar, o Grupo 1 tem auto-engajamento menor que o Grupo 3. Contudo, estes são os dois grupos menos egocêntricos, independentemente do subreddit analisado.

4.5 Análise dos Posts e Comentários

Nesta Seção apresentamos as análises dos *posts* e comentários dos usuários sob dois pontos de vista: (i) através da análise *thread* de discussão, que podem ser modeladas como da árvore de discussão para identificar o engajamento dos usuários em pedir ajuda e serem atendidos por outros usuários; (ii) através dos descritores que caracterizam os *posts* e comentários dos usuários, revelando, por exemplo, quais são os temas mais discutidos nestas comunidades e possíveis papéis dos usuários.

4.5.1 Análise das Threads

Além da análise das interações dos usuários na rede, também analisamos as *threads*, utilizando a estrutura de árvores de discussão formadas pelos *posts* e comentários.

A Figura 4.12 mostra a distribuição da profundidade das árvores de discussão em cada um dos subreddits. Note que nos subreddits Depression e SuicideWatch, quase a totalidade das árvores de discussão possuem profundidade menor que 60. Nos subreddits Anxiety e Bipolar, a quase totalidade das árvores de discussão possuem profundidade de até 20 (1/3 da profundidade dos primeiros dois subreddits).

A profundidade máxima da árvore do subreddit Depression é 142, muito maior que as do SuicideWatch (93), Bipolar (37) e Anxiety (34). Árvores de discussão com mais de cinco níveis são mais frequentes em SuicideWatch (12,09%), seguido por Bipolar (7,68%), Depression (5,32%) e Anxiety (3,73%). As árvores de discussão mais profundas tendem a estar relacionadas a lutas emocionais e problemas enfrentados pe-

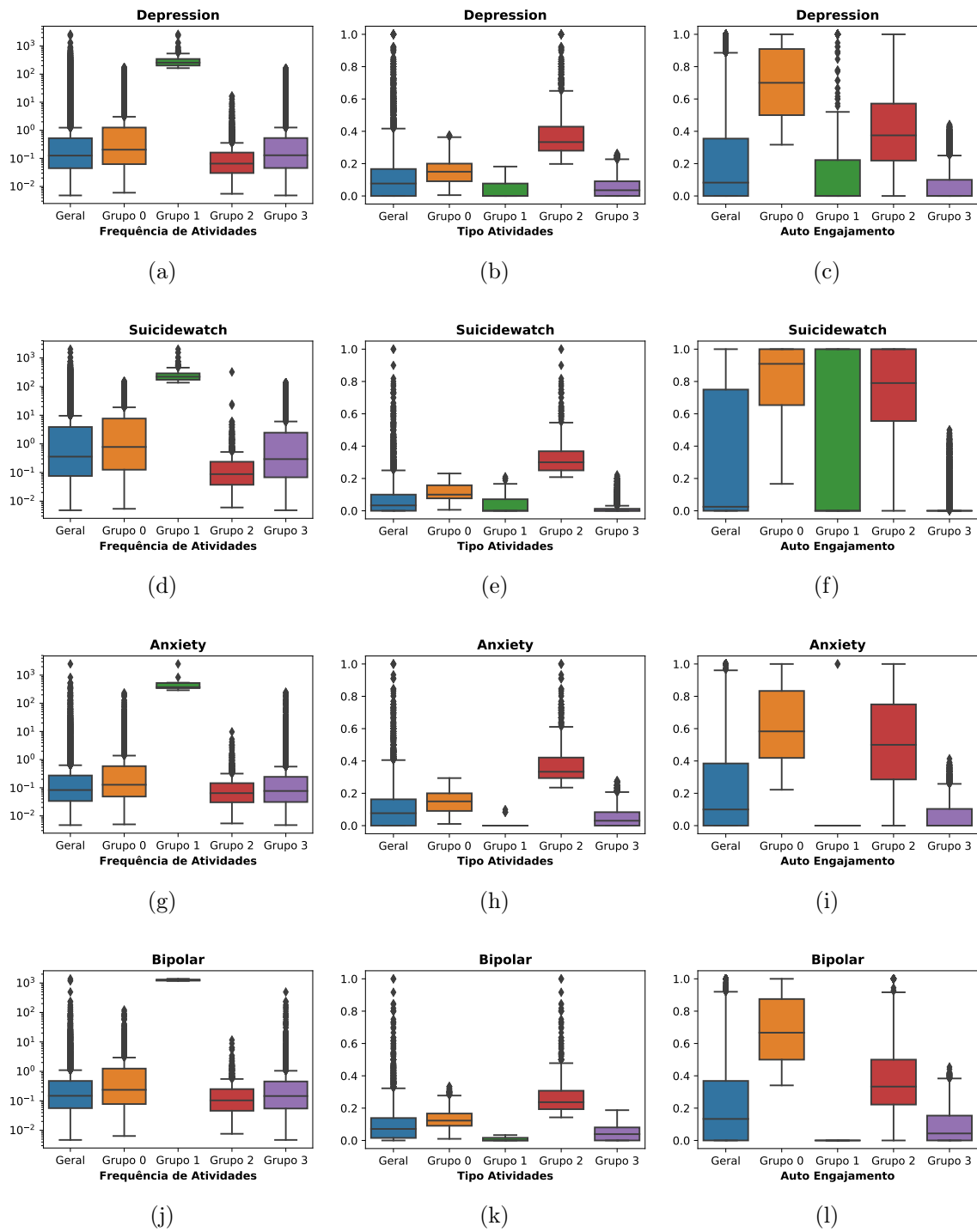


Figura 4.11: Grupos em cada Subreddit Por Característica.

los usuários e a seu pedido de apoio. Por exemplo, a árvore de discussão mais profunda da Depressão foi disparada por um usuário que é um moderador da comunidade.

Analisamos também o conteúdo presente nas 12 árvores de discussão mais pro-

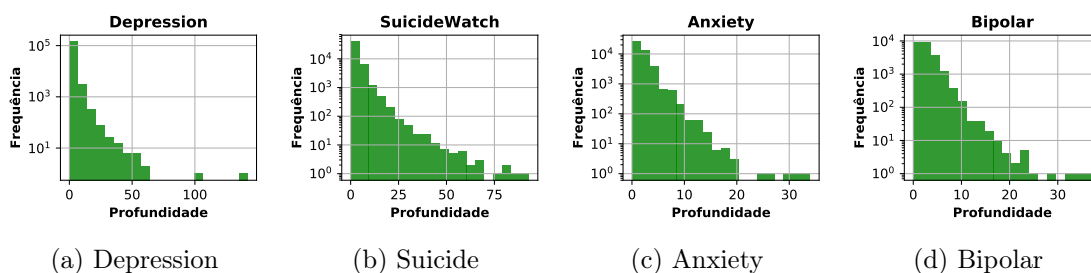


Figura 4.12: Histograma da profundidade das árvores de discussão.

fundas: uma do Depression, quatro do SuicideWatch, quatro do Anxiety e três do Bipolar. Realizamos uma pequena pesquisa pedindo que 12 participantes voluntários lessem cada um dos *posts* que geraram essas árvores e que indicassem a alternativa que melhor descreve a intenção do usuário que escreveu o *post* inicial: (i) procurando ajuda, (ii) oferecendo ajuda, (iii) indiferente ou (iv) outro -, apenas uma resposta poderia ser escolhida. A Tabela 4.5 mostra dois trechos dos *posts* e os resultados das classificações realizadas pelos participantes.

Posts	Classificação
"Help me get out of mixed state Now I've been kinda manic recently. I also foolishly despite knowing the risks went back on Zoloft a few weeks ago. I have never really got mixed states when not on Sertraline (Zoloft)... Anyway I'm totally fucking panicking about how to prevent or get myself the fuck out of going into a mixed state..."	83,3% Procurando ajuda
	16,7% Oferecendo Ajuda
	0% Indiferente
	0% Outro
"Hello fellow /r/depression posters, lurkers, and those passing through, I am very sorry about the delay between weekly posts going up.Welcome to the weekly post - a place to take a moment and share what is going on and how you are doing. If you have an accomplishment you want to talk about, or are having a tough time but don't feel..."	0% Procurando ajuda
	58,3% Oferecendo Ajuda
	33,3% Indiferente
	8,3% Outro

Tabela 4.5: Dois exemplos de itens na pesquisa sobre a intenção do post e respectivos resultados.

Os participantes da pesquisa eram alunos de Ciência da Computação e não conheciam ou tiveram contato com usuários que escreveram as postagens. A concordância entre os avaliadores medida pela estatística Kappa Fleiss [Fleiss, 1971], sendo superior a 0,3178 para todos os itens, o que indica concordância justa, no pior caso. Esta pesquisa revela intuições interessantes sobre como os membros falam sobre seus sentimentos e lutas emocionais:

1. 7 de 12 árvores analisadas foram desencadeadas por um membro em busca de ajuda.
2. As árvores de discussão mais profundas são aquelas em que as pessoas estão mais interessadas em tentar compreender melhor como a pessoa que iniciou o tópico está enfrentando seus problemas na vida real ou seus sentimentos atuais.

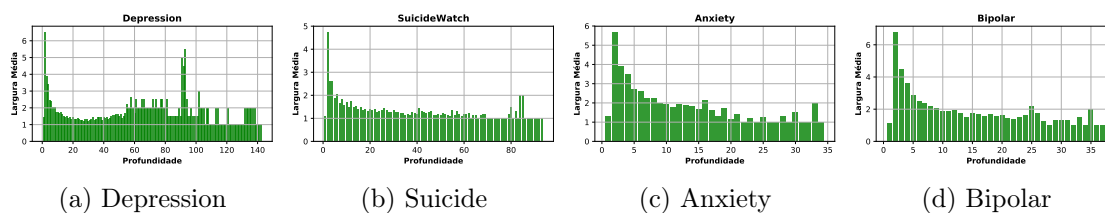


Figura 4.13: Largura média por nível das árvores de discussão (condicionado a profundidade $>$ nível). Nós de mesma profundidade estão no mesmo nível da árvore.

- Os membros que iniciam a árvore de discussão com um *post* com mais de 50 palavras (11 de 12) tendem a atrair comentários que são longos também (número médio de palavras por comentário variando de 32,05 a 259,48). Na comunidade SuicideWatch, uma árvore iniciada por um *post* com 160 palavras recebeu 88 comentários, com média de 69,61 palavras por comentário. Isto corrobora os resultados apresentados em Althoff et al. [2016].

Dando sequência à análise das árvores de discussão, a Figura 4.13 mostra a largura média das árvores de discussão para cada nível de profundidade. A largura média é a razão entre quantidade de *posts* e/ou comentários em determinado nível e a quantidade de árvores cuja profundidade é maior ou igual ao nível que está sendo considerado.

A partir da Figura 4.13 podemos observar que, exceto para o subreddit Depression, as maiores larguras médias estão concentradas nos primeiros níveis de profundidade. No caso do Depression, larguras médias maiores que 5 (cinco) são encontradas em níveis de maior profundidade (entre 85 e 95). Uma possível explicação para este resultado é a existência de moderadores que postam semanalmente incentivando os usuários a falar dos seus problemas. Assim, um comentário ao *post* pode gerar uma árvore de grande profundidade e eventualmente existirão mais comentários em cada nível, resultando em larguras diferentes.

4.5.2 Análise Textual: Descritores

A Tabela 4.6 apresenta os principais termos associados aos descritores encontrados pelo RMN quando treinado com o texto de *posts* e comentários de todos os subreddits. Os subreddits são considerados de maneira conjunta, pois desejamos obter descritores comuns a todas as comunidades. É interessante notar que o modelo captura um conjunto considerável de descritores *coerentes* (i.e., os termos principais dos descritores são relacionados), mostrando que os quatro subreddits analisados possuem uma linguagem em comum que está intimamente relacionada aos transtornos de saúde mental (veja,

Tópico	Descritores
0	realistic practical perspective perception innate objective value ideal aligned reframe
1	ideally hesitate available anytime soonest preferably appointments arrange offered pm
2	indecision duress disco uncontrolled negating catastrophising accompanying aggravates curved eliminates
3	selfish die kill neither fault blame killing anybody hurt abandon
4	clearance excite fulltime relocating salary earn savings electrician lucrative funds
5	ah stardew btw soundcloud youtu portugal awesome goo huhahaha
6	podcasts relaxing videos skate parks asmr movies distracting cafe music
7	resolve heal acknowledge overcome reassure gently encourage seek confront subtly
8	world life beautiful future inside imagine reality person somehow feels
9	met dating university hs group dated girlfriends college friendships acquaintances
10	unconfident overreacting rly teenager immature teenagers labelled ashamed unloved embarrassed
11	headaches fog fatigue heavy intense aches periods levels exhaustion moderate
12	worthless waste achievements loser useless meaningless pointless bullshit effort homework
13	disorders discriminate individuals patients prevalence uneducated definitions diagnoses diagnose clinical
14	heater cocoa headless shrimp champagne sundae brittle tray 22yrs headlight
15	accuses memy recorder reorganized uninvited drifts inserting nearer unsuspecting clone
16	hugs wishes rooting buddy xx thank bless hug hope luck
17	moved sat went dropped stayed walked drove ran home discharged
18	bupropion buspar welbutrin zoom anxiolytic tolerated zoloft geodon lamotrigine topamax
19	blacklist cdn eood 270 relationshipadvice freecompliments nytimes 736x 4chan scientificamerican
20	etc often stress anxious makes sometimes physical constantly social deal
21	14th frasier coincide ummmm roughest 2001 alternating 13th noteworthy 30th
22	doc appt pdoc dr thursday gp monday wednesday tuesday yesterday
23	flunk arsed whatcha definitively islamic physics duh honors retake ged
24	comment responses replying sounded thoughtful appreciate sincere commented reply responding
25	habit push pushing drinking avoid start habits slow stop edge
26	yrs 68 euros 20m twelve centuries trillion 204 sixteen machines
27	hii sympathies onslaught awwwww arg sterile agh youuu glee sweety
28	books book useful art published research recommend coloring journal bought
29	journey vacation luck easter smoothly rough merry holidays zealand ride

Tabela 4.6: Descritores e seus termos principais.

Tópico	Contexto
1	Posts/Comentários que oferecem suporte social
3	Posts/Comentários que relatam pensamentos suicidas
7	Posts/Comentários que refletem encorajamento
8	Posts/Comentários que refletem melhora do usuário frente ao problema de saúde mental
9	Posts/Comentários que refletem problemas de relacionamento
10	Posts/Comentários que refletem problemas de usuários durante a adolescência
12	Posts/Comentários de usuários desesperançosos
16	Posts/Comentários de usuários de agradecimentos
24	Posts/Comentários que agradecem a uma ajuda
28	Posts/Comentários que sugerem atividades para amenizar uma condição emocional negativa

Tabela 4.7: Contexto de alguns descritores.

por exemplo, o descritor 3). Os resultados apresentados a seguir focam na análise dos descritores que estão em negrito na Tabela 4.6, que foram escolhidos por representarem temas bem definidos. A Tabela 4.7 lista possíveis interpretações dos *posts* e comentários que possuem maior probabilidade de estarem relacionados aos tópicos destacados.

Cada usuário em um subreddit tem um conjunto de comentários e *posts*, denominados *spans* pelo RMN. Ao final da execução, o RMN associa cada *span* a um vetor de probabilidades sobre os tópicos. Para analisar os tópicos que compõem um subreddit, calculamos a média dos vetores associados aos respectivos spans. Esta média pode ser

interpretada como a esperança da distribuição de probabilidade sobre os tópicos ao selecionarmos um *span* de maneira uniforme aleatória. A Figura 4.14 mostra as probabilidades médias associadas aos tópicos listados na Tabela 4.7 para cada subreddit. Observamos que:

1. O **Tópico 8**, dentre os 10 analisados, é o que ocorre com maior probabilidade nas quatro comunidades. Este tópico contém descritores positivos, dando indícios de que os usuários dos subreddits oferecem suporte e ajuda uns aos outros, principalmente através de palavras de otimismo e de perspectiva de um futuro melhor. O **Tópico 7**, relacionado a termos de encorajamento, ocupa posição de destaque em todos os subreddits.
2. O **Tópico 3** é o segundo mais frequente nas comunidades Depression e SuicideWatch. Este tópico é negativo e remete a descritores de morte, ferimento, culpa, abandono, entre outros, sendo um resultado esperado em comunidades centradas na discussão de problemas de saúde mental relacionados à depressão e ao suicídio.
3. Os subreddits Depression e SuicideWatch possuem papéis bem semelhantes, quando comparamos estes 10 tópicos. No entanto, dois tópicos os diferenciam: (i) o **Tópico 12** aparece com maior probabilidade no subreddit Depression. Este tópico remete a descritores com sentimento de inutilidade e desperdício, o que é um indício que as pessoas da comunidade Depression expressam mais seu sentimento de inutilidade e; (ii) o **Tópico 16**, que remete a descritores de agradecimento, aparece em maior probabilidade no subreddit SuicideWatch, o que pode indicar que usuários desta comunidade respondem de maneira mais afetuosa ao apoio social oferecido.
4. O **Tópico 9**, que remete à lugares e grupos de relacionamentos (amigos, namoradas, universidade, encontros), aparece com maior probabilidade nas comunidades Depression e SuicideWatch, o que sugere que os problemas de saúde mental dos usuários dessas comunidades podem ter sido desencadeados, principalmente, devido à relacionamentos problemáticos com os pares.
5. O **Tópico 10** é sobre insegurança e adolescência, sendo mais prevalente na comunidade SuicideWatch, possivelmente refletindo problemas de *bullying* comuns nesta fase da vida.

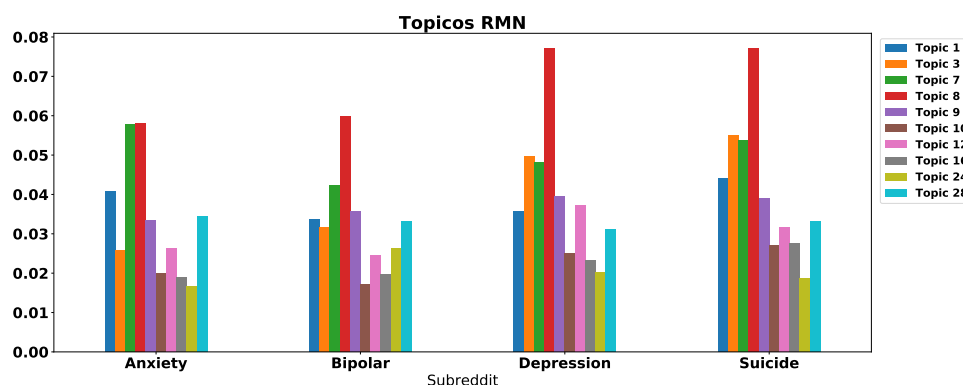


Figura 4.14: Probabilidade de cada post ou comentário ser associado a um tópico.

6. O **Tópico 28** tem probabilidade de ocorrência similar nas quatro comunidades, revelando que usuários tendem a sugerir atividades que possam ocupar a mente (por exemplo, ler um livro).

Analisamos também *posts* e comentários feitos pelos usuários, auxiliando em uma possível definição de papéis nas comunidades. Para isto, calculamos a média dos vetores associados aos *spans* de cada de usuário. Para cada tópico da Tabela 4.7, descrevemos a seguir o perfil dos cinco usuários cuja probabilidade associada ao tópico é máxima. A seguir, iremos correlacionar o comportamento destes usuários com os tópicos encontrados pelo RMN:

1. Conforme descrito anteriormente, o **Tópico 1** se caracteriza por descritores que mostram uma tendência de suporte social. Dos cinco usuários com maior massa de probabilidade associada a este tópico, quatro fazem parte da comunidade SuicideWatch e um da comunidade Depression. Ao mesmo tempo, a probabilidade de relevância dos *spans* destes usuários em tópicos com descritores mais negativos (3, 10 e 12) é pequena. Provavelmente estes usuários tem um perfil de conselheiro, oferecendo ajuda aos demais membros destas comunidades.
2. O **Tópico 3** possui descritores com alto teor de negatividade. Dos cinco usuários com maior probabilidade média de relevância neste tópico, três fazem parte da comunidade SuicideWatch e dois da comunidade Depression. Um ponto interessante é que, apesar destes usuários possuírem um discurso que tende a ser muito negativo, os mesmos possuem um probabilidade relativamente alta associada ao **Tópico 8**, que reflete um discurso de otimismo e possível melhora do transtorno sofrido. Este resultado de certa forma corrobora a análise de Althoff et al. [2016],

que os autores concluíram que pessoas que atravessam períodos de maior dificuldade são mais propensas a pensar no futuro e serem positivas, quando auxiliadas por outras pessoas que oferecem ajuda.

3. Os *spans* dos usuários mais relevantes para o **Tópico 10**, que está relacionado com problemas de insegurança e adolescência e que aparecem de forma mais pronunciada em SuicideWatch e Depression, também exibem altas probabilidades associadas a tópicos de teor negativo (10 e 12). Isto sugere a presença de um número expressivo de adolescentes que buscam as comunidades online para relatar inseguranças e pensamentos suicidas.
4. Os *spans* de usuários mais relevantes para o **Tópico 28**, que é ligado a sugestões de atividades para amenizar uma condição emocional negativa, também possuem altas probabilidades em relação aos **Tópicos 7 e 8**, revelando um possível perfil de usuários que buscam compartilhar experiências de possíveis melhoras da saúde mental através de atividades e *hobbies*.

4.6 Análise do Tom Emocional dos Usuários

A fim de mensurar o estado emocional de um usuário e verificar se há mudanças no comportamento do mesmo (QP3), caracterizamos a distribuição do tom emocional (TE) das publicações levando-se em consideração seus tamanhos, em número de palavras. Em particular, iremos considerar o TE dos *posts*, o TE dos comentários e o TE dos comentários realizados pelos autores dos *posts*.

Primeiramente, introduzimos a seguir a notação que será utilizada nesta análise. Seja $TE(p)$ o tom emocional de um *post* p ; $TE(c)$ o tom emocional de um comentário; e $TE(c_p)$ o tom emocional de um comentário feito pelo autor de uma *thread* naquela mesma *thread*.

A Figura 4.15 apresenta um mapa de calor do $TE(p)$ para diferentes tamanhos de *post*, onde os verdes mais escuros indicam menor frequência. Observe que a distribuição obtida é semelhante para todos os subreddits. Os *posts* grandes (tamanho ≥ 1000) correspondem a 1,81% no Depression, 1,9% no SuicideWatch, 0,94% no Anxiety e 0,79% no Bipolar. Note que o Depression e o SuicideWatch possuem as publicações mais longas. Além disso, pode-se observar que esses textos possuem valores extremos do TE (próximos a -1.0 e a +1.0), o que é um indício que quando os usuários realizam essas postagens maiores, transmitem uma grande carga emocional. Por exemplo, o trecho a seguir, cujo TE é igual a -0.9995, é um *post* com 6693 caracteres da comunidade SuicideWatch:

“...I am depressed, this is the short story, the easy one. The one that makes people look at me with mingled pity and disgust. I have wanted to kill myself for the better part of five years, and the only reason I haven’t is my little brother. I think about it every day, and I can’t seem to stop myself. Even on my best days, I have intrusive thoughts...”

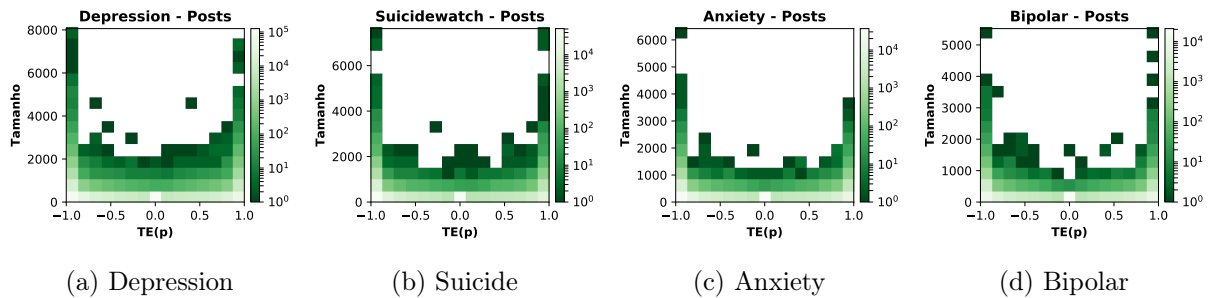


Figura 4.15: Tamanho dos *Posts* x Tom Emocional.

A Figura 4.16 apresenta o mapa de calor do tom emocional para os comentários. Novamente observamos que os *posts* são mais longos, em média, do que os comentários. É possível observar que o subreddit SuicideWatch possui maior volume de comentários longos. Além disso, os comentários grandes (tamanho ≥ 1000), correspondem a 0,05% no Depression, 0,06% no SuicideWatch, 0,05% no Anxiety e 0,04% no Bipolar.

Apesar de existirem comentários e *posts* com mais de 1000 caracteres, em uma análise minuciosa, verifica-se que pelo menos metade desses, têm aproximadamente apenas 100 caracteres nos *posts* e menos de 50 nos comentários. Ao analisar a mediana dessas comunidades é possível justificar essas afirmações: o tamanho dos *posts* no subreddit Depression é 100, SuicideWatch 105, Anxiety 105 e Bipolar 79. E dos comentários no subreddit Depression é 29, SuicideWatch 27, Anxiety 40 e Bipolar 33.

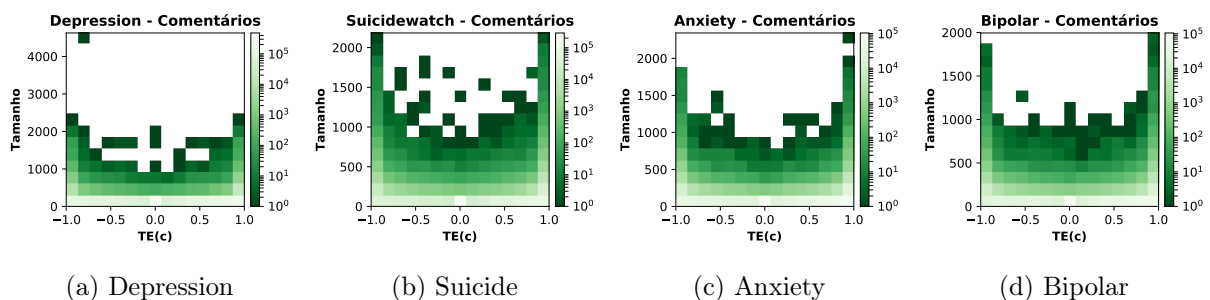


Figura 4.16: Tamanho dos Comentários x Tom Emocional.

A Figura 4.17 apresenta o mapa de calor do TE dos comentários feitos pelos autores de uma *thread* naquela própria *thread*. Ao comparar esta figura com a Figura 4.16, pode-se concluir que os maiores comentários não são de autoria dos usuários que escreveram o *post* inicial. Assim, observamos que usuários que não realizaram *posts* também escrevem grandes comentários, o que é mais um indício da existência de suporte social.

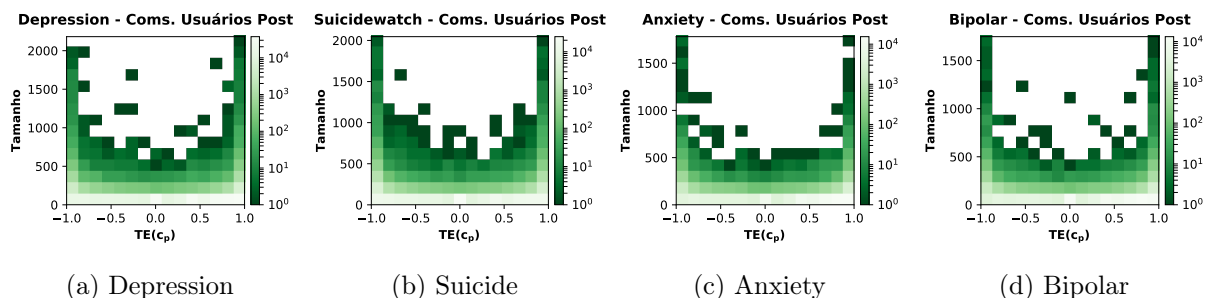


Figura 4.17: Tamanho dos Comentários dos Usuários que realizaram o post x Tom Emocional.

4.6.1 Tom Emocional e Clusters

A seguir, analisamos os grupos para verificar se existe diferença do TE. A Figura 4.18 mostra o resultado. Para investigar se a diferença do TE entre os grupos é significativa, aplicamos o teste estatístico de análise de variância, também conhecido como ANOVA [Miller Jr, 1997].

Primeiramente definimos as hipóteses (H_0 e H_1). Sejam H_0 : as amostras têm a mesma distribuição; e H_1 : as amostras não têm a mesma distribuição. Em seguida, definimos o nível de significância do teste, $\alpha = 0,05$. Depois aplicamos a padronização das variáveis e executamos o ANOVA para calcular o p-valor. O p-valor é utilizado para rejeitar ou não H_0 . Se $p \leq \alpha$, H_0 é rejeitada, as distribuições são diferentes; se $p > \alpha$, não é possível rejeitar H_0 com base nos dados.

Ao analisarmos os resultados, o valor de p foi igual a 1 em todas as comunidades. Com isso, H_0 não pode ser rejeitada, e verifica-se que o TE nos grupos não é estatisticamente relevante. Por isto, não aplicamos esses grupos na tarefa de entender a diferença do tom emocional.

4.6.2 Variação do Tom Emocional nas Árvore de Discussão

Para estudar a variação do tom emocional de um usuário nas árvores de discussão, selecionamos as árvores criadas por ele que correspondem a períodos em que o usuário

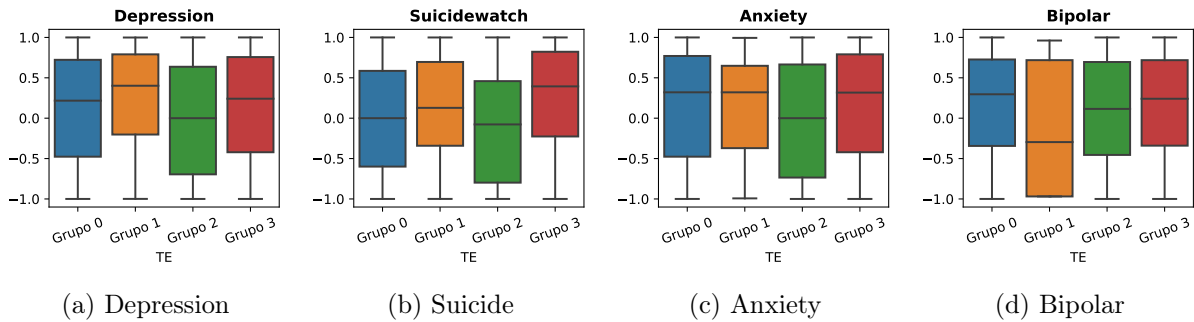


Figura 4.18: Tom Emocional x Clusters.

não realiza atividades concomitantes em outras árvores. Com isso pretendemos isolar os acontecimentos externos a *thread* que está sendo analisada. Após esta seleção, realizamos a análise do TE. Além disso, consideramos apenas sequências de comentários em (i) que exista pelo menos uma interação com outro usuário e (ii) que o intervalo entre duas interações consecutivas seja menor ou igual a 24 horas. A escolha do intervalo de 24 horas é baseada na distribuição da diferença entre duas atividades apresentada na Figura 4.7.

A Figura 4.19 ilustra a forma como as árvores de discussão foram selecionadas. O eixo horizontal corresponde ao tempo. O usuário em questão realiza atividades durante o intervalo ilustrado. As atividades realizadas por este usuário são marcadas por ticks mais grossos, identificados por P no caso de *post*, ou C_i no caso do i -ésimo comentário que ele faz na *thread*. As atividades dos outros usuários são marcadas pelos ticks mais finos. No exemplo, a *thread* 4 não foi considerada, porque antes de escrever o comentário c_2 na *thread* 2, o usuário interagiu em outras *threads*, mais especificamente, na *thread* 4. Note que a ordem em que as *threads* são avaliadas impacta no conjunto final de árvores de discussão. Neste trabalho, as *threads* são ordenadas de forma crescente segundo a data de criação do *post*.

Após a seleção das árvores de discussões, com as características descritas anteriormente, permanecemos com 87.906 árvores no Depression, 41.087 no SuicideWatch, 33.821 no Anxiety e 19.267 no Bipolar. Em percentuais, considerando o total das árvores com usuários não deletados, corresponde a 27,78% no Depression, 37,81% no SuicideWatch, 32,64% no Anxiety e 33,17% no Bipolar. Note que o percentual de árvores analisadas não é grande, porém são árvores em que o autor do *post* não interagiu em outras discussões, eliminando potenciais fatores de confusão. A seguir iremos explorar as características, relacionadas ao tom emocional, desse subconjunto de árvores.

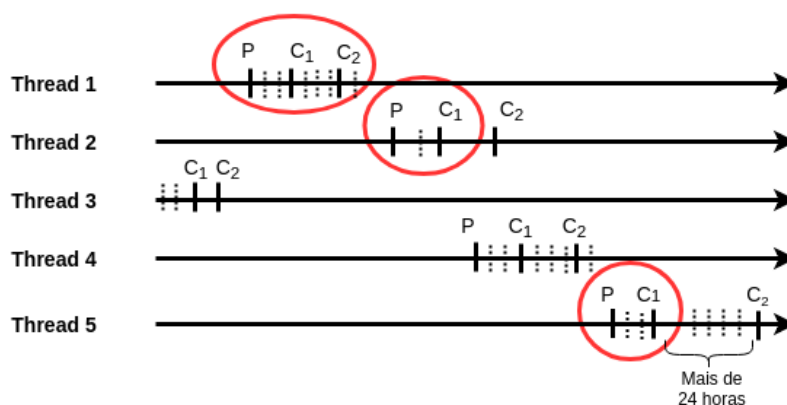


Figura 4.19: Apenas as árvores de discussão circuladas foram consideradas. O usuário *A* iniciou as *threads* 1, 2, 4 e 5. Não consideramos a sequência 4, pois o usuário ainda tinha interação na *thread* 2. A variável *p* corresponde a um *post*, As variáveis *c* são os comentários de quem realizou o *post*. Os pontilhados são os comentários de outros usuários.

Características TE Árvore de Discussão

A Figura 4.20 apresenta a distribuição do TE do *post*, do último comentário do autor do *post* e a média de TE dos comentários da árvore de discussão, sem incluir o autor do *post*. O gráfico apresentado é do tipo violino, semelhante ao boxplot, mas cuja diferença é a possibilidade de se analisar a densidade dos valores: quanto mais largo o formato, maior é a concentração de pontos naquela região. O ponto branco indica a mediana. A barra mais grossa é o intervalo interquartil.

É possível observar que, em todos subreddits, o 1º quartil e a mediana referentes ao *post* estão bem abaixo das respectivas medianas dos comentários dos últimos comentários do autor e dos comentários dos outros usuários. Além disso o 3º quartil referente ao *post* também se encontra abaixo daquele referente à média dos comentários, exceto no subreddit Bipolar. Em geral, os comentários tendem a ser mais positivos que os *posts*, exceto para o subreddit Bipolar, que apresenta também um grande volume de *posts* com o tom emocional mais positivo. Note que, em geral, há maior concentração de *posts* na região de TE negativo. Por outro lado, no caso do último comentário feito pelo autor da *thread*, há maior concentração na região de TE positivo.

Observe que a mediana do TE dos usuários que realizaram comentários nas árvores de discussão (excluindo o autor do *post*) é superior a mediana do TE dos *posts*. O 1º, 2º e 3ºs quartis referentes ao último comentário estão acima daqueles referentes ao TE do *post*, o que sugere uma melhora no tom emocional do usuário autor do *post*.

Além disso, a mediana do TE dos comentários ($TE_{c_{all}}$) é maior que o TE do último comentário do autor do *post*, exceto no subreddit SuicideWatch. Esta análise sugere

que as interações do autor de um *post* com outros usuários, ao receber comentários com TE positivo, podem contribuir para o aumento do seu TE.

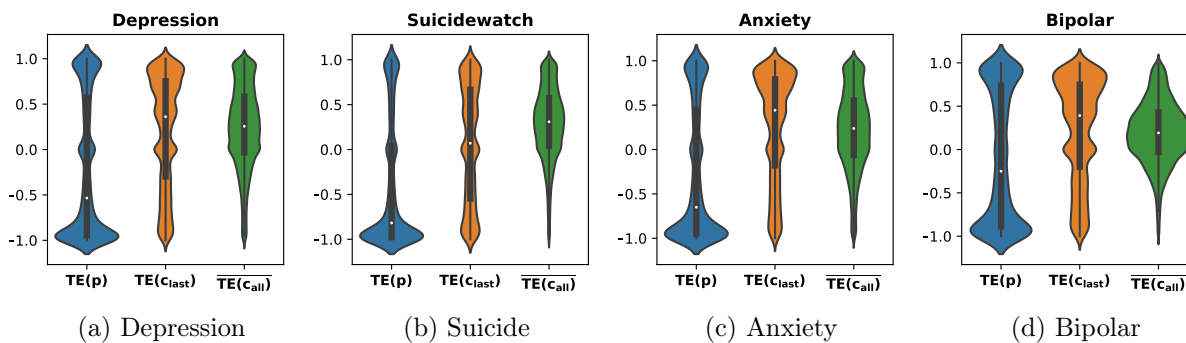


Figura 4.20: Tom Emocional.

Relação TE do Post com os Comentários

Para entender como os usuários que participam de uma árvore de discussão interagem com o autor do *post*, apresentamos a Figura 4.21, que mostra um mapa de calor do número de árvores de discussão, agrupadas de acordo com o TE do *post* (eixo x) e com o TE médio dos comentários da árvore de discussão, excluindo-se o *post* inicial. É possível observar que quando o *post* tem TE extremamente negativo (-1), a média dos comentários dos outros usuários na árvore de discussão tende a ser mais positiva (valores acima 0). Diferentemente dos demais subreddits, no Bipolar há também uma grande concentração de *posts* muito positivos (+1), que é uma característica do transtorno bipolar.

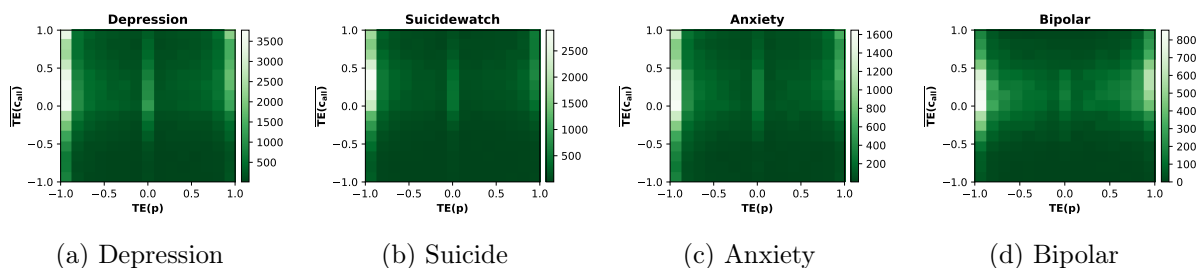


Figura 4.21: Tom Emocional dos Posts x Tom Emocional Médio dos Comentários (exceto dono do Post).

Relação TE do Post com o Último Comentário

Averiguamos também se ocorre mudança do TE entre o momento em que um usuário escreve um *post* e o momento em que faz seu último comentário na árvore de discussão, utilizando um mapa de calor semelhante ao anterior, mostrado na Figura 4.22. Note que, autores de *posts* extremamente negativos (-1) tendem a escrever comentários mais positivos ao final da *thread*. Esta variação sugere que os comentários realizados por outros usuários podem ajudar os autores que se encontram em situações difíceis.

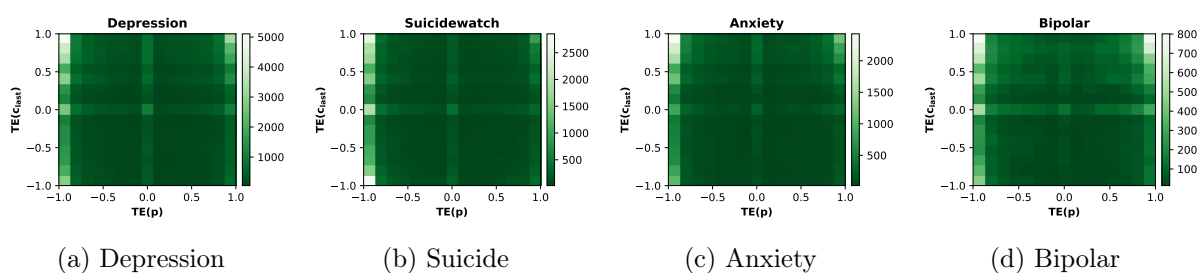


Figura 4.22: Tom Emocional dos Posts x Tom Emocional do Último Comentário.

Relação TE do Último Comentário com os demais Comentários

O passo seguinte foi investigar se o TE do último comentário do autor tem relação com a média do TE dos comentários da árvore de discussão (excluindo-se o *post* inicial), calculando-se o mapa de calor mostrado na Figura 4.23. Note que, em todos subreddits, um TE positivo no último comentário do autor do *post* está relacionado a um TE positivo médio entre os comentários *thread*. No SuicideWatch também ocorre uma grande quantidade de comentários negativos por parte dos autores da *thread*, mesmo quando estão relacionados a uma árvore de discussão com TE positivo. Isto pode ser um indício de que existem usuários nesta comunidade que são mais difíceis de melhorar seu tom emocional.

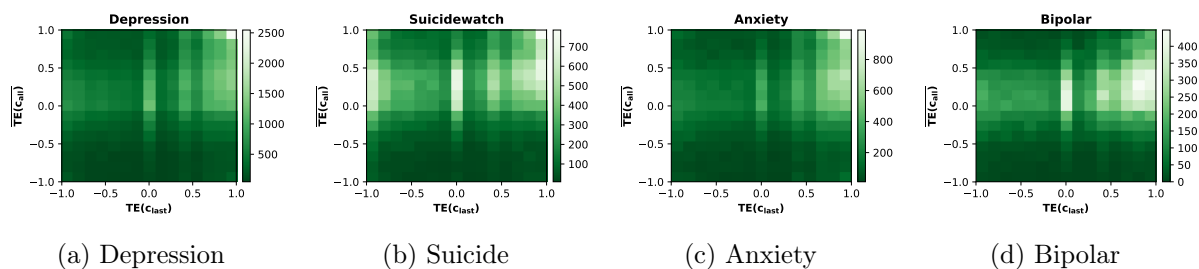


Figura 4.23: Média do Tom Emocional dos Comentários x Tom Emocional do último comentário

Para melhor entender a relação entre o TE médio dos comentários e o TE do último comentário do autor, agregamos os resultados anteriores quantizando $TE(c_{last})$ em *bins*. Estes resultados são mostrados na Figura 4.24. Como pode ser observado, existe uma correlação positiva entre essas duas variáveis: quanto maior a média do TE da árvore de discussão, maior é o TE do último comentário.

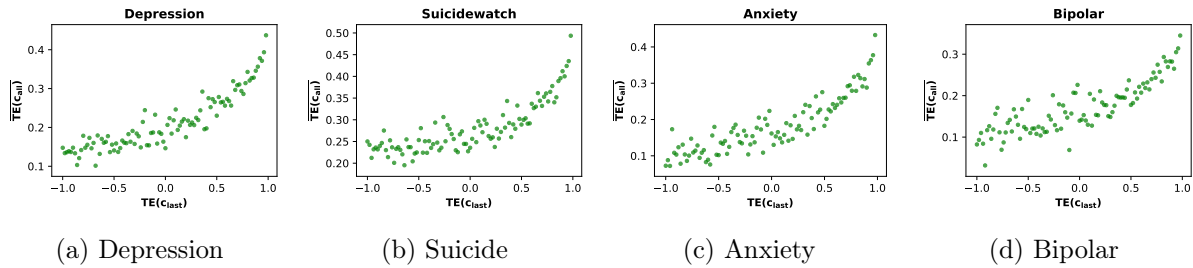


Figura 4.24: TE último comentário x Média dos comentários nas árvores de discussão

Relação TE dos Comentários com Diferença do TE do último comentário e o Post

Como observado anteriormente, o TE dos comentários das árvores de discussões estão positivamente correlacionados com o TE do último comentário. Conseqüentemente, podem estar relacionados com uma variação positiva, considerando-se a diferença entre o TE do *post* e o TE do último comentário.

Para investigar melhor estas relações, analisamos a diferença, para cada árvore de discussão, entre o tom emocional do último comentário do autor e do *post*. A Figura 4.25 apresenta a distribuição desta diferença. Observe que ocorre maior concentração nos valores acima de zero, mostrando que o usuário, na maioria das vezes, “melhora” ao final da *thread*, em relação ao TE.

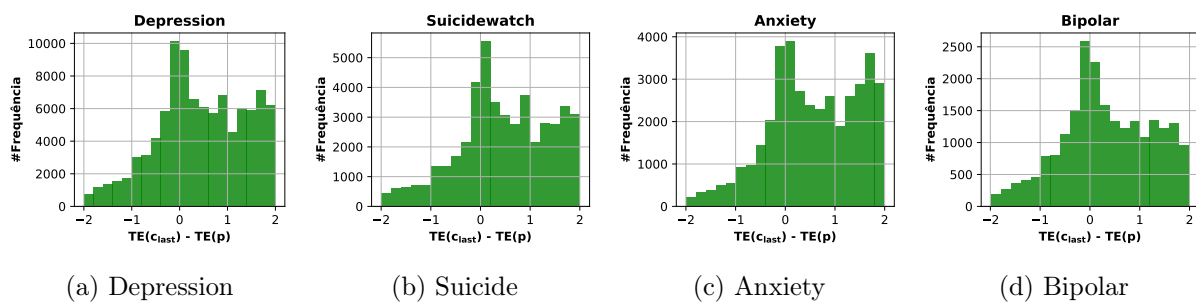


Figura 4.25: Histograma da Diferença do Tom Emocional entre comentário e post

Analizamos então a relação entre a média do TE dos comentários e a variação do TE observada pelo autor do *post*. Na Figura 4.26, notamos que quando a média é positiva, existe uma diferença positiva entre o TE do último comentário e o *post*, a qual que pode ser observada no canto superior direito. Note que quando a média é negativa, não existe um padrão claro para a diferença do tom emocional.

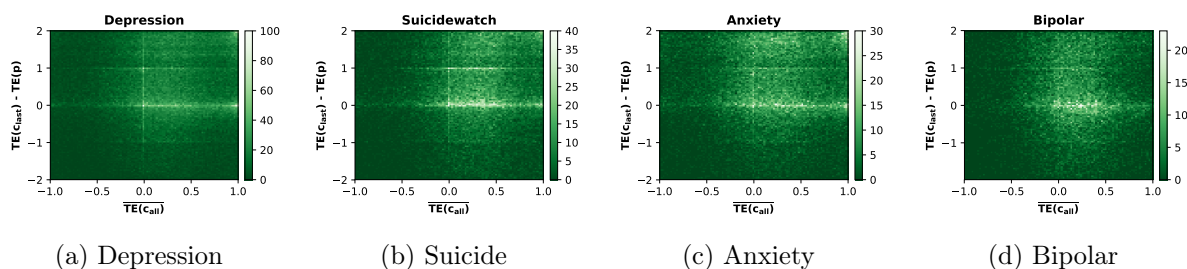


Figura 4.26: Média do Tom Emocional dos Comentários x Diferença do tom emocional entre último comentário e *post*.

Discussão

As análises realizadas fornecem um conjunto de observações relevantes a respeito da dinâmica do tom emocional de um usuário ao participar em uma árvore de discussão. Investigamos vários aspectos do TE dentro da árvore de discussão: nos *posts*, nos comentários, no último comentário do autor do *post* e na diferença do TE. Um dos resultados mais importantes é a existência da correlação entre o último comentário do usuário que realizou a postagem com os comentários dos demais usuários na árvore de discussão. Além disso, analisamos que um comentário influencia o TE do próximo comentário, um indício de que o aspecto temporal pode impactar na evolução do comportamento do usuário dentro da comunidade. Assim, pretendemos criar um modelo para entender a diferença do TE do usuário ao longo tempo, dentro da árvore de discussão.

Capítulo 5

Modelos para Previsão do Tom Emocional

Este capítulo apresenta os modelos construídos para a previsão da variação do tom emocional de um usuário entre o momento em que criou o *post* e um momento posterior em que faz um comentário na mesma árvore de discussão. A seguir, definiremos formalmente a tarefa (Seção 5.1), os atributos utilizados na modelagem (Seção 5.2) e os modelos abordados para prever o impacto de uma sequência de comentários sobre o autor da árvore de discussão (Seção 5.3).

5.1 Tarefa

Seja o *post* inicial p , com tom emocional e_p realizado por um usuário u e um conjunto de n comentários (c_1, c_2, \dots, c_n) , onde o comentário c_n , com tom emocional e_{c_n} , é o último comentário realizado pelo usuário u . Iremos prever a variação do tom emocional, $\Delta(e_{c_n} - e_p)$, do usuário u na árvore de discussão. Diante desta tarefa, nosso objetivo é verificar se houve alguma evolução do sentimento do usuário dentro da comunidade a partir da árvore de discussão (*posts* e comentários). A Figura 5.1 apresenta um resumo desta tarefa. Os *posts* e comentários serão atributos de entrada para modelos que preveem a diferença do tom emocional (TE).

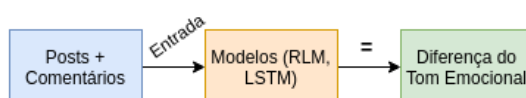


Figura 5.1: Resumo da Tarefa Desenvolvida

Embora prever a diferença do TE e prever o TE do último comentário sejam tarefas equivalentes do ponto de vista técnico, argumentamos que prever melhorias ou pioras faz mais sentido. Inclusive, considerando o aspecto social, se o usuário melhora ao longo das discussões isto pode indicar que as interações estabelecidas nestas comunidades contribuíram de forma positiva para sua saúde mental.

A tarefa escolhida possui limitações. Não é possível avaliar qualitativamente se houve realmente melhora, devido a própria característica da rede, que permite anonimidade e não possui algum recurso explícito para controlar como o usuário está se sentindo realmente. Como consequência, as métricas utilizadas não consideram aspectos externos, mais complexos da vida do usuário que poderiam nos guiar nesta avaliação.

Para desenvolver esta tarefa propomos diversos modelos. Como observamos, na etapa de caracterização, as comunidades são muito semelhantes entre si. Por isso, decidimos executar esses modelos, em fase exploratória, na comunidade Bipolar. Após a escolha do melhor modelo, o replicamos para as demais comunidades. O melhor modelo foi definido como aquele com menor MSE (*Mean Squared Error*).

5.2 Atributos

Para desenvolver a tarefa proposta, é preciso representar *posts* e comentários através de atributos. Considerando isso, definimos duas formas para capturá-los: análise textual baseada em atributos psicolinguísticas, obtidos a partir do LIWC e a análise semântica textual baseado em atributos latentes, que utilizam *Word Embeddings*. Essas representações serão utilizadas como entrada para os modelos que serão construídos.

5.2.1 Análise Textual com Atributos Psicolinguísticas

Esta análise utiliza os atributos da ferramenta LIWC 2015 (Linguistic Inquiry and Word Count). O LIWC é um recurso de análise textual, desenvolvido para extração de atributos psicolinguísticos de um texto para vários idiomas.

Com base em um dicionário, o LIWC recebe um texto de entrada e calcula um valor para cada uma de 93 categorias. Entre as características textuais extraídas estão: quantidade de palavras, dimensões linguísticas (quantidade de pronomes, de artigos, preposições, verbos auxiliares, negações, conjunções etc.); processos psicológicos (emoção negativa e positiva, ansiedade, tristeza, raiva, família, amigos, saúde, trabalho, casa, dinheiro, religião, orientação temporal, etc.) Assim, para cada um dos *posts* e

comentários obtivemos atributos que representam diversos aspectos do texto e que são utilizados para entender a variação do TE do usuário dentro da árvore de discussão.

Além dos atributos do LIWC, criamos outros dois: total de comentários e média do TE dos comentários da árvore de discussão, excluindo as atividades do autor que fez o *post*. O objetivo é entender se o tamanho da *thread* e sua média de TE tem efeito estatisticamente relevante em um modelo de regressão.

5.2.2 Análise baseada em Word Embeddings

Esta análise utiliza *Word Embeddings* (WE), também conhecida como *words vectors*, *word representations*, que representam palavras como vetores em um espaço latente. Nesse espaço, vetores similares representam palavras com o significado ou função semelhantes.

Dentre as diversas estratégias de criação de WE, optamos por utilizar o Word2Vec [Mikolov et al., 2013] que é um método estatístico para aprender a representação das palavras dentro de um corpus. Este método define dois modelos probabilísticos alternativos para se calcular a verossimilhança de uma sentença: *Continuous Bag-of-Words*, (CBOW) e *Continuous Skip-Gram*. Antes de definí-los, é preciso entender o conceito de contexto utilizado por esses modelos. Contexto são palavras que estão ao redor da palavra alvo. O CBOW aprende a representação da seguinte forma: dado um contexto, o modelo prevê a próxima palavra. O SkipGram, por sua vez, aprende prevendo o contexto dado uma palavra. Por exemplo, na frase: “Eu amo pizza”, sendo “pizza” a palavra alvo, o CBOW utiliza o contexto “Eu amo” para prever a palavra “pizza”. O SkipGram funciona de maneira oposta. Como no corpus existem diversas palavras e diversos contextos, o modelo cria as representações maximizando uma versão modificada da função de verossimilhança. Note que palavras que aparecem em um contexto semelhante, terão um vetor semelhante.

Para a criação dos *words-embeddings* usamos todos os *posts* e comentários de 2010 a 2017 de cada subreddit. No pré-processamento, utilizamos a ferramenta ekphrasis¹, a qual realiza normalização das palavras e correção ortográfica. A correção das palavras utiliza estatísticas de palavras de dois grandes corpus (Wikipédia em inglês, Twitter em inglês). No nosso contexto utilizamos a correção pelo Twitter por também ser uma rede social. Para a normalização, que consiste em substituir determinado conteúdo por uma marcação de texto, escolhemos atributos específicos, que não contribuíram para a representação. Por exemplo, e-mails, telefones, horas, etc. são atributos, na maior parte das vezes, distintivos e saber o e-mail de usuário ou algum outro dado não

¹<https://github.com/cbaziotis/ekphrasis>

agregaria nas representações. Para este caso, basta saber se o usuário informou um e-mail, assim o representamos de forma mais geral. Desta maneira, o e-mail do usuário é substituído por uma etiqueta informando que aquele texto é um e-mail. Também removemos as pontuações, caracteres especiais e excesso de espaço. Para remover as *stopwords* utilizamos *Natural Language Toolkit* (NLTK) [Bird et al., 2009].

Na geração dos *Words Embeddings* (WE) utilizamos o pacote Gensim do Python. Os parâmetros para execução do modelo foram: arquitetura *SkipGram* [Goldberg, 2017]; contexto igual à 5; considerando todas as palavras (que tenham aparecido ao menos uma vez no corpus); dimensão dos vetores é 300. Nos experimentos, também utilizamos *embeddings* de dimensão 100, mas os de dimensões 300 tiveram melhor resultado.

O WE retorna um vetor para cada palavra do *post*/comentário. Para agregar os *embeddings* associados a cada publicação, tomamos a média dos *embeddings* das palavras. Portanto, cada publicação é representada por um vetor composto pela média dos *embeddings* de todas as palavras.

A vantagem do LIWC é que os atributos são interpretáveis. Por outro lado, os atributos do Word2Vec podem carregar mais informações. Na representação WE palavras como: “bad”, “terrible”, “horrible” terão representações semelhantes. No LIWC essas palavras compactam as métricas relacionadas as categorias: processo psicológico e dimensões linguísticas padrão.

5.3 Modelos de Previsão

Nesta dissertação, consideramos dois modelos para prever a variação do tom emocional do criador da árvore da discussão: uma regressão linear multivariada (RLM) e redes neurais. Pretendemos explorar um modelo mais simples, onde o aspecto temporal não é considerado e os atributos utilizados são interpretáveis, mas não trazem informação contextual. E outro, em que aspectos semânticos e temporais serão observados.

5.3.1 Regressão com Atributos do LIWC

A RLM é uma técnica estatística clássica que utiliza duas ou mais variáveis explicativas (variáveis independentes) para prever o resultado de uma variável resposta (variável dependente) [Myers & Myers, 1990]. Nosso objetivo é modelar a relação entre as variáveis independentes e a variável dependente, que é a variação do tom emocional do autor da árvore de discussão.

Primeiramente, definimos uma regressão *baseline* com os atributos: total de comentários e média do TE da árvore de discussão, excluindo quem fez o *post* nos dois atributos. Em seguida, aplicamos uma RLM com essas variáveis e todos os atributos do LIWC. A fórmula da regressão linear múltipla é apresentada na equação 5.1:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_mx_m + \epsilon \quad (5.1)$$

onde cada (x,y) é uma árvore de discussão.

- y = variação do TE (i.e., $e_{c_n} - e_p$).
- x_j = atributos j do LIWC e/ou criados. Onde $j = 1, \dots, m$
- β_0 = intercepto.
- β_j = coeficientes dos atributos das variáveis independentes. Representam a variação da diferença do TE em função do atributo j .
- ϵ = erro do modelo (resíduo).

Utilizamos a RLM com as 93 variáveis do LIWC mais os dois atributos criados (total dos comentários e média do TE da árvore de discussão, excluindo quem fez o *post*). Como são muitas variáveis, aplicamos a regressão LASSO em uma RLM, que é uma técnica simples de regularização L1 para reduzir a complexidade do modelo, gerando modelos esparsos. Nesta abordagem, é possível obter um subconjunto de variáveis independentes que minimiza o erro da previsão da variável dependente. O LASSO faz isso impondo uma penalidade na magnitude nos parâmetros do modelo, que faz com que os coeficientes de regressão para algumas variáveis diminuam em direção a zero. Variáveis com um coeficiente de regressão igual a zero após o processo de contração são excluídas do modelo. Variáveis com coeficientes de regressão diferentes de zero são mais fortemente associadas à variável de resposta. A análise desta regressão ajuda a determinar quais dos preditores são mais importantes, reduzindo o número de parâmetros [Tibshirani, 1996].

5.3.2 Redes Neurais

O objetivo desta abordagem é verificar se a aplicação de atributos que possuem mais informações reduzem o erro do modelo. Utilizamos duas modelagens: a primeira possui as informações e não considera a ordem das publicações. Na segunda, além das informações, levamos em consideração o aspecto temporal (a ordem em que os comentários foram realizados). Assim, pretende-se separar o momento em que usuários

pedem ajuda (em uma possível postagem) e o momento que são ajudados, através de comentários de outros usuários. Como vimos na Seção 4.5.2, existem usuários que buscam dar conselhos e motivar a luta contra problemas de saúde mental. Desta maneira, queremos analisar se houve evolução, preferencialmente positiva, do TE do usuário a partir das interações com os demais participantes das comunidades, ou seja, a partir aos comentários recebidos.

Multilayer Perceptron (MLP) O MLP [Gardner & Dorling, 1998] é uma rede neural simples, que consiste em, pelo menos, três camadas de neurônios/nós, sendo uma camada de entrada, uma oculta e outra de saída. Os neurônios de entrada são os próprios atributos de cada árvore de discussão. Os neurônios da camada oculta e de saída utilizam uma função de ativação não linear. Para treinar o modelo, utiliza-se a técnica de *backpropagation* [Rumelhart et al., 1985]. Neste modelo, ainda não consideramos o aspecto temporal.

Redes Neurais Recorrentes (RNN's) Para capturar o aspecto temporal, iremos modelar a dependência de sequências com o uso de RNN's, mostrada na Figura 5.2. Neste tipo de arquitetura, a rede é formada por uma sequência de unidades RNN, que podem ser de vários tipos, incluindo vanilla RNN, GRU e LSTM. Seja a rede A onde cada unidade recebe uma entrada x_t e produz um valor de saída h_t . Estes valores são transferidos ao longo da rede, através de ligações, que permitem levar a informação, a chamada “memória” da RNN. Essas redes estão relacionadas como sequências, onde a saída é entrada para outra unidade. Neste trabalho, as sequências que formam a entrada são os comentários feitos pelos usuários.

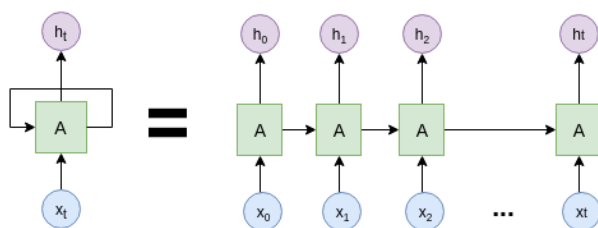


Figura 5.2: Arquitetura RNN

Um ponto negativo desta RNN vanilla é o fato de não ser capaz de levar em consideração dependências de longo prazo, devido ao problema do desaparecimento do gradiente [Hochreiter, 1998]. Para solucionar este problema, foram criadas as redes *Long Short-Term Memory* ou LSTM, um tipo especial do RNN, capazes de aprender dependências de longo prazo.

Long Short-Term Memory (LSTM) Uma unidade de LSTM (ou célula) possui 5 componentes essenciais que permitem modelar tanto dependências de longo prazo quanto de curto prazo [Gers et al., 1999]. Estes componentes são: *Cell state* (c_t); *Hidden state* (h_t); *Input gate* (i_t); *Forget gate* (f_t); *Output gate* (o_t).

A primeira etapa em uma LSTM é decidir quais informações do estado anterior h_{t-1} serão descartadas. Essa decisão é realizada pelo *Forget gate* $f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f)$, onde σ é a função sigmoide calculada elemento a elemento, $[h_{t-1}, x_t]$ é o vetor coluna formado pela concatenação de h_{t-1} e x_t , e X é a multiplicação de matriz por vetor. As entradas h_{t-1} e x_t são analisadas e gera-se um número entre 0 e 1, onde 0 mantém nenhuma informação e 1 mantém na íntegra. Neste trabalho, a primeira etapa do modelo seleciona os trechos de um comentário que serão relevantes para enviar na próxima unidade de LSTM.

A próxima etapa é decidir quais informações novas serão armazenadas na célula. Para isso utilizamos a porta de entrada $i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i)$ que através de uma sigmoide decide quais informações serão atualizadas. A seguir, uma função tanh cria um vetor de novos valores candidatos $\tilde{c}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c)$. Esta etapa, em nosso problema, é a combinação das palavras dos comentários que serão mantidas e atualizados.

Em seguida o antigo estado da célula c_{t-1} para o novo c_t . Com a definição das informações que serão esquecidas (f_t) e das que serão mantidas (\tilde{c}_t e i_t) realiza-se uma operação de atualização: $c_t = f_t * c_{t-1} + i_t * \tilde{c}_t$. Em nosso modelo, esta operação é a junção das informações relevantes do comentário anterior com o atual.

Finalmente, definimos a saída para a próxima unidade. O modelo é $o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o)$, onde decidimos quais partes da célula serão a saída. Depois, o estado oculto é atualizado para que sejam enviadas apenas as informações selecionadas $h_t = o_t * \tanh(c_t)$. Em nosso modelo, esta operação é o que será enviado do texto para a próxima célula.

A Figura 5.3 apresenta como os componentes listados anteriormente estão dispostos em uma célula LSTM. Observe que os “gates” são ligados por operações matemáticas (multiplicação e soma).

Além desta LSTM Padrão, também executamos outras versões: LSTM bidirecional (soma), LSTM bidirecional (concatenação) e LSTM Dividida. A LSTM bidirecional, é uma extensão da LSTM. A principal ideia desta abordagem é que, ao invés de treinar apenas uma camada de entrada, ela treina duas. A primeira camada é como apresentamos na LSTM padrão, a segunda tem uma cópia da entrada e a treina de forma invertida em relação a sequência da entrada [Schuster & Paliwal, 1997]. A soma e a concatenação estão relacionadas a forma de juntar as saídas de uma unidade LSTM.

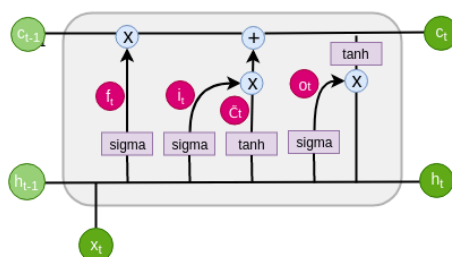


Figura 5.3: Unidade LSTM

Na soma, as saídas são adicionadas formando uma entrada de mesmo tamanho. Na concatenação, as saídas são concatenadas juntas, fornecendo o dobro do número de saídas para a próxima camada.

A LSTM dividida recebe os *posts* e os comentários em dimensões diferentes. O *post* entra na rede em uma camada densa (*Multilayer Perceptron*). Para considerar a ordem, os comentários são enviados para uma LSTM padrão. Depois de processados, juntamos as duas dimensões para retornarem a saída do modelo.

A LSTM e suas variações podem ter problemas de *overfitting*, que ocorrem quando o modelo não consegue generalizar o suficiente. Essa falha de generalização é percebida quando o resultado da métrica alvo é muito melhor na base treino do que é na base de teste [Hawkins, 2004]. O objetivo de reduzir o *overfitting* é que o modelo consiga generalizar melhor para instâncias novas.

Para minimizar os efeitos, citados anteriormente, utilizamos as técnicas a seguir: regularização utilizando *Dropout* [Srivastava et al., 2014] e o “Early stopping” [Prechelt, 1998]. O *Dropout* remove probabilisticamente unidades de LSTM durante a atualização dos pesos no treinamento da rede. Com o “Early stopping” [Prechelt, 1998], durante o treinamento avaliamos se modelo estava melhorando com o passar das épocas. Se durante dez épocas consecutivas o modelo piora, ou seja, o erro aumenta, nós interrompemos o treino, assumindo que, a partir deste ponto, o modelo não está mais aprendendo de forma “efetiva”.

A fim de validarmos a precisão dos modelos propostos, aplicamos a validação cruzada [Picard & Cook, 1984], especificamente o método *k-folds* [Bengio & Grandvalet, 2004]. Este algoritmo consiste em dividir os dados em k grupos (folds). Depois, aleatoriamente, escolhe-se um grupo para ser o grupo teste e os demais $k-1$ grupos serão utilizados no treinamento. Este método se repete até que todos os grupos tenham passado pelo grupo de teste. No final, calculamos o erro médio que cada modelo obteve. Neste trabalho, todos os experimentos são executados com 5 folds.

Capítulo 6

Análise Experimental

Este capítulo apresenta as análises experimentais dos modelos de previsão da variação do tom emocional do usuário que criou o *post*. Primeiramente, definimos três baselines simples (Seção 6.1). Em seguida, apresentamos os resultados dos modelos: RLM com regularização Lasso (Seção 6.2), *Multilayer Perceptron* (Seção 6.3) e LSTM (Seção 6.4). Por último, na Seção 6.5, discutimos os resultados dos diferentes modelos para a tarefa proposta.

6.1 Baselines

Definidos três baselines com o objetivo de obter medidas de referência em relação aos erros de previsão, necessárias para determinar o quão efetivos são os modelos propostos. Seja D a variável aleatória que representa a variação do TE sofrida pelo autor de uma árvore de discussão escolhida uniformemente ao acaso, definida por

$$D = e_{c_n} - e_p,$$

isto é, pela diferença entre e_{c_n} , o TE de seu último comentário, e e_p , o TE do *post* que iniciou a *thread*. Sejam ainda D_i a variação do TE associada à *thread* i e \hat{D}_i a respectiva previsão, retornada por um dado modelo. Considerando o *Mean Squared Error* (MSE) como medida de erro, podemos avaliar os modelos utilizando a equação

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (D_i - \hat{D}_i)^2, \quad (6.1)$$

onde as variáveis D_1, \dots, D_n representam a variação de TE associada às *threads* do conjunto de teste.

Baseline 1 (B_1): O primeiro baseline é o modelo que sempre retorna 0 para a variação do TE, ou seja, que assume que o usuário manteve seu TE na árvore de discussão. Neste caso, o MSE do Baseline 1 é dado por

$$\text{MSE}(B_1) = E[D^2].$$

Baseline 2 (B_2): O segundo baseline é o modelo que sempre retorna a variação média do TE, $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$. Neste caso, o erro de previsão é calculado por

$$\text{MSE}(B_2) = E[(D - \bar{D})^2].$$

Baseline 3 (B_3): O terceiro baseline utilizado foi uma regressão linear considerando como atributos o total de comentários e a média do TE da árvore de discussão (excluindo-se àqueles de quem fez o *post*) para prever a diferença do TE. O erro correspondente é dado pela Eq. (6.1).

A Tabela 6.1 apresenta os coeficientes estimados pelo modelo de regressão (B_3). Todos os coeficientes são estatisticamente relevantes. O coeficiente associado ao total de comentários é negativo, indicando que um número menor de comentários está correlacionado com uma variação positiva do TE. Uma possível interpretação para este fato é que, quando o usuário responde aos comentários e demonstra variação positiva, a árvore normalmente termina mais rápido.

Observa-se que o coeficiente associado à média dos comentários também é estatisticamente relevante para todas comunidades e, em especial, para Depression e SuicideWatch, nas quais apresentou maior significância com relação ao p-valor. Possíveis explicações para a particularidade destas duas comunidades é que usuários enfrentando problemas mais severos são mais suscetíveis a serem influenciados ou que os membros destas comunidades se empenhem mais em ajudar aqueles que estão passando por problemas. Além disso, apenas na comunidade Bipolar o coeficiente da média é negativo. Para este caso a diferença do TE desses usuários tende a ser inversamente proporcional à média do TE dos comentários dos usuários. Essa correlação negativa pode ser inerente à bipolaridade, mas precisa ser investigada com detalhe em trabalhos futuros.

	Intercept	Total Comentários	Média TE
Depression	0,4228 ***	-0,0014 ***	0,0619 ***
SuicideWatch	0,4354***	-0,0050 ***	0,1457 ***
Anxiety	0,5827 ***	-0,0065 ***	0,0221 *
Bipolar	0,3864 ***	-0,0062 ***	-0,0515 **

Tabela 6.1: Coeficientes da Regressão Linear. Cada ‘*’ equivale a significância estatística. Quanto mais ‘*’, maior é a significância, considerando o p-valor.

MSE dos Baselines: A Tabela 6.2 apresenta o MSE obtido para cada baseline avaliado em cada comunidade. Observe que os MSE's do baseline 3 são menores que os MSE's dos baselines 1 e 2. O modelo 3 agrega mais informações relevantes e utiliza-se de atributos específicos de cada árvore de discussão. Assim, o baseline 3 pode ser considerado um modelo mais “completo” e que consegue capturar melhor a complexidade da tarefa realizada.

Subreddit	Mean Squared Error (MSE)		
	Baseline 1	Baseline 2	Baseline 3
Depression	1,023	0,854	0,848
SuicideWatch	1,029	0,843	0,839
Anxiety	1,062	0,818	0,813
Bipolar	0,964	0,819	0,811

Tabela 6.2: MSE's dos Baselines para cada comunidade.

6.2 RLM com regularização

Neste modelo, incorporamos os atributos do LIWC aos utilizados no Baseline 3. Todos esses atributos foram normalizados antes de serem aplicados na RLM. Devido ao grande número de variáveis independentes adicionadas ao incorporar os atributos do LIWC, utilizamos a regularização L_1 (LASSO) para evitar *overfitting* e também para obter modelos mais esparsos, ou seja, com um número menor de parâmetros. Variamos a constante de regularização λ em $10^{-6}, 10^{-5}, \dots, 10^{-1}$. Os resultados sem regularização (i.e., $\lambda = 0$) foram incluídos por completude.

Consideramos quatro modelos, definidos a partir de diferentes conjuntos de variáveis independentes. Em todos os modelos, a variável dependente é a variação do TE do usuário que criou a árvore de discussão. As variáveis independentes de cada modelo são:

- Modelo 1: Média de cada atributo do LIWC sobre todos os comentários e o *post* que originou a árvore de discussão.
- Modelo 2: Média de cada atributo do LIWC sobre todos os comentários, exceto aqueles de quem criou o *post*.
- Modelo 3: Média de cada atributo do LIWC apenas sobre os comentários de quem criou o *post*.
- Modelo 4: Média de cada atributo do LIWC apenas sobre os comentários e o *post* de quem criou a *thread*.

A principal motivação para utilizar esses modelos é entender o impacto que o autor do *post* tem nas variáveis independentes, isto é, se ao incluirmos comentários do autor do *post*, variáveis passam a (ou deixam de) ser relevantes no cálculo da variação do TE de um usuário em uma árvore de discussão.

A Tabela 6.3 apresenta os resultados para a comunidade Bipolar. O modelo 4, que resulta no menor MSE, é mostrada em destaque. Neste caso, consideramos a média dos atributos do LIWC dos comentários do criador do *post*. Nesta tabela, a coluna “ $\neq 0$ ” representa o total de variáveis, que o modelo considera relevante para a previsão da variação do tom emocional. Em todos modelos, as quantidades de coeficientes diferente de zero são semelhantes, o que sugere que, para uma certa constante de regularização, os mesmos atributos são relevantes, independentemente do modelo considerado.

λ	Modelo 1			Modelo 2			Modelo 3			Modelo 4		
	Treino	Teste	$\neq 0$	Treino	Teste	$\neq 0$	Treino	Teste	$\neq 0$	Treino	Teste	$\neq 0$
0	0,7837	0,7898	95	0,7820	0,7894	95	0,7931	0,8000	95	0,6674	0,6954	95
10^{-6}	0,7837	0,7898	95	0,7820	0,7894	95	0,7931	0,8000	95	0,6674	0,6954	95
10^{-5}	0,7837	0,7898	94	0,7820	0,7894	94	0,7931	0,8000	94	0,6674	0,6954	94
10^{-4}	0,7837	0,7898	92	0,7819	0,7895	91	0,7930	0,8001	90	0,6674	0,6953	91
10^{-3}	0,7835	0,7903	79	0,7816	0,7899	80	0,7918	0,8006	78	0,6681	0,6954	78
10^{-2}	0,7861	0,7960	39	0,7845	0,7953	35	0,7873	0,8068	34	0,6739	0,6987	29
10^{-1}	0,8033	0,8173	0	0,8033	0,8173	0	0,7956	0,8196	0	0,7002	0,7249	2

Tabela 6.3: Resultados (MSE) do modelo RLM com regularização das quatro abordagens para o Bipolar (λ é a constante de regularização).

Para as demais comunidades consideramos o modelo 4, uma vez que o mesmo apresenta um melhor resultado. A Tabela 6.4 apresenta os resultados. O melhor resultado é obtido com a constante de regularização $\lambda = 10^{-4}$ (em negrito). Ao compararmos com o RLM sem regularização ($\lambda = 0$), o ganho de acurácia não é significativo, a diminuição é de 0,01% no Depression, 0,01% no SuicideWatch, 0% no Anxiety e 0,01% no Bipolar. Estes valores mostram que não há diferença significativa na aplicação da regularização.

Entretanto, ao compararmos com o baseline B_3 , verificamos que ocorre ganho. A diminuição do erro é de 12,5% no Depression, 10,8% no SuicideWatch, 12,6% no Anxiety e 14,3% no Bipolar. Isto é um indício que os atributos psicolinguísticos do LIWC contribuem para a previsão da variação do tom emocional do usuário.

Entre as variáveis que foram consideradas pelo melhor modelo, estão o “WC” (*word count*), que corresponde à quantidade total de palavras no texto, além das variáveis “Clout” e “Tone”, que estão relacionadas à sumarização dos resultados do LIWC. Um valor alto de “Clout” sugere que o autor está transmitindo propriedade e confiança no texto. Um valor baixo, sugere que ele é mais hesitante, humilde e até ansioso. Um valor alto para “Tone” está associado a um estilo mais positivo e otimista; um valor

λ	Depression			SuicideWatch			Anxiety			Bipolar		
	Treino	Teste	$\neq 0$	Treino	Teste	$\neq 0$	Treino	Teste	$\neq 0$	Treino	Teste	$\neq 0$
0	0,7356	0,7422	95	0,7538	0,7479	95	0,7047	0,7106	95	0,6674	0,6954	95
10^{-6}	0,7356	0,7422	95	0,7538	0,7479	95	0,7047	0,7106	95	0,6674	0,6954	95
10^{-5}	0,7356	0,7421	94	0,7538	0,7479	94	0,7047	0,7106	95	0,6674	0,6954	94
10^{-4}	0,7356	0,7421	92	0,7538	0,7478	93	0,7046	0,7106	90	0,6674	0,6953	91
10^{-3}	0,7359	0,7421	76	0,7542	0,7482	80	0,7045	0,7110	83	0,6681	0,6954	78
10^{-2}	0,7407	0,7467	27	0,7597	0,7525	26	0,7077	0,7175	46	0,6739	0,6987	29
10^{-1}	0,7700	0,7797	1	0,7915	0,7792	1	0,7477	0,7593	1	0,7002	0,7249	2

Tabela 6.4: Resultados (MSE) do modelo RLM com regularização da abordagem 4 para todas comunidades (λ é a constante de regularização).

baixo revela maior ansiedade, tristeza ou hostilidade. Valores de “Tone” em torno de 50 sugere ausência de emotividade [Pennebaker et al., 2015].

6.3 Modelo Rede Neural - Multilayer Perceptron

Neste modelo, ainda não consideramos o aspecto temporal dos *posts* e dos comentários; somente utilizamos informações latentes dos textos dos usuários através dos *embeddings*, que capturam características semânticas e sintáticas das palavras. O objetivo é entender se este modelo gera melhores resultados do que os obtidos com a RLM com LASSO. A Figura 6.1 apresenta a arquitetura deste modelo, onde as entradas são vetores, que possuem dimensões que variam com as abordagens utilizadas. Note que existem diversos perceptrons que “aprendem” a saída, através de atribuição de pesos em cada entrada.

Criamos duas redes neurais: (i) uma apenas com perceptrons (Figura 6.1) e (ii) outra com camadas de regularização (Figura 6.2). Consideramos 5 variações do conjunto de atributos de entrada para cada uma das redes apresentadas, discutidas a seguir. É importante ressaltar que em todos modelos, desconsideramos o último comentário do autor do *post* na árvore de discussão.

- Modelo 1 - Comentários (dimensão da entrada = 300): a entrada é um vetor de dimensão 300 que é o *embedding* médio de todos os comentários da árvore de discussão, i.e., a média elemento-a-elemento dos *embeddings* dos comentários.
- Modelo 2 - *Posts* e Comentários (dimensão da entrada = 300): a entrada é um vetor de dimensão 300 que é a média do *embedding* do *post* com o *embedding* médio de todos comentários da árvore de discussão.
- Modelo 3 - *Posts* e Comentários Separados (dimensão da entrada = 600): a entrada é composta por dois vetores, sendo um vetor de dimensão 300 que o

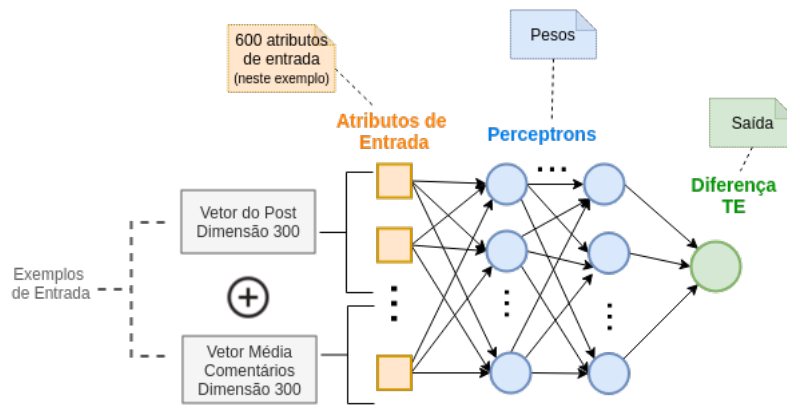


Figura 6.1: Multilayer Perceptron

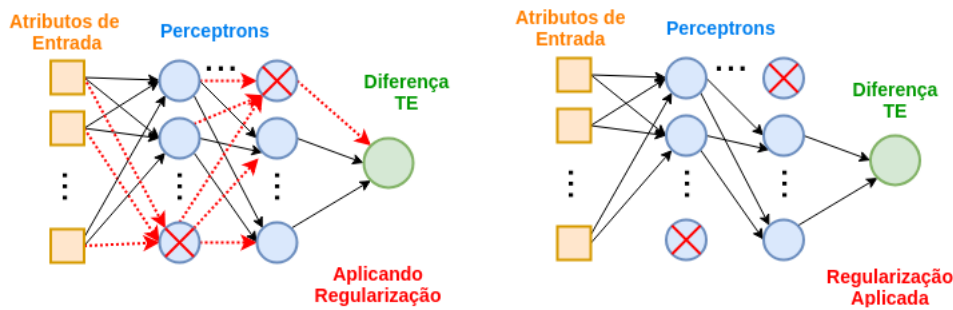


Figura 6.2: Multilayer Perceptron

embedding médio de todos comentários da árvore de discussão, e outro vetor de dimensão 300 que é o *embedding* do *post*.

- Modelo 4 - *Posts* e Comentários Separados pelo autor do *post* (dimensão da entrada = 900): a entrada é composta por três vetores, sendo um vetor de dimensão 300 que é o *embedding* médio de todos comentários (excluindo aqueles do usuário que realizou o *post*), outro vetor de dimensão 300 que é o *embedding* médio dos comentários do autor do *post*, e um vetor de dimensão 300 que é o *embedding* do *post*.
- Modelo 5 - *Posts* e Comentários Separados pelo autor do *post* (dimensão da entrada = 600): a entrada é composta por dois vetores, sendo um vetor de dimensão 300 que é o *embedding* médio de todos comentários (excluindo aqueles do usuário que realizou o *post*); outro vetor de dimensão 300 igual a média entre o *embedding* do *post* e o *embedding* médio dos comentários do autor do *post*.

A Tabela 6.5 apresenta os resultados das 10 configurações (5 modelos \times uso/não-uso de regularização) para a comunidade Bipolar. Observe que os resultados dos modelos com e sem regularização foram semelhantes. Entretanto, utilizar a regularização

é importante para evitar *overfitting*. Assim escolhemos, dentre os modelos com regularização, o modelo 4, por resultar no menor MSE dentre os demais. Note que este modelo é o que mais distingue dentre os tipos de publicação, uma vez que existe uma dimensão para os comentários dos usuários que não realizaram o *post*, outra para os comentários de quem fez o *post* e outra para o *post*. Isto mostra que separar os textos em diferentes vetores de entrada é importante para a previsão da diferença do TE.

A única diferença do modelo 5 para modelo 4 é que, no modelo 5, o vetor do *embedding* médio dos comentários do autor do *post* e o vetor dos *embedding* do seu *post* são agregados em um único vetor, que possui características textuais do mesmo autor. E o erro do modelo 5 foi 8,8% maior que o modelo 4. Apesar desses textos terem sido realizados pelo mesmo autor, é importante distinguir entre *post* e comentários do autor para melhor prever a variação do TE. Isto ocorre possivelmente devido à mudança no discurso dos comentários do autor do *post*, influenciado pelos demais usuários. Iremos aplicar o modelo 4 para as demais comunidades, tendo em vista que ele obteve o melhor desempenho.

	Sem Regularização		Com regularização	
	Media MSE	DP MSE	Media MSE	DP MSE
Modelo 1	0,8186	0,0236	0,8061	0,0192
Modelo 2	0,784	0,0182	0,7844	0,0186
Modelo 3	0,671	0,0084	0,6961	0,0446
Modelo 4	0,6781	0,0205	0,6651	0,0191
Modelo 5	0,7174	0,0121	0,7295	0,0203

Tabela 6.5: Resultados MSE das variações dos Modelos Redes Neurais Multilayer's Perceptrons para a comunidade Bipolar. Baseado em 5 folds em termos de média e desvio padrão.

A Tabela 6.6 apresenta os resultados do modelo 4 para cada uma das comunidades analisadas. A maior diferença absoluta no MSE entre comunidades é de apenas 1,01% (entre as comunidades Bipolar e Depression), mostrando que os resultados para as quatro comunidades são similares.

Em comparação com o Baseline 3, o modelo 4 obteve melhores resultados de acurácia. Na comunidade Depression, a diminuição do MSE é de 20,79%, no SuicideWatch 20,04%, no Anxiety 18,02% e no Bipolar 18,02%.

Ao compararmos com a RLM com regularização ($\lambda = 10^{-4}$), ainda obtemos uma diminuição significativa do erro médio quadrático. A redução é, para a comunidade Depression, de 9,5%, para SuicideWatch, de 10,3%, para Anxiety, de 6,2% e para Bipolar, de 4,41%. Estes resultados mostram que é possível obter previsões muito superiores usando *embeddings* do que aquelas obtidas apenas atributos psicolinguísticos do LIWC.

Subreddit	Media MSE	DP MSE
Depression	0,6719	0,0225
SuicideWatch	0,6707	0,02568
Anxiety	0,6665	0,0264
Bipolar	0,6651	0,0191

Tabela 6.6: Resultados do MSE Modelo Rede Neural Multilayer Perceptron. Baseado em 5 folds em termos de média e desvio padrão.

6.4 Modelo Long Short Term Memory (LSTM)

Diferentemente do modelo apresentado na seção anterior, o modelo de rede neural recorrente (RNN) baseado em células LSTM considera que cada comentário e *post* são enviados em vetor de dimensão $n \times 300$, onde n corresponde ao *post* e a quantidade de comentário definida em cada árvore. Estas entradas “alimentam” o modelo seguindo a sequência temporal em que as publicações foram realizadas.

O *post* e os comentários de uma árvore de discussão são representados através de *embeddings* e lidos em ordem crescente de data de publicação pelo modelo LSTM para que ele estime a variação do TE. Assim, o aspecto temporal passa a ser considerado para a previsão da diferença do TE.

Na modelagem tradicional da rede LSTM, necessitamos que cada observação usada para treinar a rede seja do mesmo tamanho que as outras. Conforme a caracterização apresentada no Capítulo 4, as árvores de discussão possuem diferentes números de comentários. Assim, precisamos definir um valor fixo de comentários para cada árvore. Diante disso, realizamos testes para determinar uma boa escolha para o número de comentários a ser utilizado como entrada na rede.

As seguintes escolhas foram avaliadas: (i) a média de comentários existentes em uma árvore de discussão, (ii) a mediana, (iii) o valor máximo, (iv) o dobro da média e (v) o triplo da média. Nas comunidades Depression, SuicideWatch e Anxiety, a média de comentários na árvore de discussão é aproximadamente 4,5. Na Bipolar a média é maior que 5. Assim, arredondamos as médias para cima e o número fixo de entrada para a rede, é 5 para as comunidades Depression, SuicideWatch e Anxiety e 6 para a comunidade Bipolar. Este número também contempla a inclusão do *post* na árvore de discussão. Entre as cinco escolhas analisadas, as que resultam em menor MSE são as que consideram a média (i) e o dobro da média (iv).

Uma vez que definimos o tamanho fixo de cada ponto do conjunto de dados, precisamos definir estratégias para lidar com os cenários de árvores com menos ou mais comentários que o tamanho definido. Assim, apresentamos duas abordagens: preenchimento com *paddings* e preenchimento com replicação.

Preenchimento com *Paddings*. Seja T o tamanho fixo da entrada em número de publicações. As árvores de discussão com tamanho menor que T comentários, são preenchidas com *paddings* (vetores zerados) no início. As que têm mais que T comentários, por sua vez, serão truncadas, utilizando-se apenas os T últimos comentários. A premissa é que preencher as entradas com vetores zerados não impacta o resultado do modelo pois eles são uma representação sem conteúdo semântico e sintático. Escolhemos os últimos comentários da árvore, pois julgamos que esses são mais relevantes para entender a diferença do tom emocional do autor do *post*, visto que esses são os mais recentes.

Preenchimento com Replicação. Utilizamos também a replicação de alguns comentários sorteados aleatoriamente para preenchimento das árvores de discussões com tamanho menor que o definido. Por exemplo, suponha que o tamanho fixo da árvore analisada será 6 comentários e que uma árvore possui 4 comentários. Assim, para o preenchimento das 2 entradas faltantes, escolhemos aleatoriamente 2 dos 6 comentários existentes e o inserimos na mesma ordem em que a publicação correspondente (comentário sorteado) apareceu. A premissa desta abordagem é que inserir comentários iguais e na mesma ordem não impacta significativamente a saída do modelo. Nas árvores com número maior do que o máximo definido, removemos os primeiros comentários.

Executamos os seguintes modelos para a comunidade Bipolar: LSTM padrão, LSTM bidirecional (concatenação e soma), rede LSTM dividida (o *post* é lido em uma camada separada). Consideramos duas formas de se definir a entrada da rede LSTM:

- **Abordagem 1:** consideramos todas as sequências, desde o *post* até os comentários, excluindo o último comentário do autor da árvore de discussão.
- **Abordagem 2:** utilizamos o *post* seguido apenas dos comentários dos usuários que não são autores da árvore de discussão em questão.

Cada uma dessas abordagens tem suas especificidades. Na abordagem 1, consideramos informações dos comentários do autor do *post* e dos demais usuários que interagiram naquela árvore. Assim é possível entender o progresso da variação do TE de um ponto de vista mais geral. Na abordagem 2, a premissa é que apenas os comentários dos outros usuários podem alterar o estado emocional codificado pelo *post* que iniciou a *thread*.

A Tabela 6.7 apresenta os resultados obtidos para a comunidade Bipolar utilizando as abordagens 1 e 2. Observamos que os melhores resultados foram obtidos

usando a rede LSTM padrão com entrada construída a partir da Abordagem 2. Embora a superioridade da escolha $T = 6$ seja estatisticamente significativa quando o preenchimento com zeros é utilizado, a diferença não é estatisticamente significativa quando a replicação de comentários é utilizada. A partir destes resultados, iremos aplicar o modelo LSTM padrão com a abordagem 2 para as demais comunidades, também considerando a validação cruzada de 5 folds.

Modelos	Tipo	Tamanho Árvore	Abordagem 1		Abordagem 2	
			Média MSE	DP MSE	Média MSE	DP MSE
LSTM Padrão	Preenchimento com zeros	6	0,6689	0,0158	0,5747	0,0257
		12	0,6796	0,0734	0,6771	0,0227
	Replicação	6	0,6898	0,0075	0,6920	0,0140
		12	0,7103	0,0143	0,7014	0,0202
LSTM Dividida	Preenchimento com zeros	6	0,7464	0,0183	0,7553	0,0147
		12	0,7361	0,0299	0,7230	0,0281
	Replicação	6	0,7312	0,076	0,7387	0,0320
		12	0,7098	0,0074	0,7122	0,0316
LSTM Bidirecional concatenação	Preenchimento com zeros	6	0,7186	0,0233	0,7108	0,0219
		12	0,7451	0,0151	0,7606	0,0256
	Replicação	6	0,7464	0,0258	0,7237	0,0166
		12	0,7716	0,0105	0,7911	0,0269
LSTM Bidirecional soma	Preenchimento com zeros	6	0,7665	0,0367	0,7323	0,0192
		12	0,7523	0,0260	0,7498	0,0516
	Replicação	6	0,7961	0,0338	0,7629	0,0111
		12	0,7579	0,0113	0,8023	0,0131

Tabela 6.7: Resultados dos modelos executados para o subreddit Bipolar (Abordagem 1 e 2) com 5 folds.

A Tabela 6.8 apresenta os resultados da abordagem 2 utilizando a LSTM padrão para as demais comunidades, os menores MSEs estão destacados. Comparando-se os resultados obtidos pela rede LSTM padrão com aqueles obtidos pelo Baseline 3, observamos ganhos de acurácia de 21,6% na comunidade Depression, 18,7% na SuicideWatch, 18,4% na Anxiety e 29,2% na Bipolar. Em comparação com os resultados dos modelos RLM com regularização observamos uma variação significativa na acurácia da predição do TE: para a comunidade Depression o ganho é de 10,4%, para comunidade SuicideWatch 8,8%, na Anxiety 6,7% e na Bipolar 17,3%. Estes resultados indicam que modelos de redes neurais com representações de texto baseadas em *embeddings* resultam em previsões mais acuradas da variação do TE do que modelos mais simples criados a partir dos atributos psicolinguísticos LIWC.

Finalmente, analisamos se a ordem das publicações (sequência de *post* e comentários) influencia na acurácia da previsão da variação do TE. Assim, comparamos os resultados obtidos com a LSTM padrão com o modelo *Multilayer Perceptron*. Na comunidade Depression, o MSE é 1,1% menor que aquele obtido com o *Multilayer Perceptron*, na SuicideWatch, -1,7% (i.e., aumentou 1.7%), na Anxiety, 0,5% e na Bipolar, 13,6%. Note que na comunidade SuicideWatch, a LSTM aumentou o erro MSE, um

Modelos	Tipo	Tamanho	Média MSE	DP MSE
Depression	Preenchimento com zeros	5	0,6648	0,0082
		10	0,6861	0,0259
	Replicação	5	0,6800	0,0148
		10	0,6921	0,0200
SuicideWatch	Preenchimento com zeros	5	0,6818	0,0028
		10	0,6994	0,0249
	Replicação	5	0,7897	0,0105
		10	0,7149	0,0303
Anxiety	Preenchimento com zeros	5	0,6631	0,01459
		10	0,6638	0,0111
	Replicação	5	0,6931	0,0036
		10	0,6690	0,0093
Bipolar	Preenchimento com zeros	6	0,5747	0,0257
		12	0,6771	0,0227
	Replicação	6	0,6920	0,0140
		12	0,7014	0,0202

Tabela 6.8: Resultados dos modelos LSTM padrão de todas comunidades abordagem 2 com 5 folds.

indício que, aparentemente, a ordem das publicações nesta comunidade não interfere na variação do TE. Uma possível explicação para a ordem não impactar a variação do TE, é que o conteúdo semântico e sintático do texto e o tom emocional dos comentários nessas árvores de discussões são muito parecidos. Assim, não considerar a estrutura da árvore de discussão, talvez contribua na obtenção melhores previsões.

6.5 Discussão dos Resultados

Nesta seção são discutidos os principais pontos da análise experimental: como os modelos propostos superam o baseline mais forte (B_3); representação dos *posts* e comentários usando *embeddings* vs. atributos psicolinguísticos; e importância do aspecto temporal das publicações na previsão da variação do TE.

Todos os modelos propostos, especificamente, RLM com regularização, *Multilayer Perceptron* (MLP), e as variações do LSTM, superam a acurácia do Baseline 3. O modelo RLM diminui em média 12,6% do MSE do B_3 em todas comunidades. Já o modelo MLP reduziu, em média, 19,22% do MSE. O modelo em que o aspecto temporal é considerado, LSTM, reduz o erro em até 22%. Estes resultados mostram que as redes neurais (MLP e LSTM) geram melhores resultados do que a RLM, por usarem representações vetoriais densas (embeddings) e modelos mais flexíveis.

Os *embeddings*, apesar de serem fatores latentes que, ao contrário dos atributos do LIWC, não permitem uma interpretação direta das dimensões, são melhores preditores. Isso significa que eles carregam mais informações sobre o texto do que o LIWC. Os modelos que utilizam os *embeddings* diminuem o MSE em ao menos 4,8% em relação aos MSE obtido pela RLM com regularização que utiliza atributos do LIWC.

Verificamos que levar o fator temporal em consideração, através das redes LSTM, levou a uma nova redução no MSE em todas as comunidades, exceto na SuicideWatch, onde a modelagem da ordem do *post* e dos comentários não diminuiu o erro. A causa desta diferença carece de mais investigação.

Em suma, percebemos que os modelos propostos são eficazes para capturar a variação do TE do usuário em uma árvore de discussão. Esses modelos podem ser utilizados no monitoramento de conversas, por exemplo, em redes sociais de suporte a usuários com problemas de saúde mental. Nosso modelo pode ser um guia para prever se o TE do autor da árvore de discussão irá ficar mais positivo diante dos comentários recebidos.

O presente trabalho apresenta limitações, visto que não é possível verificar se uma variação positiva no TE está correlacionada com uma melhora real no estado emocional do usuário, já que não temos um contato direto com usuário para averiguação. Porém, nosso modelo, abre possibilidades para o monitoramento da variação do TE do usuário. Caso o TE não esteja evoluindo, intervenções podem ser realizadas por profissionais da área.

Capítulo 7

Conclusão e Trabalhos Futuros

Esta dissertação teve como foco analisar as comunidades relacionadas a transtornos de saúde mental do Reddit. Utilizamos comunidades em que usuários discutem sobre suas experiências relacionadas a ansiedade, bipolaridade, depressão e suicídio. Primeiramente, analisamos as interações entre os usuários, através da modelagem de grafos. Por meio das métricas extraídas do grafo, verificamos que a baixa transitividade combinada com alto diâmetro e excentricidade do nó do grafo corroboram com o fato de que as interações do usuário são baseadas principalmente no conteúdo das postagens e comentários, independentemente dos usuários que os geram. Esta análise responde parte da primeira questão de pesquisa abordada nesta dissertação.

Também analisamos o volume de atividades (*posts* e comentários) realizadas pelos usuários, em acréscimo à resposta da QP1. Verificamos que a maior parte das interações são realizadas na segunda-feira, possivelmente influenciados pela rotina semanal, que gera infelicidade para várias pessoas. Além disso, foi possível observar que as atividades são menores no início do ano, provavelmente influenciados pela expectativa de um novo ano com novas oportunidades.

Para caracterizar as conversas que ocorrem nas comunidades, analisamos as *threads*, modelando-as como árvores de discussões. Em particular, realizamos uma pequena pesquisa para analisar as 12 árvores de discussão mais profundas. Apoiados por avaliadores, concluímos que a maioria das postagens são solicitações de ajuda e que, na maioria das vezes, vários usuários oferecem ajuda. A maioria das árvores mais profundas são conversas entre dois usuários, o que responde parcialmente à nossa segunda questão de pesquisa.

Encontramos grupos de usuários que desempenham papéis semelhantes dentro do Reddit, complementando a resposta à QP2. Esses grupos foram definidos pelo nível de interação dos usuários nas suas próprias árvores de discussão e nas de outros

usuários. Percebemos que existem usuários que interagem apenas em suas próprias publicações, possivelmente procurando ajuda/conselho. Há outros que não iniciam discussões (posts). No entanto, estes usuários comentam, com grande frequência, os *posts* de outros usuários, o que os caracteriza como usuários possivelmente conselheiros, os quais tentam ajudar o usuário que passa por alguma situação.

Usando o modelo RMN, detectamos automaticamente tópicos que descrevem o que é discutido com mais frequência nas quatro comunidades. Como esperado, os tópicos relacionados à ideiação suicida são mais comuns nas comunidades de Suicídio e Depressão. Além disso, descobrimos que as comunidades Depression e SuicideWatch são semelhantes para a maioria dos tópicos. Também analisamos os usuários mais relevantes para cada tópico. Por exemplo, para o tópico relacionado à superação de problemas, os usuários mais relevantes fazem parte das comunidades bipolar, suicídio ou ansiedade.

Além disso, para responder à terceira questão de pesquisa, definimos o tom emocional (TE) do usuário em uma publicação. Os resultados mostram que o TE do último comentário do autor do *post* é maior que o TE do *post* em 67% no Depression, 71% no SuicideWatch, 71,8% no Anxiety e 62,3% no Bipolar. O que mostra que, na maioria dos casos, os usuários melhoram seu humor dentro da árvore de discussão.

Dado que os comentários influenciam o TE de quem realiza uma publicação nestas comunidades, propomos modelos para capturar a variação do TE. Utilizamos atributos psicolinguísticos e atributos relacionados a informações sintáticas e semânticas, os *embeddings*. Nossos modelos identificaram com boa acurácia a variação do TE dos usuários. Comparando o baseline de maior acurácia com o melhor modelo proposto (baseado em redes neurais) o *Mean Square Error* reduz consideravelmente, o que evidencia a possibilidade de monitorar a evolução do tom emocional de tais usuários dentro das discussões realizadas nestas comunidades.

Apesar dos resultados dos modelos apresentarem uma acurácia satisfatória, os mesmos ainda podem ser melhorados. Para isso, como trabalhos futuros, pretendemos aplicar os modelos combinando mais atributos, os gerados pelo LIWC e também os *embeddings*. Além disso, desejamos aplicar outras arquiteturas de rede como o *Transformer* [Vaswani et al., 2017], que é uma arquitetura, relativamente nova, para resolver tarefas como a deste trabalho, pois ela trabalha bem com dependências de longo alcance. Ademais, planejamos utilizar camadas de atenção para explicitar quais atributos mais impactam na variação do TE da árvore.

Uma direção complementar à análise do TE nas árvores de discussão, é identificar os perfis dos usuários que participam das discussões na rede. A partir dos perfis encontrados na caracterização apresentada, podemos correlacionar estes perfis com os

tópicos abordados nas árvores de discussão e compreender se existe um perfil que domine as discussões em árvores de maior profundidade, por exemplo. Outro item, a ser explorado é a definição faixas para preenchimento dos comentários faltantes em relação ao tamanho máximo definido para a árvore.

Os resultados deste trabalho podem ser utilizados no monitoramento de usuários que sofrem de transtornos mentais. Ao identificar que o usuário não está melhorando seu tom emocional, especialistas podem intervir para evitar uma situação extrema. Para que isso ocorra, pretendemos realizar parcerias com especialistas. Também almejamos a criação de um *bot* para auxiliar os usuários que passam por algum transtorno mental. Este *bot* estabelecerá um contato com o usuário que está com o problema, e o nosso modelo servirá de monitoramento sobre a melhora do usuário, estabelecendo uma resposta mais rápida e direta sobre o tom emocional do usuário e, caso seja necessário, especialistas poderão intervir.

Referências Bibliográficas

- Alambo, A.; Gaur, M.; Lokala, U.; Kursuncu, U.; Thirunarayan, K.; Gyrard, A.; Sheth, A.; Welton, R. S. & Pathak, J. (2019). Question answering for suicide risk assessment using reddit. Em *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pp. 468–473. ISSN .
- Althoff, T.; Clark, K. & Leskovec, J. (2016). Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, 4:463.
- Althoff, T.; Jindal, P. & Leskovec, J. (2017). Online actions with offline impact: How online social networks influence online and offline user behavior. Em *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 537--546. ACM.
- Baba, T.; Baba, K. & Ikeda, D. (2019). Detecting mental health illness using short comments. Em *International Conference on Advanced Information Networking and Applications*, pp. 265--271. Springer.
- Bakshy, E.; Hofman, J. M.; Mason, W. A. & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. Em *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pp. 65--74. ACM.
- Barney, L. J.; Griffiths, K. M.; Jorm, A. F. & Christensen, H. (2006). Stigma about depression and its impact on help-seeking intentions. *Australian & New Zealand Journal of Psychiatry*, 40(1):51--54.
- Bengio, Y. & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, 5(Sep):1089--1105.
- Bholowalia, P. & Kumar, A. (2014). Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9).

- Bird, S.; Klein, E. & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc."
- Blair, J. & Abdullah, S. (2018). Supporting constructive mental health discourse in social media. Em *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '18, pp. 299--303, New York, NY, USA. ACM.
- Blei, D. M.; Ng, A. Y. & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993--1022. ISSN 1532-4435.
- Caplan, S. E. & Turner, J. S. (2007). Bringing theory to research on computer-mediated comforting communication. *Computers in Human Behavior*, 23(2):985 – 998. ISSN 0747-5632. Special Issue: Internet and Well-Being in Honor of the Memory of Michael Argyle.
- Chen, X.; Sykora, M. D.; Jackson, T. W. & Elayan, S. (2018). What about mood swings: Identifying depression on twitter with temporal measures of emotions. Em *Companion Proceedings of the The Web Conference 2018*, WWW '18, pp. 1653--1660, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- Coppersmith, G.; Dredze, M. & Harman, C. (2014). Quantifying mental health signals in twitter. Em *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pp. 51--60.
- Coppersmith, G.; Leary, R.; Crutchley, P. & Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10:1178222618792860.
- Costa, A.; Nalon, R.; Meira Jr., W. & Veloso, A. (2018). Ego-centric analysis of supportive networks. Em *Proceedings of the 10th ACM Conference on Web Science*, WebSci '18, pp. 281--285, New York, NY, USA. ACM.
- Cranford, J. A.; Shrout, P. E.; Iida, M.; Rafaeli, E.; Yip, T. & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, 32(7):917--929.
- Cunha, T.; Weber, I. & Pappa, G. (2017). A warm welcome matters!: The link between social feedback and weight loss in/r/loseit. Em *Proceedings of the 26th International*

- Conference on World Wide Web Companion*, pp. 1063--1072. International World Wide Web Conferences Steering Committee.
- Davidov, D.; Tsur, O. & Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. Em *Proceedings of the 23rd international conference on computational linguistics: posters*, pp. 241--249. Association for Computational Linguistics.
- De Choudhury, M. (2013). Role of social media in tackling challenges in mental health. Em *Proceedings of the 2nd international workshop on Socially-aware multimedia*, pp. 49--52. ACM.
- De Choudhury, M.; Counts, S. & Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. Em *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 3267--3276. ACM.
- De Choudhury, M. & De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. Em *International Conference on Web and Social Media*.
- De Choudhury, M.; Kiciman, E.; Dredze, M.; Coppersmith, G. & Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. Em *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 2098--2110. ACM.
- Dutta, S.; Ma, J. & De Choudhury, M. (2018). Measuring the impact of anxiety on online social interactions. Em *Proceedings of the 12th International AAAI Conference on Web and Social Media*.
- Falavarjani, S. A. M.; Zarrinkalam, F.; Jovanovic, J.; Bagheri, E. & Ghorbani, A. A. (2019). The reflection of offline activities on users' online social behavior: An observational study. *Information Processing & Management*, 56(6):102070. ISSN 0306-4573.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.
- Gardner, M. W. & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15):2627--2636.
- Gers, F. A.; Schmidhuber, J. & Cummins, F. (1999). Learning to forget: Continual prediction with lstm.

- Gkotsis, G.; Oellrich, A.; Hubbard, T.; Dobson, R.; Liakata, M.; Velupillai, S. & Dutta, R. (2016). The language of mental health problems in social media. Em *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*.
- Gkotsis, G.; Oellrich, A.; Velupillai, S.; Liakata, M.; Hubbard, T. J.; Dobson, R. J. & Dutta, R. (2017). Characterisation of mental health conditions in social media using informed deep learning. *Scientific reports*, 7:45141.
- Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies*, 10(1):1--309.
- Gruda, D. & Hasan, S. (2019). Feeling anxious? perceiving anxiety in tweets using machine learning. *Computers in Human Behavior*, 98:245--255.
- Hartigan, J. A. & Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100--108.
- Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1--12.
- Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107--116.
- Holtzman, N. S. et al. (2017). A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, 68:63--68.
- Hutto, C. J. & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. Em *Eighth international AAAI conference on weblogs and social media*.
- Hutto, C. J.; Yardi, S. & Gilbert, E. (2013). A longitudinal study of follow predictors on twitter. Em *Proceedings of the sigchi conference on human factors in computing systems*, pp. 821--830. ACM.
- Islam, M. R.; Kabir, M. A.; Ahmed, A.; Kamal, A. R. M.; Wang, H. & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*, 6(1):8. ISSN 2047-2501.
- Iyyer, M.; Guha, A.; Chaturvedi, S.; Boyd-Graber, J. L. & III, H. D. (2016). Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. Em *HLT-NAACL*, pp. 1534--1544.

- Jordan, K. N.; Pennebaker, J. W. & Ehrig, C. (2018). The 2016 us presidential candidates and how people tweeted about them. *SAGE Open*, 8(3):2158244018791218.
- Kavuluru, R.; Ramos-Morales, M.; Holaday, T.; Williams, A. G.; Haye, L. & Cerel, J. (2016). Classification of helpful comments on online suicide watch forums. Em *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, New York, USA.
- Laniado, D.; Tasso, R.; Volkovich, Y. & Kaltenbrunner, A. (2011). When the wikipedians talk: Network and tree structure of wikipedia discussion pages. Em *International Conference on Web and Social Media*.
- Lee, K. S.; Lee, H.; Myung, W.; Song, G.-Y.; Lee, K.; Kim, H.; Carroll, B. J. & Kim, D. K. (2018). Advanced daily prediction model for national suicide numbers with social media data. *Psychiatry investigation*, 15(4):344.
- Leis, A.; Ronzano, F.; Mayer, M. A.; Furlong, L. I. & Sanz, F. (2019). Detecting signs of depression in tweets in spanish: Behavioral and linguistic analysis. *J Med Internet Res*, 21(6):e14199. ISSN 1438-8871.
- Lopes, C. R. S.; Cunha, M.; Rodrigues, A.; Vilela, A. B. A.; Casotti, C. A. & Pereira, H. (2014). Identificando as representações sociais sobre promoção da saúde em uma rede social de trabalhadores de saúde. Em *Proceedings of the III Brazilian Workshop on Social Network Analysis and Mining, Brasília-DF Brazil*.
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Em *Advances in neural information processing systems*, pp. 3111--3119.
- Miller Jr, R. G. (1997). *Beyond ANOVA: basics of applied statistics*. Chapman and Hall/CRC.
- Murrieta, J.; Frye, C. C.; Sun, L.; Ly, L. G.; Cochancela, C. S. & Eikey, E. V. (2018). # depression: Findings from a literature review of 10 years of social media and depression research. Em *International Conference on Information*, pp. 47--56. Springer.
- Myers, R. H. & Myers, R. H. (1990). *Classical and modern regression with applications*, volume 2. Duxbury press Belmont, CA.
- Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2):167--256.

- Organization, W. H. et al. (2013). Mental health action plan 2013-2020.
- Organization, W. H. et al. (2017). Depression and other common mental disorders: global health estimates. Relatório técnico, World Health Organization.
- Pappa, G. L.; Cunha, T. O.; Bicalho, P. V.; Ribeiro, A.; Silva, A. P. C.; Meira Jr, W. & Beilegoli, A. M. R. (2017). Factors associated with weight change in online weight management communities: A case study in the loseit reddit community. *Journal of Medical Internet Research*, 19(1).
- Park, A. & Conway, M. (2018). Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: Analysis of texts from mental health communities. *Journal of Medical Internet Research*, 20(4).
- Pennebaker, J.; Booth, R.; Boyd, R. & Francis, M. (2015). Liwc 2015 operator's manual.
- Pennington, J.; Socher, R. & Manning, C. (2014). Glove: Global vectors for word representation. Em *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532--1543.
- Picard, R. R. & Cook, R. D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387):575--583.
- Prechelt, L. (1998). Early stopping-but when? Em *Neural Networks: Tricks of the trade*, pp. 55--69. Springer.
- Ramírez-Cifuentes, D.; Mayans, M. & Freire, A. (2018). Early risk detection of anorexia on social media. Em Bodrunova, S. S., editor, *Internet Science*, pp. 3--14, Cham. Springer International Publishing.
- Ramseyer Winter, V.; Gillen, M. M.; Cahill, L.; Jones, A. & Ward, M. (2017). Body appreciation, anxiety, and depression among a racially diverse sample of women. *Journal of Health Psychology*.
- Rumelhart, D. E.; Hinton, G. E. & Williams, R. J. (1985). Learning internal representations by error propagation. Relatório técnico, California Univ San Diego La Jolla Inst for Cognitive Science.
- Saha, K.; Weber, I. & De Choudhury, M. (2018). A social media based examination of the effects of counseling recommendations after student deaths on college campuses.

- Em *In Proceedings of the 12th International AAAI Conference on Web and Social Media*.
- Sahota, P. K. & Sankar, P. L. (2019). Bipolar disorder, genetic risk, and reproductive decision-making: A qualitative study of social media discussion boards. *Qualitative health research*, p. 1049732319867670.
- Schuster, M. & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673--2681.
- Seabrook, M. E.; Kern, L. M.; Fulcher, D. B. & Rickard, S. N. (2018). Predicting depression from language-based emotion dynamics: Longitudinal analysis of facebook and twitter status updates. *Journal of Medical Internet Research*, 20(5):e168.
- Shen, J. H. & Rudzicz, F. (2017). Detecting anxiety through reddit. Em *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*, pp. 58--65.
- Sigurdsson, V.; Larsen, N. M.; Sigfusdottir, A. D.; Fagerstrøm, A.; Alemu, M. H.; Folwarczny, M. & Foxall, G. (2019). The relationship between the firm's social media strategy and the consumers' engagement behavior in aviation. *Managerial and Decision Economics*.
- Souza, B. Á. S.; Almeida, T. G. A.; Menezes, A. A. M.; Figueiredo, C. M. F.; Nakamura, F. G. N. & Nakamura, E. F. N. (2017). Uma abordagem para detecção de tópicos relevantes em redes sociais online. Em *Proceedings of the VI Brazilian Workshop on Social Network Analysis and Mining, São Paulo, SP, Brazil*.
- Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I. & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929--1958.
- Stawarz, K.; Preist, C. & Coyle, D. (2019). Use of smartphone apps, social media, and web-based resources to support mental health and well-being: Online survey. *JMIR Ment Health*, 6(7):e12546. ISSN 2368-7959.
- Stone, A. A.; Hedges, S. M.; Neale, J. M. & Satin, M. S. (1985). Prospective and cross-sectional mood reports offer no evidence of a "blue monday" phenomenon. *Journal of Personality and Social Psychology*, 49(1):129.

- Tausczik, Y. R. & Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24--54.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267--288.
- Van Mieghem, P. (2011). Human psychology of common appraisal: The reddit score. *IEEE Transactions on Multimedia*, 13(6):1404--1406.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. Em *Advances in neural information processing systems*, pp. 5998--6008.
- Wang, A.; Hamilton, W. L. & Leskovec, J. (2016). Learning linguistic descriptors of user roles in online communities. Em *Proceedings of the First Workshop on NLP and Computational Social Science*, pp. 76--85.
- Weerasinghe, J.; Morales, K. & Greenstadt, R. (2019). "because... i was told... so much": Linguistic indicators of mental health status on twitter. *Proceedings on Privacy Enhancing Technologies*, 2019(4):152--171.
- Wolohan, J.; Hiraga, M.; Mukherjee, A.; Sayyed, Z. A. & Millard, M. (2018). Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with NLP. Em *Proceedings of the First International Workshop on Language Cognition and Computational Models*, pp. 11--21, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Wongkoblaph, A.; Vadillo, M. A. & Curcin, V. (2018). A multilevel predictive model for detecting social network users with depression. Em *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pp. 130--135. ISSN .
- Wongkoblaph, A.; Vadillo, M. A. & Curcin, V. (2019). Modeling depression symptoms from social network data through multiple instance learning. *AMIA Summits on Translational Science Proceedings*, 2019:44.
- Yates, A.; Cohan, A. & Goharian, N. (2017). Depression and self-harm risk assessment in online forums. *arXiv preprint arXiv:1709.01848*.
- Yusof, N. F. A.; Lin, C. & Guerin, F. (2018). Assessing the effectiveness of affective lexicons for depression classification. Em *International Conference on Applications of Natural Language to Information Systems*, pp. 65--69. Springer.