

Análise de Comunidades de Suporte a Transtornos de Saúde Mental do Reddit*

Bárbara Silveira, Ana Paula Couto da Silva, Fabricio Murai

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brazil

{barbarasilveira, ana.coutosilva, murai}@dcc.ufmg.br

Abstract. *In the last years, online social networks have gained new functionalities and goals. Initially designed for fostering friendships and for exchanging images or videos, they started to connect people willing to share experiences related to health problems, such as obesity and depression. In this work, we characterize users from four Reddit communities centered on sharing experiences related to mental health problems, named: Depression, SuicideWatch, Anxiety and Bipolar. The focus of this paper lies on analyzing (i) these users activity, (ii) the social support provided by these communities and (iii) the experiences shared through posts and comments.*

Resumo. *Nos últimos anos, redes sociais online ampliaram suas funcionalidades e objetivos. Inicialmente focadas em fomentar amizades, troca de imagens ou vídeos, passaram a conectar pessoas dispostas a trocar experiências relacionadas à problemas de saúde, como por exemplo, obesidade e depressão. Neste trabalho, caracterizamos os usuários de quatro comunidades do Reddit centradas na troca de experiências relacionadas à problemas de saúde mental, intituladas: Depression, SuicideWatch, Anxiety e Bipolar. O enfoque principal deste artigo é na análise (i) da atividade destes usuários, (ii) do suporte social oferecido por estas comunidades, e (iii) das experiências compartilhadas através de posts e comentários.*

1. Introdução

Nos últimos anos, dados da Organização Mundial da Saúde (OMS) alertam para o aumento do total de pessoas no mundo que sofrem de algum tipo de transtorno de saúde mental. Como exemplo, um relatório global lançado recentemente pela mesma organização¹, 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. No Brasil, a depressão atinge 11,5 milhões de pessoas (5,8% da população), enquanto distúrbios relacionados à ansiedade afetam mais de 18,6 milhões de brasileiros (9,3% da população).

Devido ao aumento de casos destes tipos de transtornos, políticas de saúde pública eficientes devem ser implementadas. No caso específico da depressão, a OMS é responsável pelo programa *Mental Health Gap Action Programme*, que visa ajudar os países

*The authors' work has been partially funded by the EUBra-BIGSEA project by the European Commission under the Cooperation Programme (MCTI/RNP 3rd Coordinated Call), Horizon 2020 grant agreement 690116, CAPES, CNPq and FAPEMIG.

¹<https://news.un.org/>

a aumentar os 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.

Mais ainda, a combinação entre recursos escassos (principalmente entre países de economia mais instável), o estigma social associado aos transtornos mentais e muitas vezes 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, como o suicídio. Menos da metade dos afetados no mundo (em muitos países, menos de 10%)² são diagnosticados corretamente em um quadro de depressão, por exemplo.

Este quadro geral faz com que novos recursos em busca de compreender e auxiliar os indivíduos afetados por estes transtornos sejam explorados. Por exemplo, podemos ressaltar o papel das redes sociais online. Inicialmente focadas em fomentar amizades, troca de imagens ou vídeos, passaram a conectar pessoas dispostas a trocar experiências relacionadas à problemas de saúde, como por exemplo, obesidade [Pappa et al. 2017] e depressão [Choudhury and De 2014, Kavuluru et al. 2016]. Outro trabalho [Lopes et al. 2014] utilizou uma rede social para identificar a percepção da promoção da saúde por grupo de profissionais da saúde. Neste contexto, nosso trabalho apresenta uma análise de comunidades direcionadas à discussão de transtornos de saúde mental. As observações produzidas por esta análise podem ajudar a guiar a realização mais eficiente de intervenções que auxiliem indivíduos que sofrem destas doenças.

O nosso estudo foca nas comunidades do Reddit, onde os membros compartilham suas experiências e dúvidas sobre os mais diversos assuntos. Uma 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. O Reddit é organizado em comunidades (subreddits) e em 2017 era formado por 1.204.126 comunidades com 900 milhões de comentários. Pelo menos 25 subreddits estão relacionados à transtornos da saúde mental. Nosso trabalho foca na análise dos quatro subreddits com o maior número de usuários ativos [Gkotsis et al. 2016, Gkotsis et al. 2017]: Depression (/r/depression), SuicideWatch (/r/suicide) – será abreviada como Suicide–, Anxiety (/r/anxiety) e Bipolar (/r/bipolar).

A partir do estudo das atividades dos usuários e na exploração do conteúdo compartilhado, as nossas principais contribuições são:

- Caracterizar o volume de atividades dos usuários dentro das comunidades, através do conceito de árvore de discussão.
- Analisar as experiências dos usuários aplicando 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.

²<https://nacoesunidas.org/oms-registra-aumento-de-casos-de-depressao-em-todo-o-mundo-no-brasil-sao-115-milhoes-de-pessoas/>

- Verificar indícios de apoio social oferecido por essas comunidades, a partir dos tópicos e descritores derivados pelo RMN.

Este artigo está organizado da seguinte forma: a Seção 2 descreve os principais trabalhos relacionados; a Seção 3 detalha os dados e métodos utilizados; os resultados da análise dos subreddits são apresentados na Seção 4; as implicações deste trabalho e trabalhos futuros são discutidos na Seção 5.

2. Trabalhos Relacionados

A aplicação de ferramentas para auxiliar pessoas que sofrem de diferentes transtornos de saúde mental, 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.

No entanto, o avanço dos meios de comunicação, sejam por telefones celulares ou através da Internet, está revolucionando o suporte oferecido às pessoas que sofrem de transtornos como depressão e ansiedade. Assim, alguns trabalhos na literatura buscam entender como novas tecnologias são eficazes em minimizar os sintomas destas doenças. A seguir, apresentamos os trabalhos que são mais relevantes ao escopo da análise que fazemos neste artigo.

Os autores em [Althoff et al. 2016] analisam conversas, via SMS, de indivíduos auxiliados por conselheiros ligados à organização *Crisis Trends*³. 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 [Gkotsis et al. 2016] investigou a linguagem dos *posts* do Reddit relacionados à saúde mental e identificou várias das características linguísticas dessas comunidades. Esses mesmos autores fizeram outro trabalho [Gkotsis et al. 2017] analisando as postagens do Reddit para desenvolver classificadores que reconheçam e classifiquem postagens relacionadas à doença mental, através da técnica de *deep learning*. Além disso, existem trabalhos [Souza et al. 2017, Wang et al. 2016] que abordam a caracterização de informações, através de extração de tópicos a fim de descrever o contexto analisado.

Considerando o uso do Reddit e comunidades direcionadas à saúde mental, o trabalho em [Choudhury and 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 *self-disclosure* em torno do tópico estigmático da saúde mental, uma vez que não existe a preocupação em ser identificado. Adicionalmente, postagens com menos inibição são as que possuem mais atenção e discutem problemas de relacionamentos e de saúde. Essas postagens reúnem maior apoio comunitário, através de votos e comentários. Os autores em [Kavuluru et al. 2016] focam a análise do subreddit *SuicideWatch*, com o objetivo de classificar, de forma automática, comentários que possam

³<https://crisistrends.org/>

impactar positivamente o comportamento de indivíduos com pensamentos suicidas (comentários *helpful specific*).

As análises apresentadas neste artigo diferem dos artigos descritos anteriormente em dois pontos principais. Primeiro, analisamos as atividades dos usuários em quatro comunidades, com o principal foco nos textos das publicações, isto é, dos *posts* e comentários. Avaliamos o engajamento dos usuários através das árvores de discussão construídas a partir das publicações. A segunda diferença é a aplicação do RMN (*Relationship Modeling Network*), proposto por [Iyyer et al. 2016], ao texto das publicações para modelar o papel dos usuários do Reddit nas comunidades das quais eles participam. Diferentemente do LDA (*Latent Dirichlet Allocation*) que associa documentos a um conjunto de tópicos latentes extraídos do corpus, o RMN tem como objetivo modelar a relação entre um par de entidades que aparece múltiplas vezes em diversos documentos ou contextos. Mais do que isso, o RMN modela a trajetória de um relacionamento dentro de cada contexto. Neste artigo, utilizamos o RMN para estudar a relação entre usuários do Reddit e suas respectivas comunidades.

3. Metodologia

Nesta seção descrevemos o conjunto de dados analisados (Seção 3.1), a maneira como os *posts* e comentários feitos pelos usuários serão analisados, através da criação de árvores de discussão (Seção 3.2), e a definição de descritores utilizando o *Relationship Modeling Network* (Seção 3.3).

3.1. Conjunto de Dados

Os dados de usuários do Reddit analisados neste trabalho foram recuperados online⁴. Existem mais de 25 subreddits que focam na discussão de transtornos mentais, entretanto selecionamos os quatro subreddits com o maior número de *posts* e comentários (desconsiderando o Opiates (*/r/opiates*), ligado à dependência de rémédios) [Gkotsis et al. 2016, Gkotsis et al. 2017]: Depression (*/r/depression*), SuicideWatch (*/r/suicide*), Anxiety (*/r/anxiety*) e Bipolar (*/r/bipolar*).

Todas as publicações dos usuários destas comunidades (*posts* e comentários) realizadas no ano de 2017 foram extraídas dos dados disponibilizados, totalizando 261.511 *posts* e 1.256.669 comentários de 207.683 usuários únicos. A Tabela 1 apresenta as principais características de cada um dos subreddits analisados.

	Depression	SuicideWatch	Anxiety	Bipolar
Usuários Únicos	43.322	13.940	105.879	44.542
Posts	44.288	24.349	145.072	47.802
Comentários	188.045	185.957	624.578	258.089
Posts por dia	121,34	66,71	397,46	130,96
Comentários por dia	515,19	509,47	1.711,17	707,09

Tabela 1. Estatísticas básicas dos subreddits analisados.

⁴<http://files.pushshift.io/reddit/>

3.2. Árvores de Discussão

Uma maneira de analisar a atividade gerada em torno dos tópicos iniciados pelas *threads* (em termos de *posts* e comentários) é a construção de árvores de discussão. Consideramos como árvore de discussão, toda a troca de informação que se inicia por um post (raiz da árvore), seguido de comentários feitos pelo próprio usuário que gerou o post ou por outros usuários na comunidade.

A análise da largura e profundidade destas árvores pode revelar 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. Encontrar tais usuários em um subreddit é importante para projetar intervenções direcionadas que podem auxiliar na melhoria da saúde mental de um grande número de participantes.

3.3. Relationship Modeling Network (RMN)

O RMN é uma rede neural recursiva projetada para modelar relacionamentos entre pares de entidades a partir de texto. Um relacionamento em um dado instante é representado como um vetor de pesos sobre K descritores. As entidades não precisam ser da mesma classe: por exemplo, neste trabalho, as entidades são os usuários e as comunidades nas quais eles interagem. Cada post ou comentário corresponde a um momento diferente.

As palavras são representadas por *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, geramos os *embeddings* usando o 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 deles. A representação de 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} .

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 suavidade $\alpha \in (0, 1)$ 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 $r_t = R^\top d_t$. 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 funcionamento do RMN.

Os dados de entrada para o RMN foram pré-processados da seguinte forma. Primeiramente, foram removidos todos os *posts* e comentários marcados como *[deleted]* ou *[removed]*, as *stopwords* (aplicando a biblioteca NLTK) e as pontuações. Segundo, foram considerados somente os *posts* e comentários dos usuários que realizaram no mínimo 50 atividades (*posts/comentários*) em cada subreddit, seguindo a metodologia apresentada em [Wang et al. 2016]. Por fim, foram selecionadas todas as palavras que aparecem nos quatro subreddits analisados, buscando encontrar semelhanças na forma com que as pessoas se expressam quando discutem sobre transtornos de saúde mental. Assim, o sub-

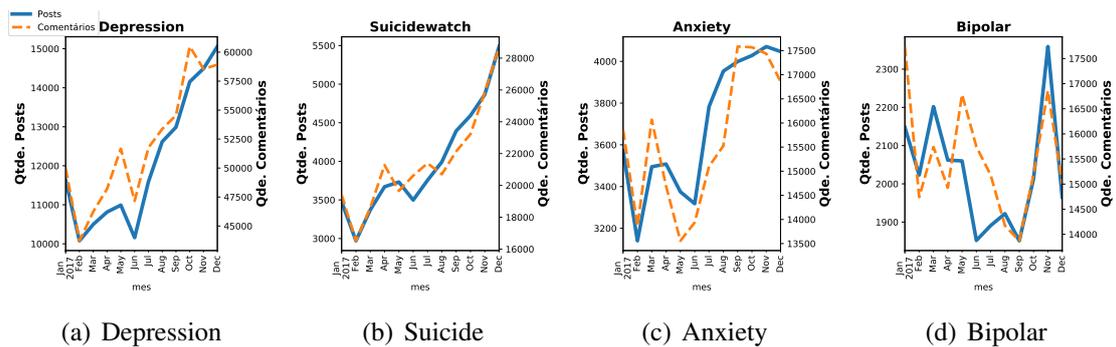


Figura 1. Evolução do total de *posts* e comentários em cada subreddit.

conjunto de dados analisados pelo RMN engloba 18.020 palavras únicas, 25.101 *posts* e 401.428 comentários.

4. Análise dos Subreddits

Nesta seção apresentamos a análise dos usuários dos quatro subreddits relacionados à transtornos de saúde mental. O primeiro conjunto de resultados descreve o volume de atividades dos usuários (Seção 4.1). A seguir, apresentamos uma análise detalhada das discussões dos usuários, dos tópicos que mais aparecem nos *posts* e comentários, bem como de alguns perfis de usuários extraídos da análise destes tópicos (Seção 4.2).

4.1. Atividades dos Usuários

A Figura 1 apresenta o volume mensal de *posts* e comentários em cada um dos subreddits analisados. É interessante notar que, exceto para o subreddit Bipolar, o número mensal de publicações manteve uma tendência de crescimento durante quase todo o ano de 2017. No caso da comunidade Anxiety, o mês de Maio revela uma diminuição no volume de *posts* e comentários, seguido de uma crescimento substancial. Outro ponto a ser ressaltado é que os picos de atividades nestas comunidades podem estar correlacionados com eventos reais que geram grande comoção entre pessoas. Por exemplo, se considerarmos o subreddit Suicide, entre os meses de Fevereiro e início de Abril o total de *posts* aumentou $\approx 16\%$ enquanto o total de comentários aumentou em $\approx 30\%$. 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 adolescentes ⁵.

Uma outra maneira de quantificar o volume de atividades 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 2 apresenta o histograma contendo o número de pares encontrados quantizados por números de interações. Podemos observar que a grande maioria dos usuários estabelecem poucos diálogos entre si (primeira coluna dos histogramas apresentados). Este resultado corrobora a filosofia da rede social do 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).

⁵<https://www.theatlantic.com/entertainment/archive/2017/08/13-reasons-why-demonstrates-cultures-power/535518/>

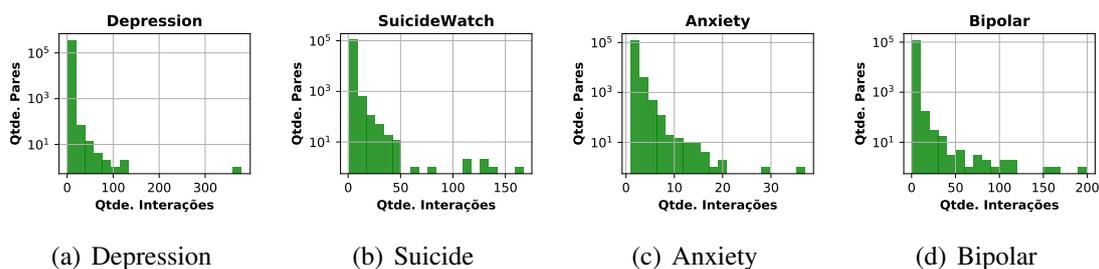


Figura 2. Histograma de números de pares de usuários quantizados por volume de interação.

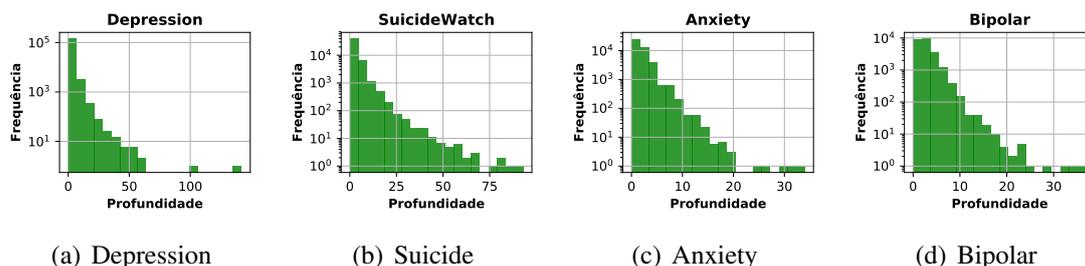


Figura 3. Histograma da profundidade das árvores de discussão.

4.2. Posts e Comentários

A seguir apresentamos as análises dos *posts* e comentários dos usuários sob dois pontos de vista: o primeiro, análise da árvore de discussão para identificar o engajamento dos usuários em pedir ajuda e serem atendidos por outros usuários; o segundo, através dos descritores que caracterizam os *posts* e comentários dos usuários, revelando, por exemplo, quais são os temas mais discutidos nestas comunidades e possíveis perfis dos usuários.

Árvores de Discussão

A Figura 3 mostra a distribuição da profundidade das árvores de discussão em cada um dos subreddits. Um ponto a ser ressaltado é que nos subreddits Depression e Suicide, quase a totalidade das árvores de discussão possuem profundidade de até 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). Uma possível explicação para este resultado é que as discussões nos dois primeiros subreddits tendem a ter maior complexidade ou gravidade, fomentando, assim, maior suporte social entre os usuários.

A Tabela 2 traz uma investigação detalhada das árvores de maior profundidade de cada subreddit. É interessante notar que os usuários realmente buscam ajuda nestes tipos de comunidades, relatando abertamente os problemas pelos quais estão passando (veja abaixo, por exemplo, o post que gera a maior árvore de discussão no subreddit Suicide). Podemos ressaltar também a importância do papel de um membro que atua como moderador. Este perfil de membro pode motivar os demais a compartilharem o estado emocional, fazendo com que os outros usuários iniciem um processo de aconselhamento e ajuda.

Com o objetivo de melhor caracterizar as árvores que fomentam maior engajamento dos usuários, realizamos uma análise manual daquelas com as maiores profundida-

Subreddit	Publicações	Profundidade	Análise
Depression	769.650	142	Post feito pelo moderador da comunidade (“skyqween”) incentivando os usuários a relatarem como estão. Alguns usuários oferecem ajuda, gerando suporte social. Pelo total de <i>posts</i> e comentários feitos, podemos inferir que os usuários compartilham seus estados emocionais quando são motivados.
Suicide	305.891	93	Usuário relata estar sofrendo de pensamentos suicidas. Um outro usuário estabelece um diálogo, questionando o estado emocional do usuário que fez o post, oferecendo ajuda.
Anxiety	232.333	34	Usuário relata ansiedade por ter feito uma entrevista de emprego e demais usuários trocam experiências similares, oferecendo ajuda.
Bipolar	210.306	37	Usuário pede ajuda e um diálogo é iniciado com um outro usuário.

Tabela 2. Análise das árvores com maior profundidade.

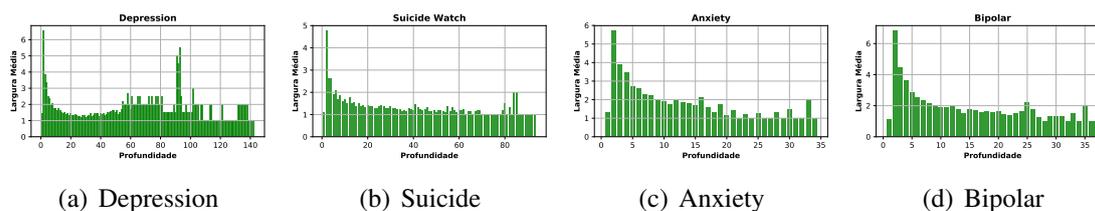


Figura 4. Largura média por nível das árvores de discussão (condicionado a profundidade > nível). Nós de mesma profundidade estão no mesmo nível da árvore.

des. Através da inspeção de 16 árvores de discussão, obtivemos as seguintes conclusões:

1. Usuários que iniciam a árvore de discussão tendem a escrever *posts* procurando por ajuda. A partir disso, um diálogo se inicia, e se estabelece, na maior parte das vezes, com somente um usuário em específico.
2. Árvores de discussão de maior profundidade são aquelas em que as pessoas que estão dispostas a ajudar demonstram mais interesse no problema da pessoa que pediu ajuda. O interesse pode ser mostrado, por exemplo, através de perguntas que buscam entender melhor a situação real do usuário que pediu ajuda.
3. Usuários que iniciam uma árvore de discussão com um *post* formado por muitas palavras, tendem a ter árvores de discussão de maior profundidade onde o usuário que está ajudando faz comentários longos. Este resultado, de uma certa forma, é semelhante ao do artigo [Althoff et al. 2016], pois uma das características de um bom conselheiro é esclarecer problemas de ambiguidade. Assim, membros com perfil de conselheiros tendem a escreverem mais a fim de deixar bem claro o problema pelo qual o usuário que iniciou a árvore de discussão tem vivido.

Seguindo a análise das árvores de discussão, a Figura 4 mostra a largura média a cada nível de profundidade das árvores de discussão. A largura média é 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.

A partir da Figura 4 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 resultados é a existência de moderadores que postam semanalmente incentivando os usuários a falarem 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.

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 meny 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 sweetie
28	books book useful art published research recommend coloring journal bought
29	journey vacation luck easter smoothly rough merry holidays zealand ride

Tabela 3. 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. Contexto de alguns descritores.

Descritores

A Tabela 3 apresenta os termos principais associados aos descritores encontrados pelo RMN quando treinado com o texto de *posts* e comentários de todos os subreddits. Os subreddits são considerados em conjunto, 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 um linguagem em comum e que está intimamente relacionada aos transtornos de saúde mental (veja, por exemplo, o descritor 3). Os resultados apresentados a seguir focam na análise dos descritores que estão em negrito na Tabela 3. A Tabela 4 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*, deno-

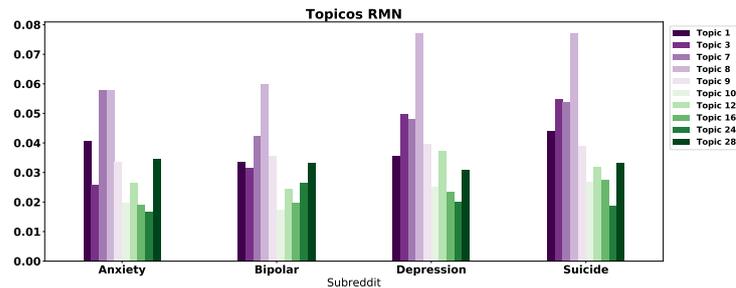


Figura 5. Probabilidade de cada post ou comentário ser associado a um tópico.

minados *spans* pelo RMN. Ao final da execução, o RMN associa a cada *span* 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 5 mostra as probabilidades médias associadas aos tópicos listados na Tabela 4 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 Suicide. 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 Suicide possuem perfis 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 Suicide, 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 Suicide, 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 este assunto mais frequente na comunidade Suicide e possivelmente refletindo problemas de *bullying* comuns nesta fase da vida.
6. O **Tópico 28** tem a probabilidade parecida de ocorrência nas quatro comunidades, revelando que usuários tendem a indicar atividades que possam ocupar a mente (por exemplo, ler um livro).

Analisamos também os *posts* e comentários feitos pelos usuários, auxiliando em uma possível definição de perfis nas comunidades. Para isto, calculamos a média dos vetores associados aos *spans* de cada de usuário. Para cada tópico da Tabela 4, descrevemos a seguir o perfil dos cinco usuários cujas probabilidades associadas 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 probabilidade média de relevância neste tópico, quatro fazem parte da comunidade Suicide 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 Suicide 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], onde os autores concluem 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 Suicide 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*.

5. Conclusão e Trabalhos Futuros

O presente trabalho apresentou uma análise da árvore de discussão dos subreddits: Anxiety, Bipolar, Depression e Suicide. Analisando as 16 árvores de discussão com maiores profundidades percebe-se que a maior parte dos *posts* são pedidos de ajuda e, que na maior parte das vezes, mais de uma pessoa oferece ajuda. As árvores mais profundas, em sua maioria, são compostas por dois usuários (o que fez o *posts* e o que ofereceu ajuda).

Através do modelo RMN foram gerados tópicos que descrevem aquilo que é frequentemente discutido nas quatro comunidades. Tópicos relacionados a pensamentos suicidas são mais frequentes na comunidade de Suicide e Depression. Além disso, constatou-se que os subreddits Depression e Suicide são parecidos em relação a maior parte dos tópicos. Também foi possível analisar os usuários mais relevantes para cada tópico. Por

exemplo, o tópicos 7, que está relacionado à superação de problemas, tem os usuários mais relevantes nas comunidades Bipolar, Suicide e Anxiety.

Como trabalho futuro pretendemos explorar as trajetórias dos usuários em termos dos descritores relacionados às suas publicações. Além disso, pretendemos analisar as características de sucesso em uma discussão.

Referências

- Althoff, T., Clark, K., and 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.
- Choudhury, M. D. and De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *International Conference on Web and Social Media*.
- Gkotsis, G., Oellrich, A., Hubbard, T., Dobson, R., Liakata, M., Velupillai, S., and Dutta, R. (2016). The language of mental health problems in social media. In *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., and Dutta, R. (2017). Characterisation of mental health conditions in social media using informed deep learning. *Scientific reports*, 7:45141.
- Iyyer, M., Guha, A., Chaturvedi, S., Boyd-Graber, J. L., and III, H. D. (2016). Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *HLT-NAACL*, pages 1534–1544.
- Kavuluru, R., Ramos-Morales, M., Holaday, T., Williams, A. G., Haye, L., and Cerel, J. (2016). Classification of helpful comments on online suicide watch forums. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, New York, USA.
- Lopes, C. R. S., Cunha, M., Rodrigues, A., Vilela, A. B. A., Casotti, C. A., and Pereira, H. (2014). Identificando as representações sociais sobre promoção da saúde em uma rede social de trabalhadores de saúde. In *Proceedings of the III Brazilian Workshop on Social Network Analysis and Mining, Brasília-DF Brazil*.
- Pappa, G. L., Cunha, T. O., Bicalho, P. V., Ribeiro, A., Silva, A. P. C., Meira Jr, W., and Beleigoli, 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).
- Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- Souza, B. Á. S., Almeida, T. G. A., Menezes, A. A. M., Figueiredo, C. M. F., Nakamura, F. G. N., and Nakamura, E. F. N. (2017). Uma abordagem para detecção de tópicos relevantes em redes sociais online. In *Proceedings of the VI Brazilian Workshop on Social Network Analysis and Mining, São Paulo, SP, Brazil*.
- Wang, A., Hamilton, W. L., and Leskovec, J. (2016). Learning linguistic descriptors of user roles in online communities. In *Proceedings of the First Workshop on NLP and Computational Social Science*, pages 76–85.