



MONITORAMENTO E ANÁLISE DE *TWEETS* RELACIONADOS À ANSIEDADE E À DEPRESSÃO

German Sack
Juliana Campos
Luciana Lacerda
Marco Milanez
Vitor Mendonça



Especialização em Big Data – FATEC

Rio de Janeiro | outubro de 2020

1.

CONTEXTO

Motivação do problema pesquisado

2.

OBJETIVO

3.

METODOLOGIA

4.

MÉTRICAS

5.

ANÁLISE DOS RESULTADOS

6.

CONCLUSÃO

AGENDA

CONTEXTO

- A partir do cenário de isolamento social como consequência do coronavírus, estudos constaram impactos não apenas econômicos, mas também na saúde mental dos indivíduos.
- No Brasil, esse cenário potencializou o nível de ansiedade e depressão das pessoas como apurado pela LaNCE/UERJ.
- Aumento do uso das redes sociais com o objetivo de amenizar o isolamento social.
- Popularização do *Text Mining* devido ao crescimento da internet e do conteúdo gerado pelos usuários.

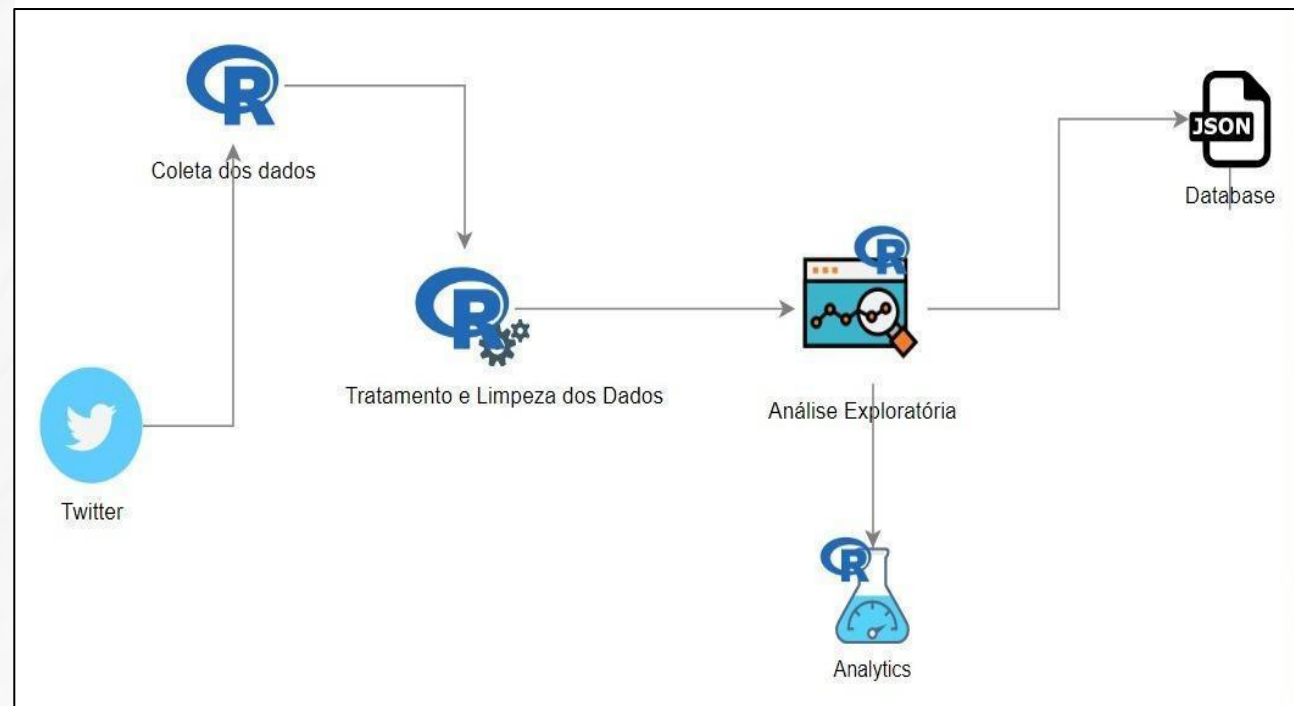
OBJETIVO

- Monitorar os tweets que continham as palavras “ansiedade” e/ou “depressão”.
- Realizar análise de sentimentos para identificar as polaridades das postagens.
- Verificar a existência de padrões nas postagens e construir a evolução das publicações ao longo dos dias coletados.
- Fazer uma análise semântica para melhor compreender os contextos dos tweets.

METODOLOGIA

- A coleta foi realizada tendo como base os *tweets* que contivessem pelo menos um dos termos a seguir: Ansiedade e/ou Depressão.
- Utilização da API do Twitter em conjunto com o software R.
- Período de coleta: **27 de julho a 18 de agosto de 2020.**

Diagrama de Arquitetura do Trabalho



Fonte: Elaboração própria.

METODOLOGIA

- É necessário uma etapa de pré-processamento, que vai estruturar os textos que, até então, eram não estruturados.
- A etapa de indexação e normalização envolve a remoção de *stopwords*, pontuação, numeração, links, espaços desnecessários, emojis, caracteres especiais, quebra de linhas, tweets duplicados, etc.
- Necessidade de criação de dicionários aderentes ao cenário analisado para melhor ajuste da análise de sentimentos.

Etapas do processo de Mineração de dados



Fonte: Adaptado de Aranha e Passos (2006).

MÉTRICAS

■ Contagem de tweets: **567.382**

■ Percentual de tweets por polaridade:

Positivo: 7%
Neutro: 12%
Negativo: 81%

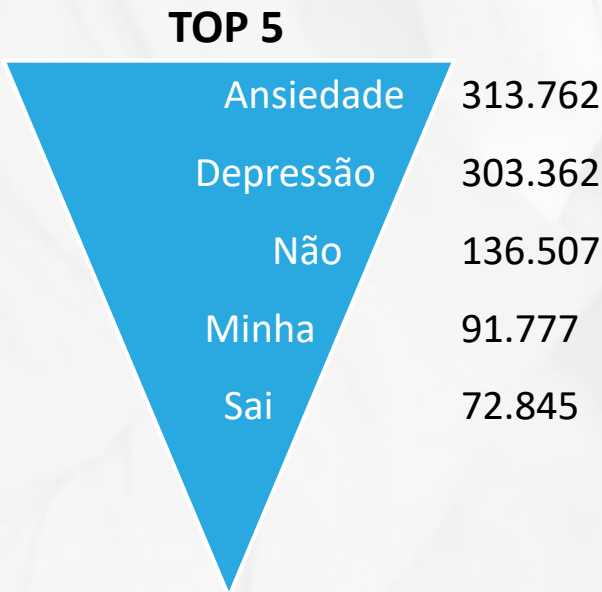
■ Contagem de tweets por polaridade e por dia:

Dia da semana	Negativo	Neutro	Positivo	Total
Domingo	54.260	12.974	5.133	72.367
Segunda	74.152	18.040	7.426	99.618
Terça	70.978	17.991	6.978	95.947
Quarta	59.056	14.638	5.797	79.491
Quinta	58.630	13.597	5.500	77.727
Sexta	52.567	12.688	5.176	70.431
Sábado	54.221	12.550	5.030	71.801

Fonte: Elaboração própria a partir dos dados do Twitter.

MÉTRICAS

■ Contagem de palavras:



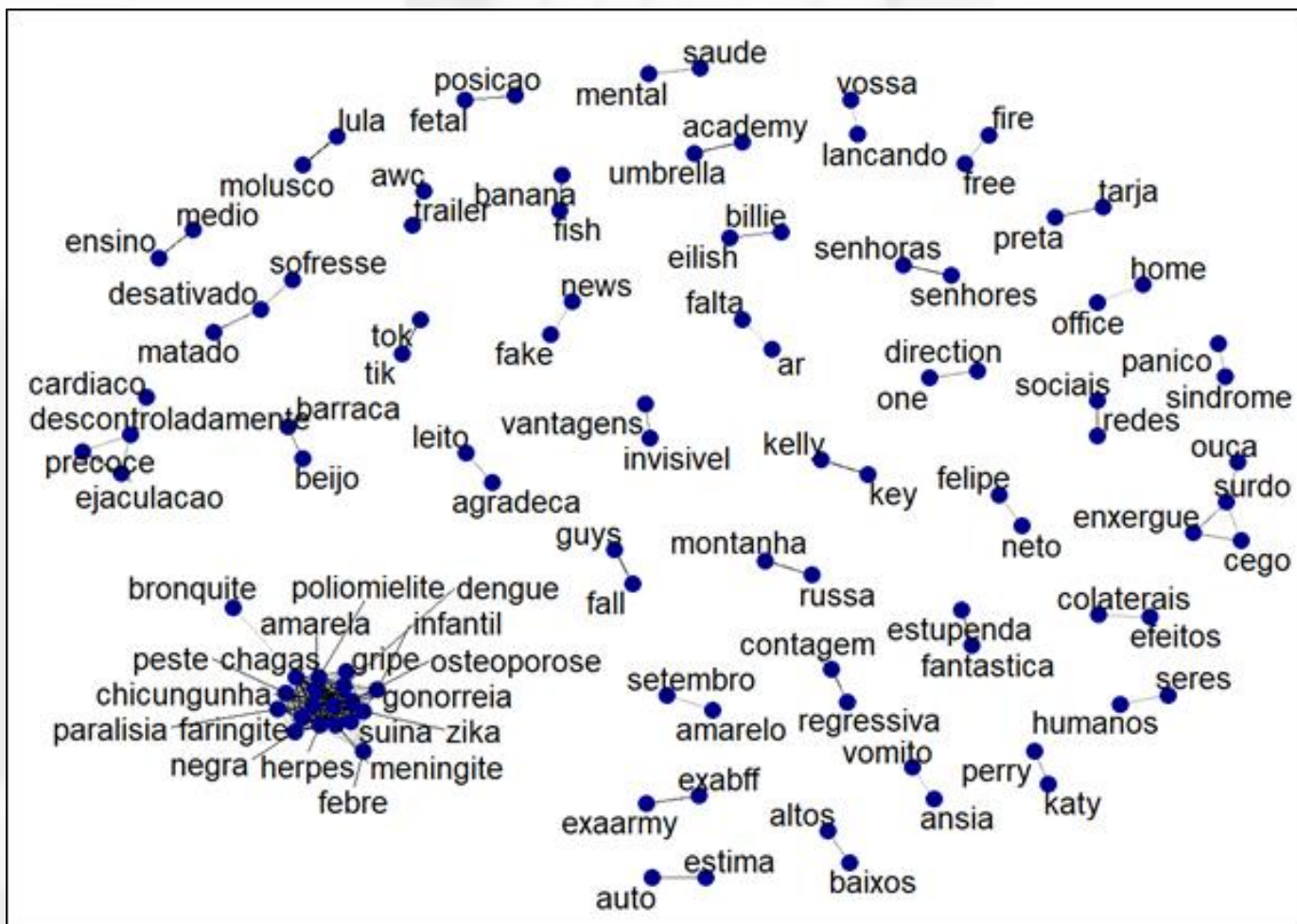
■ Média de *tweets* por dia da semana:

Dia da semana	Média de <i>Tweets</i>
Domingo	24.122
Segunda	24.905
Terça	23.987
Quarta	26.497
Quinta	25.909
Sexta	23.477
Sábado	23.934

Fonte: Elaboração própria a partir dos dados do Twitter.

ANÁLISE DOS RESULTADOS

Diagrama de correlação entre as palavras



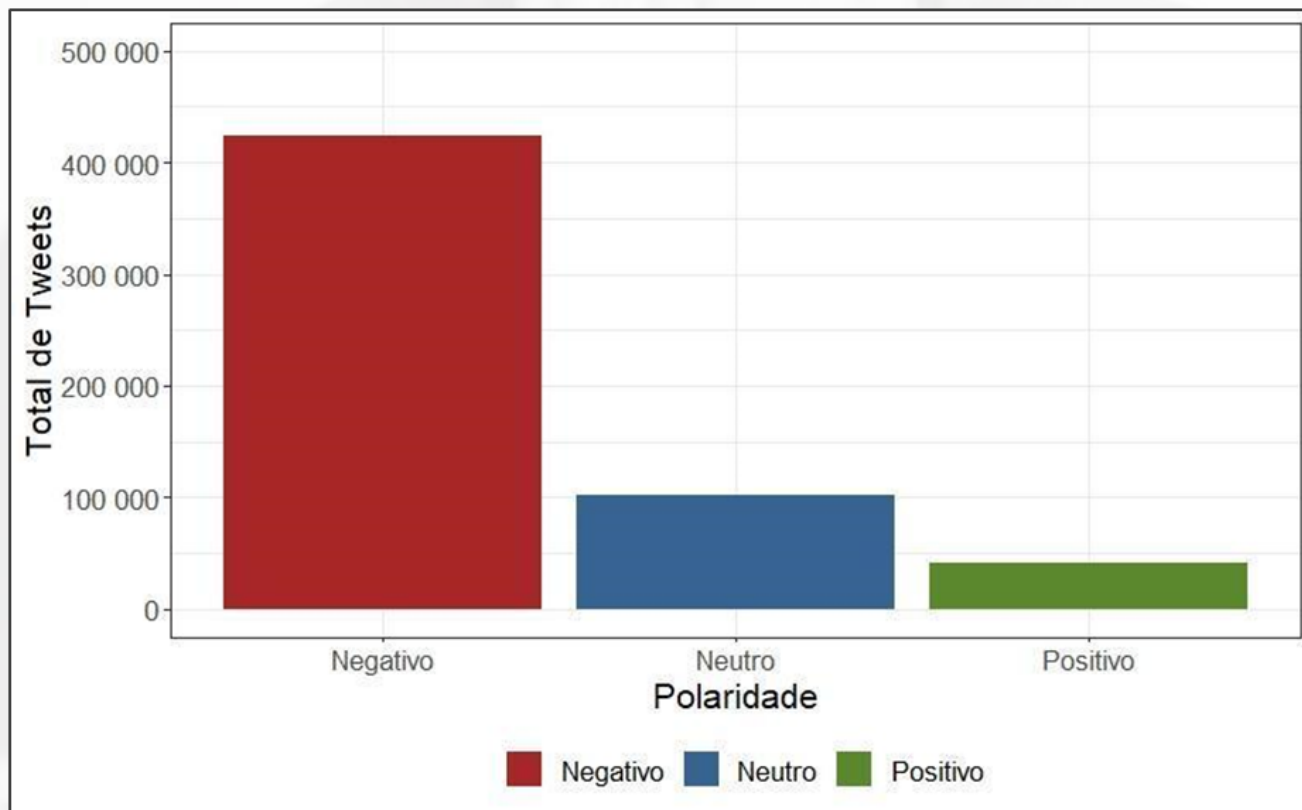
- Correlação entre diversas doenças.
- Associações com algumas personalidades (como o Felipe Neto, Billie Elish, Katy Perry, One Direction) ou entretenimento (Umbrella Academy, Barraca do Beijo).
- Termos como Setembro Amarelo, Saúde mental também apareceram.

Fonte: Elaboração própria a partir dos dados do Twitter.

Nota: correlação foi feita para os termos mencionados mais de 100 vezes.

ANÁLISE DOS RESULTADOS

Análise de Sentimentos – *Tweets* por polaridade

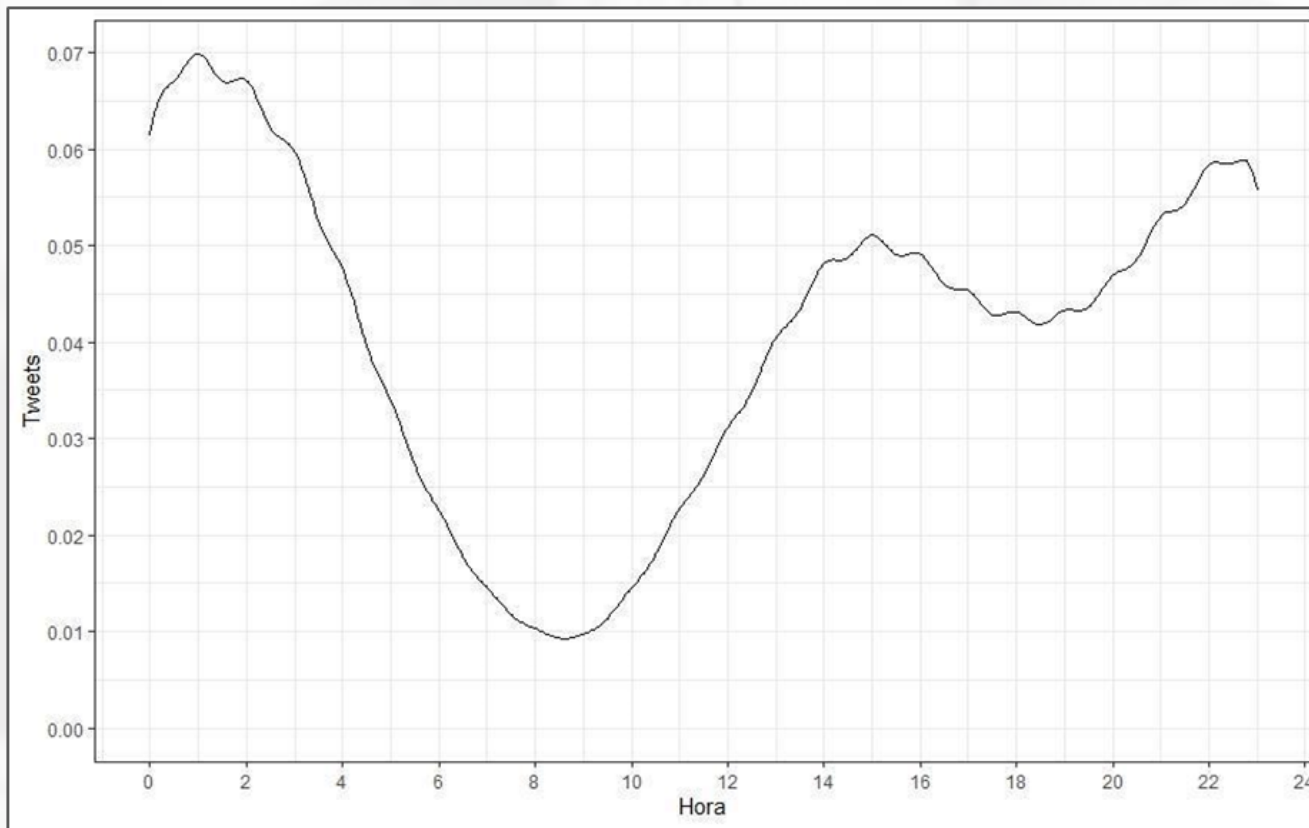


Fonte: Elaboração própria a partir dos dados do Twitter.

- Considerou os dicionários de termos positivos e negativos criados.
- Devido ao contexto analisado, era esperado que *tweets* com polaridade negativa tivessem proporção superior às demais polaridades.

ANÁLISE DOS RESULTADOS

Análise de Sentimentos – Densidade de Tweets por hora



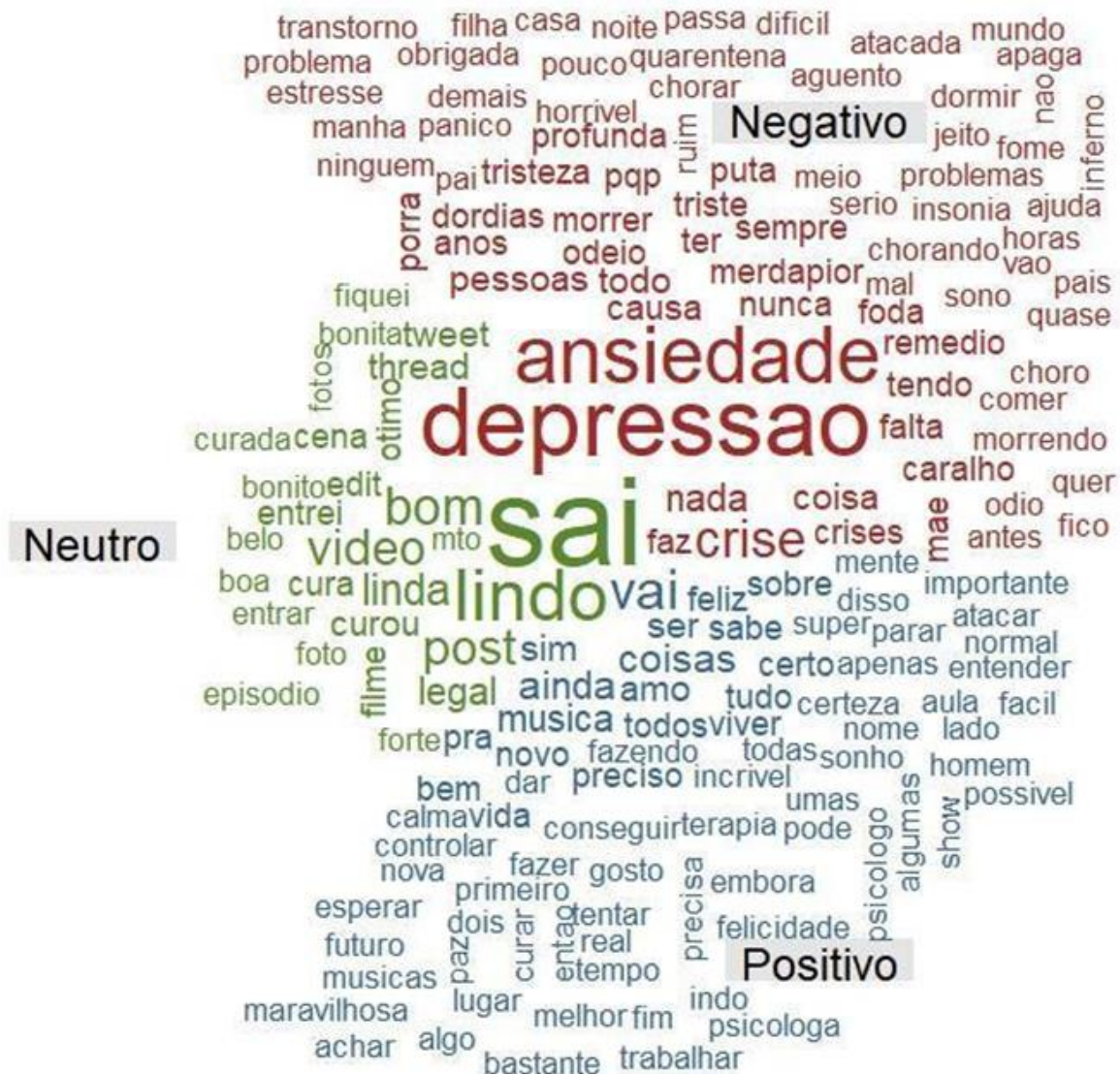
- A curva de densidade permite uma análise de performance do processo de coleta.
- Pico das postagens entre meia noite e duas horas da manhã.
- Ponto de atenção: volume de tweets às 3 horas é semelhante ao das 22h.

Fonte: Elaboração própria a partir dos dados do Twitter.

“Ansiedade” e “Depressão” foram as palavras mais frequentes nos tweets classificados como negativos.

Relevância de outras palavras associadas ao universo analisado como, por exemplo, “stress”, “quarentena”, “insônia”, “crise”, “insônia” etc.

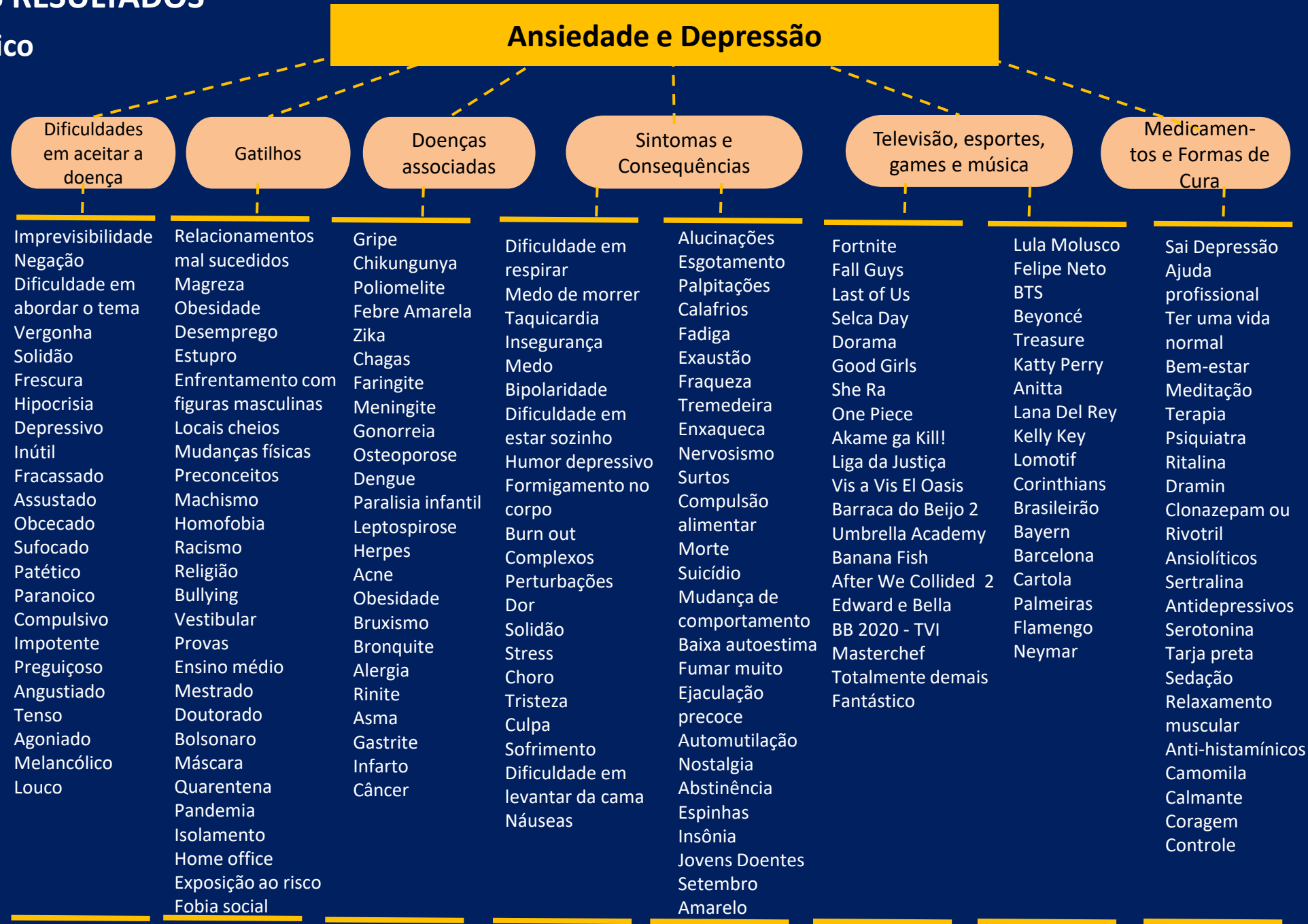
Palavras como “lindo” e “linda” foram classificadas na polaridade Neutra. Observou-se que muitas publicações com esses termos estavam relacionadas a um contexto de ironia.



Fonte: Elaboração própria a partir dos dados do Twitter.

ANÁLISE DOS RESULTADOS

Mapa Semântico



ANÁLISE DOS RESULTADOS

Em resposta a

Eu tenho crise de ansiedade e **graças a Pandemia comecei a desenvolver** depressão e transtorno do Pânico



· 3 de ago

Contei via WhatsApp pra minha mãe que tenho ansiedade, transtorno alimentar e baixa autoestima, não consigo parar de pensar no que ela vai falar quando acordar...



2 de ago

Quando você já nasce ansioso **#ansiedade #ansioso #anxiety**



4

· 3 de ago

· 10 de ago

Toda vez que eu tenho uma crise de ansiedade eu desenho, pra vê se consigo me reencontrar. Pra mim a ansiedade é como um demônio que te atormenta com a preocupação de um futuro, mas ele não deixa seu passado, ele usa isso para alavancar mais ainda sua preocupação. **#ansiedade**



8

· 28 de jul

Mas nera nada demais, só que a ansiedade não ajudou, quem disse que consegui me concentrar para estudar?



1

... · 14 de ago

Jogadores viraram viúvas de treinador, depois que o JJ saiu tem um querendo sair outro em **depressão** outros engordando, outros querendo ganhar salário milionário ser ter futebol para justificar ganhar tanto. Enquanto isso o **Flamengo** fica em segundo, terceiro, quarto plano.



2

· 3 de ago

Pandemia+transtorno de ansiedade+morar no Brasil= crise quase todo dia.



8

vou ver crepúsculo pra melhorar ansiedade bjs



· 3 de ago

Pais fazem isso

Filhos:

- criam basicamente uma dupla personalidade (uma pra qnd tá em casa, outra qnd tá na rua)
- **aprendem a mentir habilmente**
- criam transtorno de ansiedade
- ficam com mania de perseguição
- vêem os pais como inimigos
- confiam mais em amigos/professores



1

CONCLUSÃO

- A escolha dos termos “ansiedade” e “depressão” ocorreu devido aos desdobramentos que a pandemia do novo Coronavírus acarretou diretamente a saúde mental.

- Foi muito importante desenvolver um dicionário aderente ao tema pesquisado, para que a classificação das polaridades esteja adequada à realidade.

- Acho que vale falar desse outro contexto de ansiedade na conclusão do PPT tb: No período de coleta, os usuários publicaram mais quarta e quinta feira, o que refuta a hipótese de que aos finais de semana as pessoas se sentem mais solitárias e publicariam mais. No entanto, reforça um outro contexto de ansiedade: expectativa por competições esportivas, que ocorrem nestes dias da semana.

- O pico das publicações durante a madrugada dá indícios para uma das consequências da ansiedade e da depressão: a insônia.

- A nuvem de palavras permitiu observar que aspectos relacionados ao tratamento da ansiedade ou da depressão aparecem de maneira recorrente.

- A análise semântica evidenciou que os usuários também demonstram estarem ansiosos por estímulos não diretamente relacionados ao coronavírus e também por fatos não atrelados a doença, como ansiedade por uma atividade de entretenimento.

- Gatilhos relacionados ao contexto mundial como, desemprego, instabilidade política e isolamento social também aparecem, convergindo, pela ótica da mineração de textos, com os estudos consultados.