



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS POLÍTICAS E JURÍDICAS
ESCOLA DE ADMINISTRAÇÃO

FABIO LUIS DE SOUZA CARVALHO SOARES

Análise textual de discursos de criminalidade através do Twitter de comunidades virtuais, Polícia Militar, Civil e Guarda Municipal do Rio de Janeiro.

FABIO LUIS DE SOUZA CARVALHO SOARES

Análise textual de discursos de criminalidade através do Twitter de comunidades virtuais, Polícia Militar, Civil e Guarda Municipal do Rio de Janeiro.

Trabalho de Conclusão de Curso de graduação apresentado ao Instituto da Universidade Federal do Estado do Rio de Janeiro, como requisito parcial para a obtenção do grau de Bacharel em Administração Pública.

Orientador: Steven Dutt-Ross

RIO DE JANEIRO, RJ – BRASIL
2022

D118 DE SOUZA CARVALHO SOARES, FABIO LUIS
Análise textual de discursos de criminalidade
através do Twitter de comunidades virtuais, Polícia
Militar, Civil e Guarda Municipal do Rio de
Janeiro. / FABIO LUIS DE SOUZA CARVALHO SOARES. --
Rio de Janeiro, 2022.

Orientador: Steven Dutt-Ross.
Trabalho de Conclusão de Curso (Graduação) -
Universidade Federal do Estado do Rio de Janeiro,
Graduação em Administração Pública, 2022.

1. segurança publica. 2. estatística. 3.
criminalidade. 4. violência. 5. análise textual. I.
Dutt-Ross, Steven, orient. II. Título.

FABIO LUIS DE SOUZA CARVALHO SOARES

Análise textual de discursos de criminalidade através do Twitter de comunidades virtuais, Polícia Militar, Civil e Guarda Municipal do Rio de Janeiro.

Trabalho de Conclusão de Curso de graduação apresentado ao Instituto da Universidade Federal do Estado do Rio de Janeiro, como requisito parcial para a obtenção do grau de Bacharel em Administração Pública.

Aprovado em _____ de _____ 2022.

BANCA EXAMINADORA

Prof. Dr. Steven Dutt-Ross (Orientador)
Universidade Federal do Estado do Rio de Janeiro

Prof. Dr. José Geraldo Pereira Barbosa (Avaliador)
Universidade Federal do Estado do Rio de Janeiro

Prof. Dr. Vinicius Pinheiro Israel (Avaliador)
Universidade Federal do Estado do Rio de Janeiro

*A Eliane, Carlos e Pedro, que apesar das
dificuldades sempre me guiaram pelos caminhos
corretos.*

AGRADECIMENTOS

Agradeço aos meus queridos e amados mãe, pai e irmão, por todo amor e dedicação que tiveram por mim, por sempre me guiarem por caminhos honestos e justos e por fazerem de tudo para me proporcionar as melhores condições possíveis dentro de nossas realidades.

Agradeço à minha melhor amiga e companheira de vida Lany, por ter sempre me apoiado na escolha de cursar o que eu realmente gostava, que sempre me dá suporte em momentos bons e ruins, pela paciência e por todo amor.

Agradeço por toda minha família, em especial Claudia, Daniele, Joana, Lucas e Valdeir, que estão sempre presentes e torcendo para o meu futuro.

Agradeço aos amigos que fiz na UNIRIO, especialmente ao Antônio, Alexandre, Gustavo, Gama, João, Karen, Marcela, Marcella, Mariana e Pinho, por terem feito parte dessa experiência e a tornarem ainda mais agradável.

Agradeço a todos os mestres que encontrei pelo caminho, desde o ensino básico até o superior, que através do conhecimento ajudaram a moldar a pessoa que sou hoje.

Agradeço ao meu orientador Steven Dutt-Ross, por todo o apoio, paciência e dedicação no desenvolvimento deste trabalho.

[...] não sobrecarreguem seus corações pensando no melhor caminho. Pode ser que as trilhas nas quais cada um de vocês deve pisar já estejam diante de seus pés, embora talvez não consigam enxergá-las.

(J. R. R. Tolkien)

RESUMO

A violência urbana é constante na vida de muitas pessoas, principalmente na vida da população do estado do Rio de Janeiro. O avanço constante de grupos armados por todo o território, a omissão do Estado e fatores históricos, sociais e econômicos foram fundamentais para a expansão do crime no Rio de Janeiro. Dito isto, a população está em constante alerta, sobretudo pelas contantes notícias vinculadas nas mídias digitais e sociais e televisão. Mesmo com todo o apelo existente em cima dessa violência, o poder público segue ineficaz, ausente e que pouco se comunica com a população. Com isso, foram criadas comunidades virtuais no Twitter, com intuito de alertar as pessoas sobre a violência existente em diversos pontos do estado. O objetivo principal deste trabalho é analisar dados dessas comunidades virtuais e entender quais são os indicadores de criminalidade mais relevantes abordados pela população. Além disso, procura-se analisar a comunicação de órgãos de segurança pública do estado através do Twitter e entender os principais pontos abordados por estes e suas semelhanças e diferenças de comunicação em relação as comunidades virtuais. Os resultados da pesquisa indicam que os perfis das Comunidades Virtuais no Twitter visam alertar a população e indicar os locais onde há ocorrências de crimes e violência, enquanto os perfis dos órgãos de segurança pública se comunicam de forma mais institucionalizada. A análise também indica que os temas mais relevantes abordados pelas comunidades virtuais possuem relação direta com dados de criminalidade do Estado do Rio de Janeiro.

Palavras-chave: Twitter. Violência. Análise Textual. Segurança Pública. Comunidades Virtuais.

ABSTRACT

The urban violence is constant in the lives of many people, especially in the lives of the population of the state of Rio de Janeiro. The expansion of armed groups throughout the territory, the omission of the State and historical, social and economic factors were fundamental for the expansion of crime in Rio de Janeiro. The population is on constant alert, especially due to the constant news linked in digital and social media and television. Even with all the appeal that exists on top of this violence, the public power remains ineffective, absent and have a poor communication with the people. With this, virtual communities were created on Twitter with the aim of alerting people about the violence existing in every part of the state. The main objective of this work is to analyze data from these virtual communities and understand which are the most relevant crime indicators addressed by the population. In addition, it seeks to analyze the communication of public security agencies in the state through Twitter and to understand the main points addressed by them and their similarities and differences in communication in relation to virtual communities. The research results indicate that the profiles of Virtual Communities on Twitter aim to alert the population and point out the places where crimes and violence occur, while the profiles of public security agencies communicate in a more institutionalized way. The results also indicate that the most relevant topics addressed by the virtual communities are directly related to crime data from the State of Rio de Janeiro.

Palavras-chave: Twitter. Violence. Textual Analysis. Public security. Virtual Communities.

LISTA DE FIGURAS

Figura 1 - Exemplo de um tweet existente na base de dados	35
Figura 2 - Nuvem de palavras das quatro bases de dados	42
Figura 3 - Monitoramento de criminalidade da Região Metropolitana e Norte Fluminense do Rio de Janeiro	46

LISTA DE GRÁFICOS

Gráfico 1--Tipo de dispositivo utilizado para acesso à internet	27
Gráfico 2 - Histogramas das bases utilizadas	38
Gráfico 3 - Termos mais frequentes - TF.....	40
Gráfico 4 - Termos com maior TF-IDF	44

LISTA DE TABELAS

Tabela 1 - Quantidade de termos existentes em cada documento	37
Tabela 2 - Frequência dos termos das quatro bases utilizadas	39
Tabela 3 - Função tf-idf nos termos mais utilizados	43
Tabela 4 - Palavras chaves utilizadas na busca da primeira base de dados	55
Tabela 5 – Lista de Stopwords	56

LISTA DE QUADROS

Quadro 1 - Exemplos de tweets que continham os termos pré-selecionados, mas que não eram relacionados com assuntos referentes a crimes, ocorrências, etc.	31
Quadro 2 - Exemplos de dados da base Comunidades Virtuais	59
Quadro 3 - Exemplos de dados das bases das polícias militar, civil e guarda municipal	60

SUMÁRIO

1 INTRODUÇÃO	15
1.2 OBJETIVO DE PESQUISA	17
1.3 OBJETIVO SECUNDÁRIO	17
2 REVISÃO DE LITERATURA	18
2.1 VIOLÊNCIA E CRIMINALIDADE	18
2.2 A SUBNOTIFICAÇÃO DE CRIMES E A PESQUISA DE VITIMIZAÇÃO.....	19
2.3 A VIOLÊNCIA URBANA	21
2.4 A MÍDIA E A PERCEPÇÃO DA VIOLÊNCIA.....	23
2.5 O TWITTER.....	25
2.6 COMUNIDADES VIRTUAIS	26
3 MÉTODO	29
3.1 MINERAÇÃO TEXTUAL.....	29
3.2 CORPUS	29
3.3 COLETA DE DADOS	30
3.3.1 <i>base de dados</i>	30
3.3.2 <i>Seleção da Base de Dados</i>	30
3.3.3 <i>Pré-processamento de dados</i>	33
3.3.3.1 <i>Remoção de pontuação, caracteres especiais e ajuste do texto</i>	33
3.3.3.2 <i>Tokenização</i>	34
3.3.3.3 <i>Stopwords</i>	34
3.4 TF-IDF	35
3.5 NUVEM DE PALAVRAS.....	36
4 ANÁLISE DE RESULTADOS	37
5 LIMITAÇÕES, CONCLUSÃO E TRABALHOS FUTUROS	48
6 REFERÊNCIAS	52
ANEXOS	55
ANEXO A – PALAVRAS UTILIZADAS PARA BUSCA DA BASE PRIMÁRIA	55
ANEXO B – LISTA DE STOPWORDS	56
ANEXO C – EXEMPLOS DE DADOS DAS BASES ANALISADAS	59

1 INTRODUÇÃO

O crime organizado vem expandindo cada vez mais seus territórios e sua força pela cidade do Rio de Janeiro. Segundo um estudo de uma parceria entre Instituto Fogo Cruzado, o Disque Denúncia, GENI-UFF, NEV-USPIBGE e a plataforma digital Pista News (2020), a milícia e o tráfico [SR1] estão presentes em 96 dos 163 bairros do município, contemplando cerca de 3,76 milhões de habitantes que vivem nessas regiões. O aumento da violência urbana é noticiado constantemente em todas as mídias, assim como roubos, crimes violentos e constantes confrontos entre policiais e bandidos.

Um dos principais instrumentos utilizados para o registro e denúncia de crimes é o boletim de ocorrência, responsável pela maior fonte de dados utilizados como forma de medir a criminalidade. Porém, é de conhecimento geral que muitos crimes e delitos ocorridos acabam não sendo reconhecidos pela polícia, visto que muitas das vítimas de crimes optam por não fazer a denúncia e nem o registro de ocorrência. Murray et al (2013) relata em seu estudo que apenas 48% das vítimas de furto e 38% das vítimas de roubo relataram o incidente à polícia no Brasil em 2009. Já outra pesquisa de vitimização feita em Belo Horizonte, aponta que apenas 29% das pessoas entrevistadas que foram vítimas de furto acionaram a polícia, 27,4% entre vítimas de roubo, 25,7% de vítimas de agressão física e 14,5% de vítimas de agressão sexual (CRISP, 2002).

Disto isto, entende-se que a subnotificação é um problema recorrente no que diz respeito às pesquisas que tentam mensurar a criminalidade. Considerando que muitos dos crimes não são notificados à polícia, os estudos oficiais acabam sofrendo críticas pois acabam não retratando a criminalidade que realmente é praticada. Isso também acaba gerando outros problemas de análise de dados oficiais, pois a relação entre a taxa de criminalidade real e as taxas divulgadas pelas instituições policiais é medida por elementos como o nível de confiança da população na polícia, a capacidade da polícia de processar os registros e até mesmo os conceitos associados a cada tipo de delito (HAGEN, 2008). Isso significa que aumentos de crimes em dados oficiais não necessariamente significam que esse tipo de crime aumentou, mas

também pode significar que a percepção da sociedade sobre um tipo de crime mudou e estimulou as vítimas a denunciarem o crime, diminuindo a subnotificação e consequentemente aumentando os números oficiais.

Nos últimos anos, os órgãos de segurança pública criaram diversas ferramentas de denúncias e atendimentos online que visavam facilitar o acesso do cidadão aos serviços policiais, diminuir o tempo de resposta dos mesmos e agilizar os atendimentos. Esses órgãos também estão presentes em redes sociais, visando se aproximar cada vez mais da população e aumentar sua confiabilidade e reconhecimento.

As redes sociais são estruturas que atuam em diversos níveis e facilitam o relacionamento entre pessoas e o compartilhamento de informações de forma veloz e prática. Além de serem ferramentas que conectam as pessoas, as mídias sociais também podem ser utilizadas como uma grande fonte de informações.

O Twitter, por exemplo, é uma das mais utilizadas redes sociais do mundo e de acordo com o relatório financeiro disponibilizado pela empresa em 28 de abril de 2022, a rede social possui cerca de 214 milhões de usuários ativos diariamente. Essa grande quantidade de usuários gera um grande volume de dados que podem servir como base para diversos tipos de pesquisas.

Com a violência crescente na região metropolitana do Rio de Janeiro e a sensação de abandono dos órgãos de segurança pública, os cidadãos utilizam cada vez mais as redes sociais para expor, compartilhar e denunciar os problemas da cidade, com o objetivo de espalhar as informações para o maior número possível de pessoas e de maneira rápida. Esse movimento foi crescendo principalmente no Twitter, onde até foram criadas comunidades virtuais que possuem como objetivo alertar e informar a população em tempo real sobre problemas que estão ocorrendo nas mais diversas áreas da cidade, sobretudo compartilhando informações sobre tiroteios, assaltos e crimes em geral que acontecem na cidade.

Órgãos como a Polícia Militar do Estado do Rio de Janeiro (PMERJ), Polícia Civil do Estado do Rio de Janeiro (PCERJ) e a Guarda Municipal do Rio de Janeiro (GMRio) possuem contas oficiais no Twitter e visam se comunicar melhor com a população e apresentar melhor seus trabalhos e ações do dia a dia. Todavia, essa

comunicação aparentemente não é eficaz pois possui um tom mais institucional e padrão, destoando completamente da forma de como os cidadãos se organizam e comunicam através da mesma rede e conseqüentemente acarretando pouco engajamento e interações entre a população e os órgãos de segurança pública nas redes.

1.2 OBJETIVO DE PESQUISA

O objetivo principal deste trabalho é analisar textualmente os tweets relacionados à violência e criminalidade feitos por comunidades virtuais e compará-los com os tweets feitos pelas principais polícias do Estado do Rio de Janeiro (Polícia Militar, Civil e Guarda Municipal).

Através de ferramentas e análises estatísticas, podemos entender melhor os temas mais relevantes e abordados pelas comunidades virtuais no Twitter, sobretudo na região suburbana do estado do Rio de Janeiro, no que diz respeito à violência e criminalidade. A análise desses dados visa também entender um pouco melhor as semelhanças e diferenças na comunicação feita através dos tweets feitos pelas comunidades virtuais e polícias do estado do Rio de Janeiro.

1.3 OBJETIVO SECUNDÁRIO

Como objetivo secundário deste trabalho, entender se os dados gerados e analisados podem ser aproveitados como indicadores nos setores de inteligência das polícias do estado, servindo como base para análises de criminalidade e políticas de combate à violência e criminalidade nas mais diversas áreas urbanas do estado do Rio de Janeiro.

2 REVISÃO DE LITERATURA

2.1 VIOLÊNCIA E CRIMINALIDADE

A violência é inerente ao ser humano e está presente desde tempos antigos na nossa sociedade. O conceito de violência é bastante complexo e existem diversas teorias filosóficas e sociológicas que tentam explicar esse assunto. Até mesmos dicionários possuem diferentes descrições sobre o significado da palavra violência. Para o Novo dicionário Etimológico de Língua Portuguesa, de Rodrigo Fontinha (s/d), violência significa:

Violência. s.f (Lat. Violentia (m)) - Qualidade de violento; ação ou resultado da ação de violentar ou violentar-se; ímpeto veemente nas ações; força com que se pretende obrigar uma pessoa a fazer aquilo que ela não quer; ato violento; abuso da força e do poder; opressão; grande impulso; veemência, impetuosidade; coação; constrangimento (FONTINHA, s/d).

Existem diversos tipos de violência, desde a violência física até a violência psicológica. A palavra violência em latim é violentia, que significa a ação de violar o outro ou se violar. Esse termo também indica algo fora do natural do comportamento humano, ligado diretamente ao uso da força e que gera danos físicos, mas que também de outros pontos de vista como o filosófico pode tolher a liberdade a liberdade e/ou a vontade de outro indivíduo.

[...] Uma, de caráter mais psicológico, é que elas são assim denominadas, com frequência, muito mais pelo impacto emocional que produzem no imaginário das pessoas do que por razões objetivas consistentes. Outra razão, de caráter mais filosófico, é a dificuldade de encontrar um princípio racional que explique essas ações, particularmente sob o impacto emocional dos seus efeitos. Outra, de caráter mais antropológico, é que a qualificação das ações como violentas permite desqualificar seus autores, tornando-os a expressão máxima da desumanidade, rebaixando-os, equivocadamente, ao nível da animalidade, mundo onde não há lugar para a violência por não existir nele liberdade, intencionalidade, nem consciência, todas elas características da condição humana dos homens (PINO, 2007, v. 28, p. 763).

Todavia, nem todo ato violento configura um crime, assim como nem todo crime se configura como uma violência. O crime, segundo Durkheim (2004, p.97), consiste num ato que ofende certos sentimentos coletivos dotados de uma energia e de uma clareza particulares. Isso significa que o crime está diretamente ligado à sociedade, ou seja, são os fatos sociais que definem o que se enquadra em crime ou não. Podemos tomar como exemplo disso o fato de que um ato pode se configurar

crime em algum país e em outro não e possuírem diferentes formas de sanções, penas ou até mesmo a ausência de consequências de tal ato.

Dito isto, entende-se que violência e crime possuem certo grau de relação, porém são termos bastante distintos entre si e que possuem um amplo e complexo significado, dependendo do contexto social em que está inserido. No exemplo do Brasil, existem centenas de artigos em sua Constituição Federal que tipificam os tipos de crime no país para que seja possível haver uma aplicação legal da lei, pois a própria Carta Magna cita que não existe crime e nem pena sem uma lei anterior que defina esse crime, conforme é dito no Art. 1º da Lei nº 7.209, de 11 de julho de 1984 (BRASIL, 1984).

2.2 A SUBNOTIFICAÇÃO DE CRIMES E A PESQUISA DE VITIMIZAÇÃO

A escassez de dados e a confiabilidade dos registros oficiais de crimes são constantemente contestados pelos pesquisadores. A dificuldade em coletar dados consistentes prejudica o trabalho com dados oficiais, visto as complexidades existentes no processo de registro de ocorrência. Segundo Muniz (2000), o boletim de ocorrência é a primeira “entrevista com o crime”, ou seja, o primeiro recurso de notificação de um delito, precedendo, inclusive, as atividades policiais de verificação, qualificação e fundamentação das informações obtidas inicialmente.

No município do Rio de Janeiro, a polícia militar faz o registro de suas ocorrências nos chamados B.O. (boletim de ocorrência). O procedimento da polícia militar é repassar para a polícia civil todos os registros de boletins em que se configuram crime, para que haja investigação. Porém, grande parte das notificações não são constituídas como crime e, por consequência, não são enviadas para a polícia civil. Um dos principais instrumentos utilizados pela polícia civil é o registro de ocorrência (R.O), porém, segundo CANO (2000), os registros de ocorrência não abrangem o conjunto total de crimes e, portanto, não podem ser usados como base de um sistema de informação criminal.

Para se registrar um crime, ele deve ser identificado e notificado à polícia e registrado em seu boletim de ocorrência. Acontece que, por existirem diferentes tipos de padrões de classificação de delitos (motivados por interesses institucionais,

interesses pessoais dos agentes públicos de segurança, por falta de credibilidade da polícia, desprezo econômico ou moral ao crime), gera-se uma grande discrepância entre o número de crimes divulgados por dados oficiais com o número de crimes efetivamente praticados. CATÃO (2008, p27) aponta que muitas pessoas deixam de registrar um delito por motivos como: o medo de represálias, falta de credibilidade na polícia, por não achar importante, por conseguir reparar o problema por outro meio e outros motivos.

Como alternativa para mensuração da criminalidade, surge como alternativa a pesquisa de vitimização, que busca coletar informações junto às vítimas de crimes e reduzir o problema da subnotificação (CATÃO, 2000). Para Campagnac, Borges e Soares (2008), a pesquisa de vitimização visa medir e estimar com maior precisão o crime e o perfil das vítimas, além de medir o impacto do crime na vida das pessoas e permitindo a avaliação da percepção da vítima em relação à efetividade e confiança nas organizações de segurança pública. Surgida na década de 60 nos EUA, a pesquisa de vitimização foi se espalhando por diversos países europeus e, no Brasil, apenas no ano de 1988. Segundo Miranda (2008), cerca de apenas 23 pesquisas de vitimização foram feitas entre os anos de 1988 e 2006 no Brasil, onde poucas dessas pesquisas podem ser comparadas por conta da falta de padronização na metodologia e formas de coleta de dados (MIRANDA, 2008; CATÃO, 2008)

Esse tipo de pesquisa consiste em uma abordagem do tipo survey¹, aplicada em uma amostra populacional, para medir e conhecer aspectos, hábitos e práticas cotidianas da população no que diz respeito à segurança pública e aos sentimentos de vitimização, confiança nas instituições de segurança pública e medo da criminalidade. Segundo LIMA et al (2008), é possível obter variáveis intermediárias entre as vitimizações e o registro policial. Isso significa que a pesquisa de vitimização não possui apenas como objetivo identificar e justificar as diferenças entre dados oficiais e crimes realmente praticados, mas também visa fornecer indicadores que mostrem aspectos ligados à etiologia e ao contexto do fenômeno da criminalidade, que aspectos relacionados ao perfil das vítimas (SENASP, 2013, p.4).

¹ A pesquisa survey é um tipo de investigação quantitativa, que busca coletar dados e informações a partir de características e opiniões de grupos de indivíduos. Caso os grupos sejam representativos da população, os resultados obtidos podem ser extrapolados para todo o universo de estudo.

Conforme afirmado por Zauberman (2008), a pesquisa de vitimização permite avaliar a relação que a vítima possui com as instituições públicas, sua sociabilidade, confiança nos vizinhos, estilo de vida, sociabilidade e política de registro de ocorrência oficial das instituições de segurança. Também é dito por Costa e Lima (2018) que outro fator medido nas pesquisas de vitimização é a confiança das pessoas na polícia, pois é um fator diretamente associado ao registro de ocorrência, uma vez que a falta de confiança pode significar baixa colaboração com a polícia e, conseqüentemente, gerando subnotificação. Alencar e Oliveira (2015) corroboram com essa visão, sugerindo que percepções negativas possuem efeitos significativos na opção de não procurar a polícia após a ocorrência de um crime.

2.3 A VIOLÊNCIA URBANA

A violência urbana é um fenômeno que ocorre em todo o mundo. No Brasil, há raízes sociais e históricas impetradas em nossa sociedade que ajudam a entender um pouco melhor esse problema recorrente. A industrialização na década de 50 ocasionou também no aumento da densidade demográfica nas cidades. A industrialização permitiu a construção de rodovias, ferrovias, construção de indústrias, avanço na medicina, etc., o que intensificou a urbanização em diversas partes do país.

Por conta da grande concentração de pessoas nas principais áreas urbanas do país, nasce o caos urbano, a crise sanitária, a crise habitacional, a multidão e a pobreza (VAZ, 1994, p.44). Para reagir a esta crise urbana, surgem as habitações populares, sobretudo os cortiços, que visavam abrigar um grande número de pessoas. Contudo, essas habitações muitas vezes possuíam situações precárias e insalubres, sendo constantemente vetores de doenças.

Por conta de medidas impostas pelo governo para controle da saúde pública e higiene no século XX, cortiços e bairros inteiros foram demolidos, e com isso, para Abreu (1997, p.67), “grande parte da população foi, então, forçada a morar com outras famílias, a pagar aluguéis altos (devido à diminuição da oferta de habitações) ou a mudar-se para os subúrbios, já que pouquíssimas foram as habitações populares construídas pelo Estado em substituição às que foram destruídas”.

Isso permitiu o nascimento de imensas periferias desprovidas de qualquer infraestrutura urbana ou equipamentos e serviços públicos, que marcou todas as cidades brasileiras, nas décadas seguintes, promovendo um enorme ônus às gerações subsequentes (CARVALHO, 2002, p.49). A desigualdade social, omissão do estado, péssimas condições de vida e má distribuição de renda são refletidas até hoje em nosso país, foram fatores determinantes no aumento da criminalidade urbana.

Perdurando até os dias de hoje, as condições das favelas e periferias no país são em sua maioria espaços com pouca infraestrutura, desorganizados estruturalmente e socialmente, onde a maioria da população é sobretudo negra e possui baixa escolaridade. Além disso, a taxa de desemprego é elevada e muitos dos que trabalham recebem pagamentos mínimos. Somado à omissão do estado nessas áreas, as favelas e periferias tornam-se o ambiente ideal para o surgimento de organizações criminosas, virando o foco do uso da força, da venda de drogas e armas de fogo. Isso acaba afetando não somente aquela área, mas toda a região, conforme essas organizações vão se tornando mais poderosas e influentes. Contudo, é importante dizer que apesar da pobreza e desigualdade contribuírem para a violência urbana, isso não é o que explica a violência urbana em si. Conforme Carrion (2008, p.119, tradução nossa) conclui:

Embora a variável urbana não tenha a condição determinística atribuída pelas correntes etiológicas da violência, não se pode ignorar que a produção social do território é um elemento importante no comportamento de determinados tipos de violência. Existe uma geografia da violência que não é apenas a manifestação dos atos violentos no território (cenário), mas que também é um elemento relevante na sua produção (violência urbana). Esse elemento parte da divisão social do espaço e de um urbanismo particular, que pode produzir um tipo particular de violência de acordo com a segregação urbana (localização residencial, densidade e atividades) e suas implicações sociais (estranheza e medo) [...].²

² No original: Si bien la variable urbana no tiene la condición determinista asignada por las corrientes etiológicas de la violencia, no se puede desconocer que la producción social del territorio sí es un elemento importante en el comportamiento de ciertos tipos de violencias. Existe una geografía de la violencia que no es sólo la manifestación de los hechos violentos en el territorio (escenario) sino también un elemento relevante en la producción de los mismos (violencia urbana). Este elemento viene de la división social del espacio y de una lógica particular de urbanismo que puede producir un tipo de violencia particular acorde a la segregación urbana (localización residencial, densidad y de actividades) y a sus implicaciones sociales (foraneidad y temor) [...]

2.4 A MÍDIA E A PERCEPÇÃO DA VIOLÊNCIA

A violência está presente desde os tempos mais primórdios da sociedade. Existem incontáveis registros antigos de guerras, assassinatos, estupros, roubos, etc, que são estudados e analisados nas escolas do mundo todo.

Nos dias de hoje, a violência sempre está sendo destacada nos principais jornais do país. Notícias de crimes estão sempre sendo destacadas nos programas televisivos, jornais, rádios e sites a todo momento, e muitas vezes sendo tratadas de forma sensacionalista apenas para atrair mais telespectadores/leitores. Para Rolim (2006, p.193), esse sensacionalismo constante em notícias sobre crimes saturam a população, trazendo um senso de ameaça constante e insegurança e conseqüentemente construindo uma chamada “realidade invertida”.

Para Hohlfeldt (1997, p.44), as mídias de comunicação influenciam o telespectador em médio e longo prazo. O constante fluxo de informações com o tempo vai se consolidando na memória e vida das pessoas e, apesar de não conseguir impor ao telespectador uma opinião, consegue influenciar diretamente sobre o que ele fala e pensa. Isso significa que a mídia influencia diretamente nas agendas públicas, sociais e individuais.

Isso reflete claramente no nosso dia a dia, onde podemos ver casos e casos de pessoas que vivem apreensivas e inseguras, pessoas desenvolvendo crises de pânico e com medo de andar nas ruas. A falta de contextualização dos crimes e violências que são divulgados nas grandes mídias também afetam a opinião da população, pois a falta de explicação do contexto daquele crime gera a sensação de que aquilo pode ocorrer em qualquer lugar e com qualquer pessoa. Como exemplo, vemos claramente no Brasil uma grande percepção das pessoas sobre o município do Rio de Janeiro, onde muitos acham que existem tiroteios todos os dias e que tudo está tomado pelas favelas. A violência nessa cidade tem aumentado nos últimos anos, mas por ser transmitido exaustivamente pelos veículos de comunicação em todo o país em esclarecer contextos, a percepção sobre a cidade acaba se tornando irreal.

Contudo, a mídia pode exercer funções que contribuam para a antiviolença. Para Rolim (2006, p.206), uma mídia comprometida é aquela que fiscaliza e é crítica em relação às políticas de segurança pública, casos de corrupção, má utilização de recursos públicos, etc. Rolim (2006. p.206-207) ainda conclui:

Seria, de toda maneira, injusto não assinalar que a mídia tem, em várias oportunidades, assumido partes dessa agenda antiviolença. A luta pelos direitos humanos no Brasil e em tantos outros países não teria alcançado muitas de suas conquistas sem a presença corajosa de jornalistas que realizaram denúncias e revelaram problemas graves antes desconhecidos da maioria da população. O balanço crítico a ser feito sobre as relações da mídia com a violência e o crime, A SÍNDROME DA RAINHA VERMELHA 206 então, não poderá ignorar contribuições tão valiosas que continuam sendo oferecidas. Esse papel contraditório com o contexto das críticas que faço deve mesmo servir para que, ao utilizarmos a expressão “mídia”, tenhamos presente que o conceito não pressupõe uma realidade homogênea e indiferenciada. É bem verdade que boa parte das abordagens equivocadas sobre o tema nos meios de comunicação diz respeito a um padrão jornalístico que costuma ser reproduzido naturalmente sem que isso corresponda, necessariamente, à existência de pressupostos teóricos ou ideológicos mais definidos.

Para Gomis (1991, p.187), a mídia multiplica, generaliza e ativa as dinâmicas sociais, pois a imagem que as mídias transmitem se tornam referências de mudança social, estimulando o público a participar ativamente e colaborar ativamente dos eventos que mais interessam a eles, colaborando com a antiviolença.

Existem diversos exemplos que corroboram com essa análise, como por exemplo, o surgimento de denúncias anônimas e/ou testemunhas em caso que foram para a mídia, postagens em redes sociais denunciando diversos tipos de crimes, violências e irregularidades. Inclusive isso é utilizado como estratégia tanto para a polícia quanto para os canais de comunicação, que veiculam notícias sobre crimes e informam telefones e canais de contato para eventuais denúncias, testemunhas e informações que contribuam para o caso.

Isso significa que os canais de comunicação possuem influência direta na percepção de violência das pessoas. É inegável que as constantes chamadas midiáticas causam sensação de insegurança e terror na população. Mas a mídia também colabora nas medidas antiviolença, influenciando os indivíduos a exercerem sua cidadania e auxiliar, de sua maneira, no combate contra a violência, corrupção, criminalidade, etc.

2.5 O TWITTER

O Twitter é uma rede social que foi fundada em 21 de março de 2006 por Jack Dorsey, Evan Williams e Biz Stone, em São Francisco, Califórnia, Estados Unidos (TWITTER..., 2022). Ela funciona como uma espécie de microblog, onde através da internet as pessoas podem gratuitamente criar perfis, enviar mensagens e receber atualizações em tempo real de outros usuários. É possível se conectar com outros perfis através do botão “seguir”, que funciona como uma assinatura que permite você acompanhar as publicações dessas pessoas “seguidas”. Uma das principais características do Twitter é o limite de caracteres existente para suas publicações, que inicialmente eram de 140 caracteres, porém em novembro de 2017 esse limite aumenta para 280 caracteres. Isso deve-se ao fato de que o Twitter possuía originalmente o conceito de uma plataforma baseada em comunicação via SMS (serviço de mensagens curtas).

Além da possibilidade de publicar mensagens em seu próprio perfil, é possível manter uma conversa com outros usuários através das “replies”, que permite fazer comentários em uma publicação de outro usuário. Outra função que também é bastante utilizada é o “retweet”, que permite ao usuário a publicar em seu perfil algo que outra conta publicou (exibindo a conta que postou originalmente a mensagem).

Sendo uma das redes sociais mais conhecidas do mundo, o Twitter atualmente conta com cerca de 214 milhões de usuários ativos diariamente, segundo relatório fiscal do primeiro trimestre de 2022 (TWITTER, 2022) e nesse período teve uma receita de US \$1,2 bilhão. Atualmente, o Twitter é utilizado pela grande maioria das empresas do mundo como um canal de comunicação entre a companhia e seus consumidores, assim como é utilizado pela maioria dos veículos de comunicação, artistas, políticos, atletas, etc., devido à velocidade que se espalha a informação através dessa rede.

Com cerca de 500 milhões de tweets diários, o Twitter virou uma enorme fonte de informação que é utilizada por empresas de diversos ramos para entender melhor o seu consumidor, seus nichos, investir em publicidade e propaganda dentro das tendências atuais e etc. O Twitter ultimamente também é bastante utilizado por movimentos políticos e religiosos, que visam atrair mais pessoas para o seu

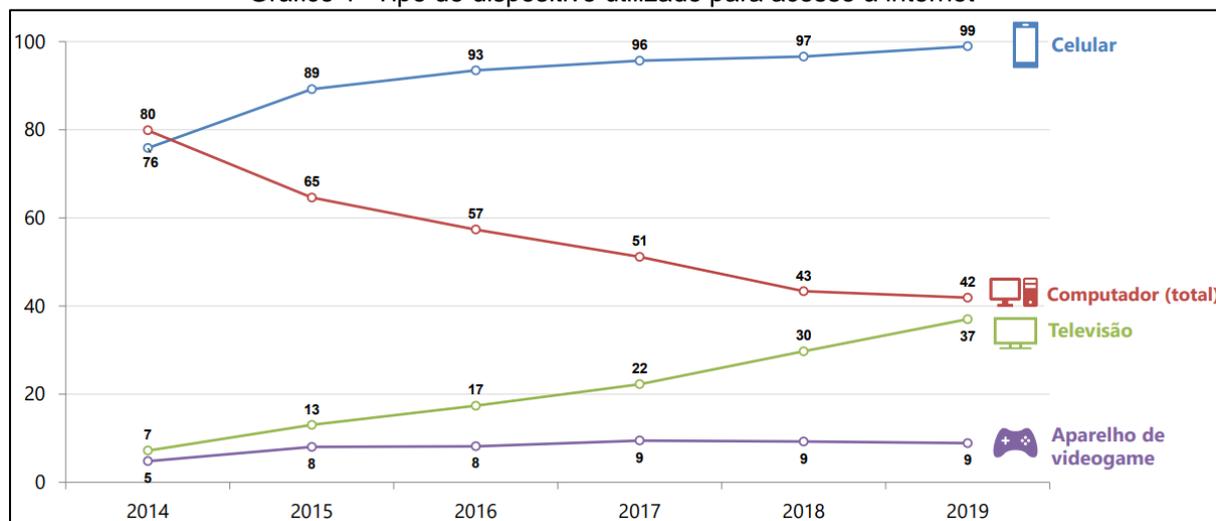
movimento e angariar votos, influenciando diretamente nas eleições de toda a parte do mundo. Inclusive é utilizado de forma antiética, onde usuários espalham um grande volume de informações falsas -conhecidas como fake News - que funcionam como um subterfúgio para difundir desinformação e manchar a reputação de adversários políticos, artistas, ideologias, etc.

A API (Interface de Programação de Aplicativo) atual do Twitter permite que a empresa abra os seus dados e funcionalidades de seu aplicativo/site para terceiros, parceiros de negócios, departamentos internos da companhia, etc. Através disso, é possível que produtos e serviços se comuniquem e aproveitem dados e funcionalidades uns dos outros por meio de uma interface documentada (IBM, 2020). Através da API do Twitter, podemos utilizar aplicativos de terceiros que consolidam dados gerados na rede social e utilizá-los para fazer análises sobre os mais diversos tipos de assuntos e pesquisas.

2.6 COMUNIDADES VIRTUAIS

O desenvolvimento da internet e das tecnologias de comunicação transformaram o mundo nas últimas décadas. As transformações ocorridas podem ser consideradas como uma revolução contemporânea da ascensão digital e da informação, que possibilitou que a informação se tornasse um processo global, mudou o tempo e o espaço, encurtou distâncias e acelerou práticas (KOHN, MORAES, 2007). No Brasil, o acesso à internet cresceu bastante nos últimos anos. Em 2019, três a cada quatro brasileiros acessam a internet, equivalente a 134 milhões de usuários (TIC Domicílios, 2019). Grande parte desses acessos se dá por meio dos aparelhos celulares, que foi encontrado em 99,2% dos domicílios que possuíam acesso a internet. Entre as pessoas que usam exclusivamente o celular para acessar a internet, esse percentual é de 58%. O gráfico 1 mostra o evolutivo do tipo de dispositivo utilizado para acesso à internet no país:

Gráfico 1--Tipo de dispositivo utilizado para acesso à internet



Fonte - Fonte: CGI.br/NIC.br, Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br)

A internet encurtou barreiras entre as pessoas, conectando e mantendo próximas pessoas de diferentes partes do país e do mundo, de diferentes credos, culturas, etc. Isso possibilitou a criação de vínculos entre pessoas sem a necessidade de estarem presentes fisicamente. Conseqüentemente, indivíduos que possuíam interesses em comum estabeleceram relações e se organizaram em comunidades virtuais, utilizando as ferramentas disponíveis na internet (sites, fóruns, blogs, redes sociais, etc.) para trocar experiências, informações, compartilhamento de valores, apoio mútuo sobre os mais diversos e variados assuntos. Segundo Schlemmer & Carvalho (2005), as comunidades virtuais são redes eletrônicas de comunicação interativa autodefinidas, organizadas em torno de um interesse ou finalidade compartilhados.

Nessa ótica, as comunidades virtuais permitem que indivíduos de diferentes partes do mundo possam debater, opinar, trocar informação e expressar suas ideias sobre determinados assuntos. Quando há troca de informações e construção de conhecimento em uma comunidade virtual, ela passa a ser considerada uma comunidade virtual de aprendizagem, sendo articulada por seus assuntos, objetivos, tarefas e pesquisas em comum (Lotito, 2005, p4).

Wenger (2006) introduz o conceito de comunidade prática como um grupo de pessoas que compartilham um interesse ou uma paixão sobre algo que eles fazem e tentam aprender como fazer isso de uma melhor forma à medida que eles interagem regularmente. A comunidade prática não existe sem uma comunidade virtual de

aprendizagem, porém nem toda comunidade de aprendizagem é uma comunidade prática. A comunidade de prática possui três características fundamentais:

- Domínio – a competência, foco ou assunto daquela comunidade que une os indivíduos e permite que eles se unam para compartilhar informações e conhecimento.
- Comunidade – referente a interação dos membros dessa comunidade, ou seja, a participação dos indivíduos nas discussões, atividades, compartilhamento de ideias e informações, etc. Isto é, a construção de um relacionamento e sentimento de dono para com a comunidade, procurando a resposta para a solução dos problemas juntos.
- Prática – é a prática que o indivíduo possui sobre aquele assunto, isto é, a vivência, a experiência de vida, histórias e ferramentas que a pessoa possui sobre determinado problema e o compartilhamento dessas experiências práticas com a comunidade, que busca aumentar a gama de práticas e informações para enfrentar os problemas recorrentes.

Gouvêa (2006, p.42) diz que na comunidade prática a participação é auto selecionada, isto é, as pessoas percebem quando e porque devem juntar-se àquela comunidade, estimulando o aprendizado, mudanças e dividir o aprendizado.

Atualmente, temos inúmeros exemplos de comunidades virtuais existentes na internet. Blogs, fóruns, sites de compra e venda de produtos, sites voltados a perguntas que outros usuários podem responder, redes sociais, etc. Dentro do objetivo proposto neste trabalho, iremos utilizar como base perfis em redes sociais que funcionam como uma comunidade virtual, visando o compartilhamento e troca de informação constante.

3 MÉTODO

3.1 MINERAÇÃO TEXTUAL

A mineração textual (Text Mining) é um processo muito utilizado para “traduzir” a linguagem natural de documentos de textos para dados numéricos, permitindo análises matemáticas e estatísticas e encontrar padrões e relacionamentos de termos com base na quantidade de vezes que ele é encontrado no documento analisado.

Para Sukanyal e Biruntha (2012, p.269), existem três passos para o processo de mineração textual: processamento de texto, aplicação de técnicas de mineração textual e análise textual. O primeiro passo trata-se de selecionar o texto a ser estudado, desde que esteja pré-processado (formatado estruturalmente sem perda de características naturais). O segundo passo trata-se da aplicação das técnicas de análise, como aplicação do algoritmo, técnicas matemáticas e estatísticas, etc. O último passo é a análise dos resultados, onde o usuário pode navegar pelos resultados construídos e alcançar os objetivos e perspectivas.

Diversas áreas científicas aplicam a mineração textual para gerar conhecimento e informações para gerar melhores estratégias. Atualmente, devido ao grande volume de informações geradas, os dados provenientes das redes sociais são frequentemente utilizados como modelo de base de pesquisas e para isso são utilizadas técnicas de mineração de texto para leitura apropriada desses dados.

3.2 CORPUS

Para Dahlet (2002, p.128), o corpus se apresenta como materialidade que é necessário fazer significar, destacando-se do pano de fundo composto entre cruzamento da problemática com a teoria. Já Sinclair (2005, p.23) diz que corpus é uma coleção de textos linguísticos em formato eletrônico que representam uma variação da língua como fonte de dados para uma pesquisa linguística.

Sendo assim, podemos definir como o corpus deste trabalho o conjunto de todos os dados obtidos que serão estudados e classificados. Todos os tweets existentes nas bases de dados estudadas compõem o corpus desta pesquisa.

3.3 COLETA DE DADOS

O Twitter pode ser tratado como uma grande fonte de informação, porém a análise desses dados não é uma tarefa fácil. Por se tratar de um campo aberto onde os usuários podem escrever de qualquer forma e sobre qualquer assunto, é necessário que os dados sejam tratados de maneira cautelosa. O padrão textual não é seguido e há presença de muitos caracteres especiais, abreviações, gírias, adaptações gramaticais, links, propagandas etc., que geram problemas na interpretação dos dados de tweets (publicações na plataforma). Sendo assim, para termos uma base de dados coesa, é necessário realizar procedimentos de limpeza da base.

3.3.1 BASE DE DADOS

Para gerar um banco de dados através de postagens no Twitter, foi programado um script em linguagem de programação R e, através do RStudio (R Core Team, 2014), compilado o código que realizava uma busca de todos os tweets (publicações no Twitter) feitos nas últimas 24 horas que continham palavras chaves que poderiam ser pré-definidas. Os tweets eram organizados em um único arquivo, em formato .xlsx (excel). Esse arquivo criado funciona como um banco de dados, contendo todas as informações dos tweets. Os termos pré-definidos eram de livre escolha e foram escolhidas palavras relacionadas a crimes e violência em geral, como por exemplo: roubo, assalto, crime, furto, homicídio, etc. Essas palavras também foram utilizadas em outros tempos verbais para aumentar a abrangência. Todas as palavras utilizadas podem ser encontradas no Anexo A. No total, foram coletados 15.796 tweets no período de 21 de novembro de 2021 até 29 de novembro de 2021.

3.3.2 SELEÇÃO DA BASE DE DADOS

Nesta etapa, foi realizada uma verificação dos dados obtidos. Cumpre registrar que teve uma falta de especificação da população alvo. Ocorreu que grande parte do volume de dados eram relacionados com assuntos totalmente diferentes daquilo que era relevante para o trabalho. Eram temas relacionados à esportes (onde usuários

alegavam que o time no qual torcia foi “roubado” ou “assaltado”), jogos, brincadeiras, conversas informais, piadas, etc., vide exemplos no quadro 1:

Quadro 1 - Exemplos de tweets que continham os termos pré-selecionados, mas que não eram relacionados com assuntos referentes a crimes, ocorrências, etc.

Usuário	Tweet
Lessaflu	Esse juiz apostou no Bahia,tão assaltando o Cuiabá
Jhon_PedroR	Estão assaltando o meu Lakers, LeBron é inocente
gallomineiro	Estão assaltando o Galo na final do futebol feminino.
brubritess	acordada assaltando minha própria geladeira 😏
bruno_amaralc	Estão furtando o meu Fluminense
VictorDeVerdade	Amo esse pokémon. Ele sai furtando vários itens pra gente pelo caminho
012bielsccp	@Palmeiras Roubando é fácil
d_udacec	@Cruzeiro tao roubando meu cruzeiro

Fonte: Elaborado pelo autor.

Pelo grande volume de dados, é extremamente difícil analisar cada um dos tweets coletados para analisar se o assunto era relacionado ao escopo do trabalho. Para evitar encontrar resultados enviesados em que outros assuntos não interferissem de forma considerável na análise dos dados, a base foi descartada. Sendo assim, foi necessário avaliar outros procedimentos mais adequados para gerar informações relacionadas ao tema.

Para gerar outro banco de dados, foi realizada uma busca por comunidades do Twitter (perfis que atuam como uma comunidade virtual) que tinham como objetivo fazer denúncias de crimes e informar a população de ocorrências pela região metropolitana do Rio de Janeiro, o que assegurava que a maior parte das informações da base de dados fossem relativos ao tema proposto no trabalho. Esses perfis funcionam de forma coletiva, na qual pessoas por toda cidade alertam a página sobre ocorrências pela cidade (através de e-mail, mensagens privadas no Twitter, através de tweets ou até mesmo por aplicativo). Procurou-se identificar os perfis mais relevantes (isto é, com mais seguidores e interações) e que atuavam com alertas sobretudo pela região suburbana do Rio de Janeiro. No fim, foram selecionados o total de 11 perfis: @AlertaAssaltoRJ; @BlogCrimesnews1; @catumbialerta; @DDalertaRio; @FogoCruzadoRJ; @Informacoes_RJ; @informe_rj; @legal_rj; @PadreMiguelNews; @PenhaNewsRJ; @RJ_OTT.

Para gerar a base de dados dos órgãos de segurança pública, foram selecionadas as páginas (perfis) oficiais da polícia militar, civil e guarda municipal do Rio de Janeiro: @PMERJ; @PCERJ; @GMRio;

Foi utilizado a API do website Vicinitas para criar todas essas fontes de dados, que possibilitou exportar as últimas 3200 interações feitas pelas páginas selecionadas (tweets, retweets, replies, porém no trabalho é considerado apenas os tweets). Essas informações foram exportadas em formato xlxs (planilha do excel) e consolidadas em um único arquivo, contendo um total de 16.453 tweets distintos.

Já nos bancos de dados das polícias militar, civil e guarda municipal, foi desenvolvido o script de busca utilizado inicialmente, pois nessas páginas não havia necessidade em avaliar os temas dos tweets. O script fez a busca pelos últimos 3250 tweets de cada um desses perfis. Sendo assim, cada banco de dados (pmerj, pcerj, gmrio) possui 3250 linhas.

Foi feita uma limpeza e tratamento da base de dados gerada e desta vez foi identificado que a maioria dos assuntos eram relacionados a denúncias, crimes e ocorrências. Ainda havia tweets que não eram relevantes para o tema, como por exemplo, anúncios de publicidade, links, propagandas, etc. Esses tweets foram excluídos da versão final da base de dados, que continha o total de 15.616 tweets únicos. É importante citar que, por conta do tamanho da base, é extremamente difícil identificar todas as informações presentes no corpus que não possuem relevância para o trabalho. Contudo, ao compilar os scripts que irão processar e analisar os dados, será possível identificar os termos não relevantes e retirá-los da análise na medida em que eles aparecem.

Conforme dito anteriormente, os scripts e API's exportaram os últimos 3250 tweets de cada página. Portanto, não há uma definição de data na base de dados, pois depende da atividade de cada perfil. Contudo, grande parte das informações são do ano de 2021 e 2020, porém contém informações (em menor volume) de outros anos, sendo o mais antigo 10 de outubro de 2015 e o mais recente sendo o dia em que os dados foram captados, 12 de dezembro de 2021. No Anexo C, destaco exemplos de linhas das fontes de dados extraídas.

3.3.3 PRÉ-PROCESSAMENTO DE DADOS

Ao analisar textos oriundos de redes sociais, é necessário compreender que por se tratar de campos abertos, muitas vezes os textos inseridos não são relevantes e informativos pois se tratam de expressões ou termos não desejados. Sendo assim, é necessário realizar um trabalho para ajustar o corpus para desconsiderar estes termos.

3.3.3.1 REMOÇÃO DE PONTUAÇÃO, CARACTERES ESPECIAIS E AJUSTE DO TEXTO

As ferramentas de análise textual não compreendem certos tipos de caracteres. Tampouco analisar esse tipo de caractere é relevante para o trabalho. Sendo assim, através do script programado foi retirado todas as pontuações, acentos, caracteres especiais, etc., ou seja, todos os caracteres não alfanuméricos.

Além disso, foi realizado um trabalho para colocar todas as letras do corpus em minúsculo, pois em análises tipográficas, os softwares são sensíveis ao tipo de caractere (maiúscula/minúscula). Isto significa que há distinção entre palavras maiúsculas e minúsculas, mesmo que se trate do mesmo termo. Por exemplo: a palavra “Assalto” seria tratada de forma distinta da palavra “assalto”, pois a primeira começa com um A maiúsculo, diferentemente da segunda.

Outro tratamento utilizado foi a modificação de alguns termos e transformando-os em um só, pois apesar de se tratar de termos com mais de uma palavra, elas traziam o significado de apenas uma única coisa. Neste trabalho, temos como exemplo localidades do Rio de Janeiro, como “Nova Iguaçu”. O compilador do código analisaria os termos “nova” e “Iguaçu” de forma separada. Sendo assim foi necessário transformar essas passagens em apenas um termo, e, neste caso, “Nova Iguaçu” foi transformado em “Nova_iguaçu”. Isso ocorre com outros termos presentes no corpus.

3.3.3.2 TOKENIZAÇÃO

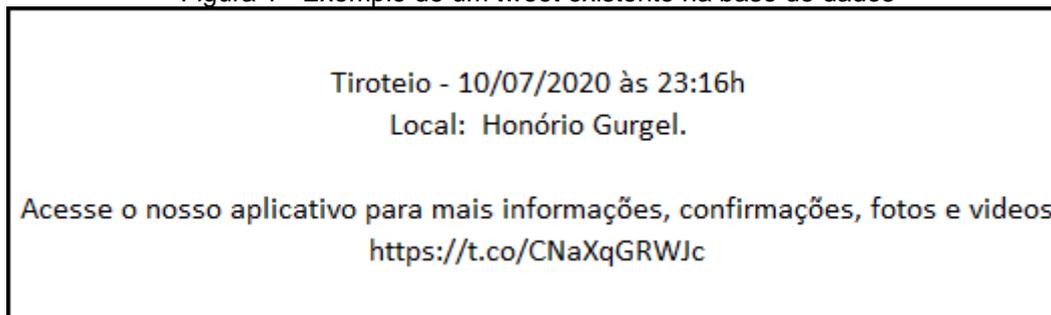
O processo de tokenização é um dos mais importantes, pois através dele que cada palavra em nosso banco de dados é transformada em um único token, ou seja, cada palavra escrita em um tweet transforma-se em um termo e assim a contagem desses termos poderá ser analisada pelo computador. Isso garante que o computador não leia cada tweet como um único termo. Através do pacote `tidytext` (SILGE; ROBINSON, 2017) é feita essa conversão, transformando a base de dados em uma tabela onde cada linha possui um único token.

3.3.3.3 STOPWORDS

Podemos definir as *stopwords* como as palavras que não possuem relevância e sentido para o texto. Essas palavras geralmente são bastante frequentes no corpus, porém não trazem significado relevante ao analisá-los. Geralmente, essas palavras são artigos como “o”, “a”, “uma”, “uns”, etc. Também podem ser palavras de uso rotineiro como “tá”, “está”, “também”, “tenho”, etc. Para realizar a remoção desses termos foi utilizado o pacote *stopwords* (BENOIT; MUHR; WATANABE, 2021), que contém uma grande coleção das *stopwords* mais comuns em diversas línguas.

Dependendo do contexto analisado e também da fonte, percebe-se que alguns termos também podem ser considerados como *stopwords*, pois representam “vícios” na linguagem da base analisada. Portanto, à medida em que o código era executado, era possível notar a presença de diversas palavras que não traziam sentido para o trabalho. Muitas dessas palavras tratavam-se de hashtags, que funcionam como palavras-chaves e são bastante utilizadas em redes sociais para gerar engajamento e divulgação do perfil. Outros termos identificados como *stopwords* eram parte de links e mensagens padrões que eram enviadas junto com os tweets, como mostra a figura 1:

Figura 1 - Exemplo de um tweet existente na base de dados



Fonte – Elaborado pelo autor

Neste exemplo, a passagem “Acesse o nosso aplicativo para mais informações, confirmações, fotos e vídeos <https://t.co/CNaXqGRWJc>” é uma mensagem padrão bastante utilizada nos tweets desta página, aumentando a frequência desses termos no documento analisado. Sendo assim, esses termos foram identificados e adicionados à lista de *stopwords*.

Também foi possível identificar diversos termos que eram números. Ao analisar tais termos, entendia-se que eram referentes à horários e datas. O fato de transformar cada termo em um token elevou a complexidade em analisar tais números, pois eles eram tratados separadamente e ficaria difícil identificar o horário e data completa de cada tweet, visto o grande volume de dados. Sendo assim, esses números também foram adicionados à lista de *stopwords*. A lista completa de *stopwords* pode ser vista no Anexo B.

3.4 TF-IDF

O TF-IDF (term frequency–inverse document frequency) é uma ferramenta estatística tradicionalmente utilizada na mineração textual para medições de similaridade de textos não-estruturados ou semiestruturados (CASTRO, LIMA, 2012). Esse método possibilita quantificar documentos textuais e mensurar a importância que um termo possui dentro de um documento. Primeiramente, calcula-se a frequência dos termos (tf), isto é, quantidade de vezes em que os termos são repetidos dentro de um documento.

$$tf(x) = \frac{\text{número de vezes que o termo } x \text{ é citado no documento}}{\text{total de palavras em um documento}}$$

Ocorre que em documentos muito extensos, existem palavras que são frequentemente repetidas e não são relevantes no contexto geral. Essas palavras geralmente são artigos, como: “ou”, “a”, “o”, “as”, “os”, “de”, “um”, “umas”, etc. Essas palavras, como definido na sessão, podem variar para cada documento, dependendo do contexto. Como esses termos não são diferenciais para a classificação de documentos, é necessário comparar os termos com a proporção inversa daquela palavra em todo o documento (idf).

$$idf(x) = \log \frac{\text{número total de documentos}}{\text{número de documentos que contém o termo } x}$$

Por fim, combinando a frequência do termo (tf) com o inverso da frequência nos documentos (idf), temos o TF-IDF, que é representado pela fórmula:

$$TF - IDF = f(x, d) \cdot \log(D / f_{x,D})$$

Onde $f(x, d)$ significa o número de vezes que o termo x aparece no documento d , D é o tamanho do documento (total de documentos) e $f_{x,D}$ é o número de documentos D em que o termo x está presente.

Utilizando este método, conseguimos definir o quanto é relevante uma palavra em um documento. O valor TF-IDF de um termo aumenta de acordo de vezes em que essa palavra é repetida no documento, porém seu valor é balanceado pela frequência desse termo no corpus (documento ou conjunto de documentos, textos etc.). Assim é possível os termos que são relevantes para o tema proposto.

3.5 NUVEM DE PALAVRAS

A nuvem de palavras é um recurso visual utilizado para demonstrar o grau de frequência de termos em um texto. Cada palavra possui um tamanho e intensidade diferentes, baseados na frequência na qual essa palavra aparece no corpus. Quanto maior a frequência de uma palavra no documento, maior é o seu tamanho e a intensidade de sua cor.

Quanto mais o termo é utilizado no documento, mais chamativa ela se torna nesse gráfico, indicando a relevância de uma palavra dentro daquele contexto. Esse

gráfico é bastante utilizado, pois permite uma identificação fácil e rápida sobre os aspectos de uma pesquisa. Também é uma ótima maneira para comparar textos, onde podemos analisar diferentes documentos e buscar suas similaridades.

4 ANÁLISE DE RESULTADOS

Os procedimentos foram realizados em quatro bases diferentes. Uma delas continham os tweets referentes às comunidades virtuais e as outras três eram tweets das páginas oficiais da polícia militar, civil e guarda municipal do Rio de Janeiro. A tabela 1 mostra a quantidade de termos existentes em cada base de dados, após o processo de tokenização:

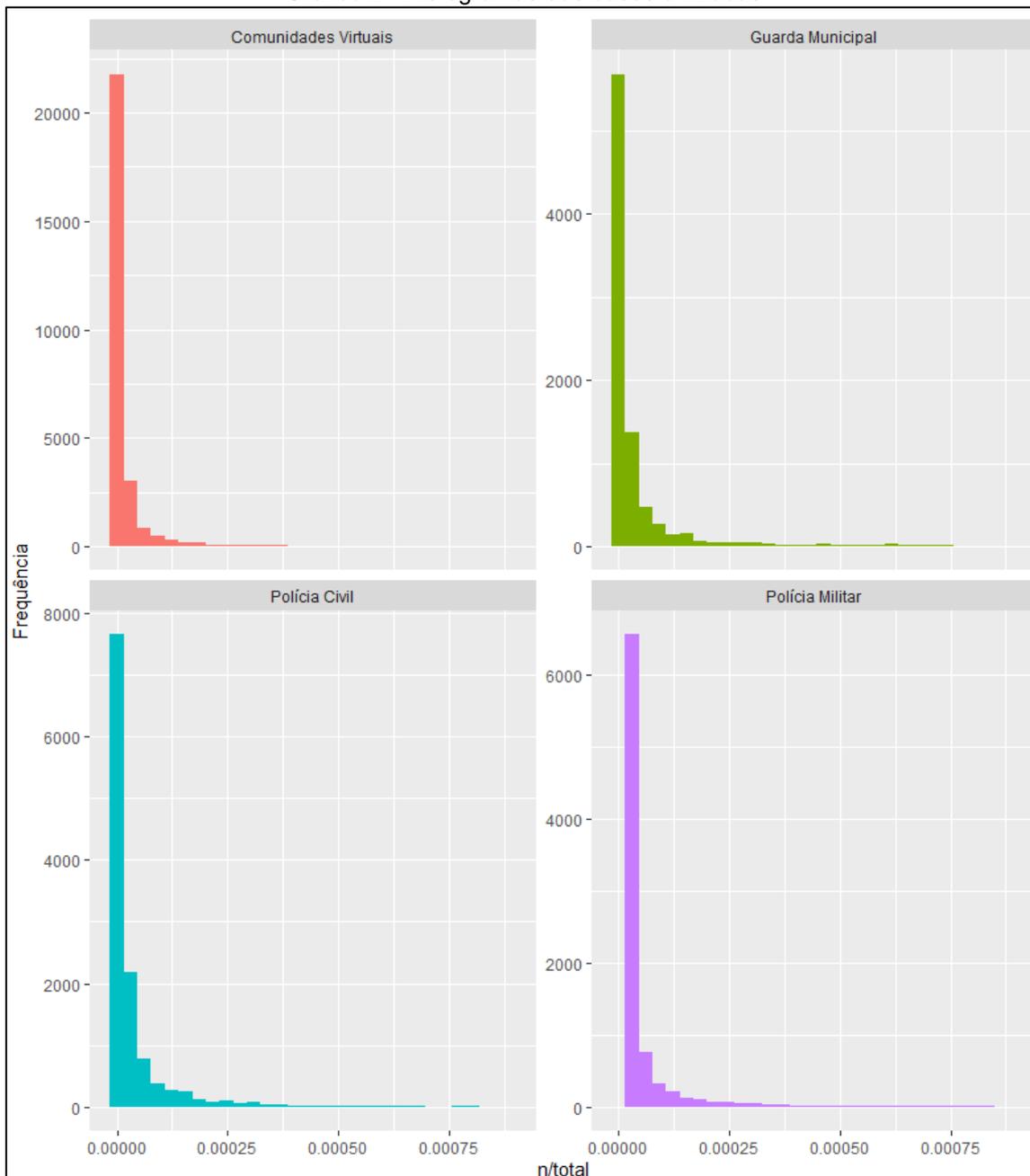
Tabela 1 - Quantidade de termos existentes em cada documento

Base	Total de Termo
Comunidades Virtuais	229480
Guarda Municipal	71396
Polícia Civil	71073
Polícia Militar	62832

Fonte: Elaborado pelo autor

Por conter mais perfis analisados (total de 11), o banco de dados das Comunidades Virtuais possui mais termos que as outras bases. O histograma abaixo (gráfico 2) ajuda entender melhor a distribuição dos termos nas quatro bases de dados utilizadas no trabalho:

Gráfico 2 - Histogramas das bases utilizadas



Fonte: Elaborado pelo autor

Estes gráficos representam a distribuição das frequências dos termos (tf), ou seja, o número de vezes que uma palavra aparece no documento dividido pelo total de termos existentes no documento. “n” é o número de vezes que a palavra é usada em um documento e total é o total de palavras nesse documento (n/total).

Podemos notar que a distribuição das quatro bases utilizadas é similar, onde há a presença de termos frequentemente utilizados (colunas com alto valor, referente a termos bastante repetidos nos documentos que pode ser visto na Tabela 2), mas o alongamento à direita do gráfico mostra que há muitas palavras que são raramente

usadas e são pouco frequentes. A tabela abaixo (tabela 2) mostra os termos mais frequentes de cada base:

Tabela 2 - Frequência dos termos das quatro bases utilizadas

Base	Termo	n	Total
Comunidades Virtuais	https	10631	229480
Comunidades Virtuais	t.co	10631	229480
Polícia Civil	t.co	3318	71073
Polícia Civil	https	3319	71073
Guarda Municipal	t.co	3275	71396
Guarda Municipal	https	3274	71396
Polícia Militar	https	3137	62832
Polícia Militar	t.co	3173	62832

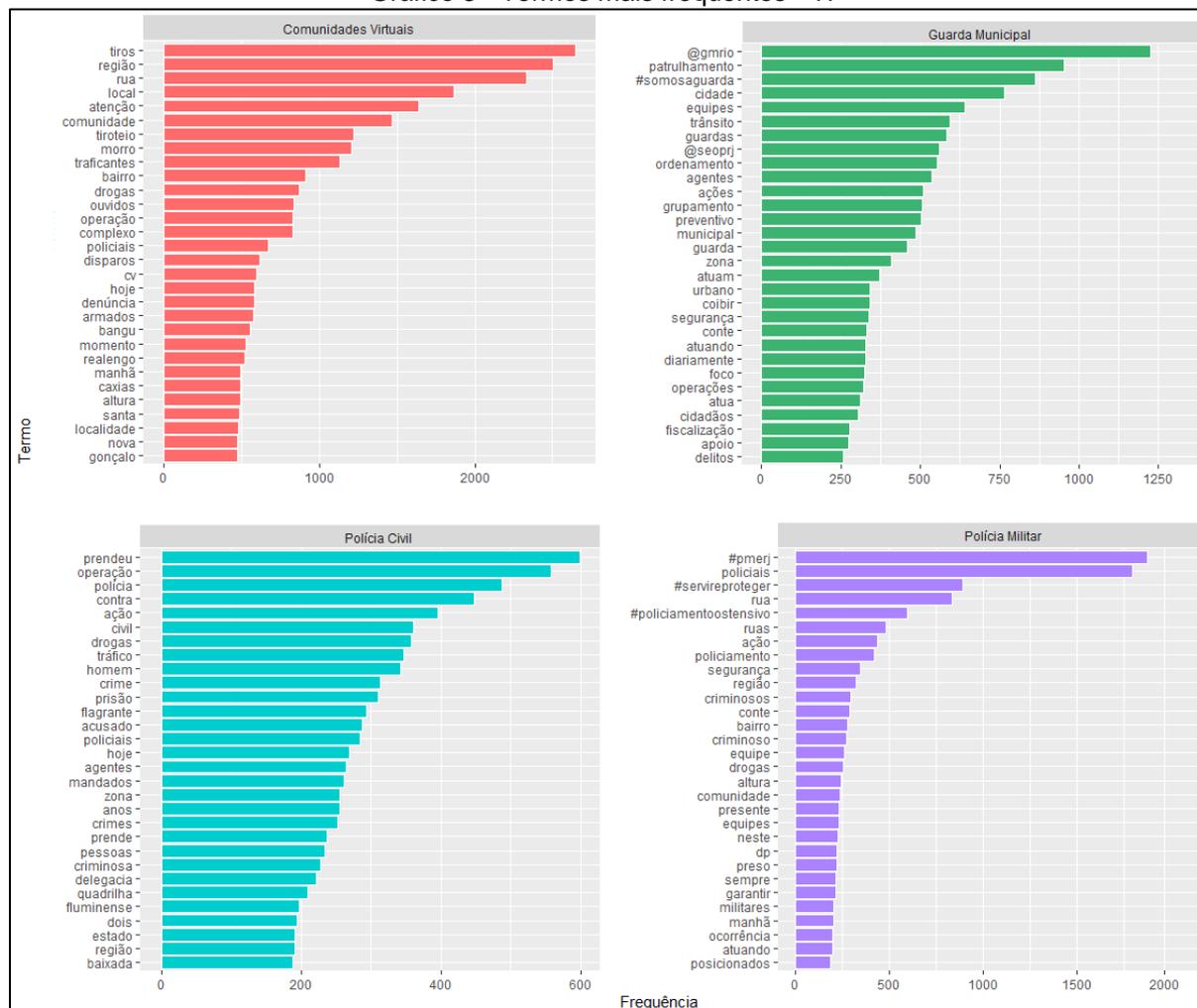
Fonte: Elaborado pelo autor

A tabela 2 mostra os termos mais frequentes existentes nas quatro bases analisadas. A coluna “n” mostra o número de vezes que o termo aparece na base e a coluna “Total” representa o total de termos existentes na base. Podemos ver claramente tratam-se de termos que não trazem significado para a análise, pois são referentes à links. Isso acontece porque grande parte dos tweets possuem links que redirecionam para alguma pagina no próprio Twitter, site com a notícia, links de propagandas, vídeos, etc. O termo “https” significa protocolo de transferência de hipertexto seguro, que é um protocolo de segurança utilizado na URL (uniform resource locator) de grande parte dos sites da internet. Já o termo “t.co” refere-se ao encurtador de links disponibilizado pelo próprio Twitter, que encurta automaticamente todo link em seus tweets para um link <http://t.co/>. Isso também explica as semelhanças nos valores dos termos “https” e “t.co”, pois ambos vêm acompanhados.

Esses resultados mostram a importância do tf-idf, que visa encontrar as palavras mais relevantes nos documentos, diminuindo o peso dos termos mais comuns e aumentando o peso das palavras menos utilizadas no corpus. Ou seja, encontrar palavras importantes, mas que não são muito comuns.

Sendo assim, através de funções do script, algumas palavras foram banidas da análise (adicionadas à lista de stopwords) pois não tinham valor significativo na análise. O gráfico 3 representa os 30 termos mais frequentes em cada base após a retirada dos termos indesejados:

Gráfico 3 - Termos mais frequentes - TF



Fonte: Elaborado pelo autor

O gráfico 3 traz resultados interessantes. Na base “Comunidade Virtuais”, é notável o tom de alerta e o caráter informativo. As cinco palavras mais frequentes possuem relação direta com avisos e alertas (“tiros”, “atenção”) e localidades (“região”, “rua”, “local”). As palavras seguintes mantêm o mesmo princípio, com termos relacionados diretamente a ocorrências e localidades.

Percebemos também que há muitas referências para termos relacionados a armas de fogo, como: “tiros”, “tiroteio”, “ouvidos” (refere-se à tiros ouvidos/escutados), “armados” e “disparos”. Também há bastante indicadores de locais, como por

exemplo: “comunidade”, “complexo”, “rua”, “local”, “bangu”, “realengo”, “caxias”. Isso mostra que grande parte dos tweets feitos por essas comunidades virtuais buscam sinalizar ocorrências e os locais que elas ocorrem, com objetivo de alertar seus leitores/seguidores para que eles não se encontrem em uma situação de risco de vida.

A base da polícia militar e a base da guarda municipal possuem semelhanças em alguns aspectos. Primeiramente, podemos ver que ambas as páginas utilizam bastantes hashtags para gerar engajamento e menções a si próprias. Os termos “#somosguarda”, “diariamente”, “@seoprij” (Secretaria de Ordem Pública do Rio de Janeiro), “#pmerj”, “servireproteger”, “#policiamentoostensivo”, “conte” e “presente” são bastante utilizadas para reafirmar o trabalho dos órgãos e passar uma sensação de presença constante.

Outra semelhança entre essas duas páginas é a existência de termos relacionados à prevenção, como: “patrulhamento”, “ordenamento”, “ações”, “preventivo”, “coibir”, “policiamento”, “segurança”, “ação” (referente à ação preventiva).

Também podemos notar que existem muitas referências no que se diz respeito ao patrulhamento, com a utilização de termos como: “equipes” (frequente em ambas as bases), “agentes”, “grupamento”, “atuam”, “fiscalização”, “apoio”, “ação”, “policiamento”, “atuando”, “posicionados”.

Essas semelhanças são reflexo das funções de ambas as polícias, que se dá pela execução de policiamento ostensivo (isto é, patrulhamento devidamente uniformizado e de fácil reconhecimento para coibir a infração de leis), preventivo e aparelhado e garantir a proteção da população e de bens e serviços.

Enquanto isso, os termos mais utilizados pela página da polícia civil possuem características judiciárias, isto é, são mais relacionadas com a execução de prisões e expedições de mandados, como podemos ver nos termos: “preendeu”, “prisão”, “flagrante”, “acusado”, “mandados”, “prende”. Isto se dá ao fato de a polícia civil ser mais voltada para o lado da investigação de crimes, enquanto a polícia militar e guarda municipal possuem um papel de presença mais ostensiva do Estado.

Tabela 3 - Função tf-idf nos termos mais utilizados

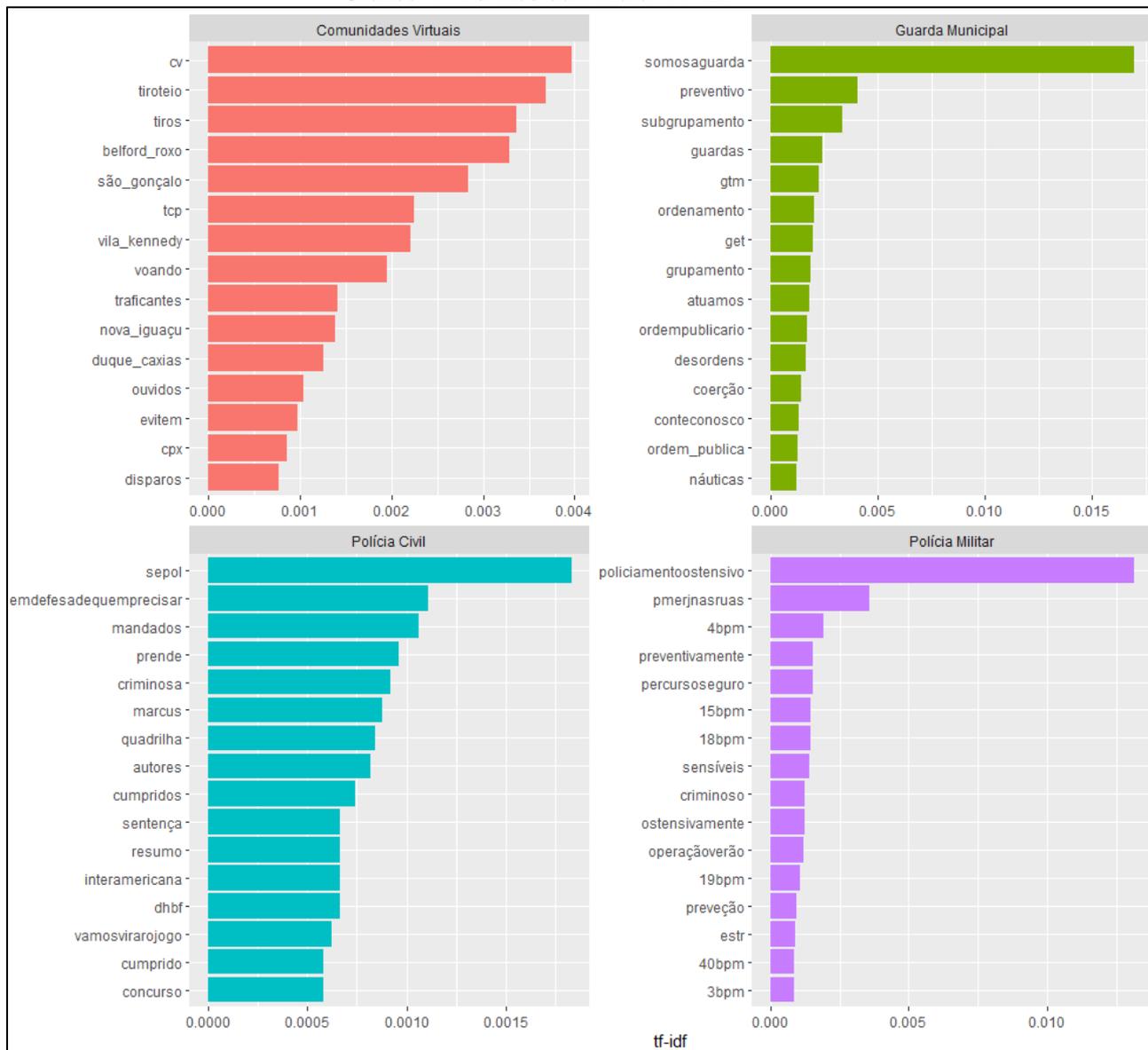
Base	Termo	n	Total	tf	idf	tf-idf
Comunidades Virtuais	https	10631	229480	0.0465	0	0
Comunidades Virtuais	t.co	10631	229480	0.0465	0	0
Polícia Civil	t.co	3318	71073	0.0467	0	0
Polícia Civil	https	3319	71073	0.0467	0	0
Guarda Municipal	t.co	3275	71396	0.0459	0	0
Guarda Municipal	https	3274	71396	0.0459	0	0
Polícia Militar	https	3137	62832	0.0499	0	0
Polícia Militar	t.co	3173	62832	0.0499	0	0

Fonte: Elaborado pelo autor

A tabela 3 mostra a função tf-idf sendo aplicada para os termos mais frequentes das bases analisadas. Podemos observar que as palavras possuem valores altos na frequência do termo (tf), porém valor 0 no idf e tf-idf. Lembrando que, $idf(t,D) = \log(N/nt)$. Assim, com esses termos ocorrendo em todas as quatro bases, $idf = \log(4/4) = 0$. Como $tf-idf = tf * idf$, qualquer multiplicação por zero será zero. Logo, $tf-idf = 0$.

Sendo assim, é necessário reorganizar os dados para que mostrem de forma crescente os maiores valores tf-idf das quatro bases analisadas. No gráfico 4, podemos ver os 15 termos com maior tf-idf de cada base:

Gráfico 4 - Termos com maior TF-IDF



Fonte: Elaborado pelo autor

Podemos observar a presença de novos termos em todas as bases. Assim, com tf-idf, podemos observar os termos mais representativos (mais exclusivos) de cada um dos quatro grupos. Porém, também existem termos que são bastante frequentes nas bases e ainda possuem alto tf-idf, o que mostra a relação linear entre frequência x tf-idf.

Na base “Comunidades Virtuais”, o termo com maior valor tf-idf é “cv”, que é uma sigla bastante utilizada para se referir à facção criminosa “Comando Vermelho”, existente em todo o Brasil, majoritariamente no Rio de Janeiro. No ranking, ainda

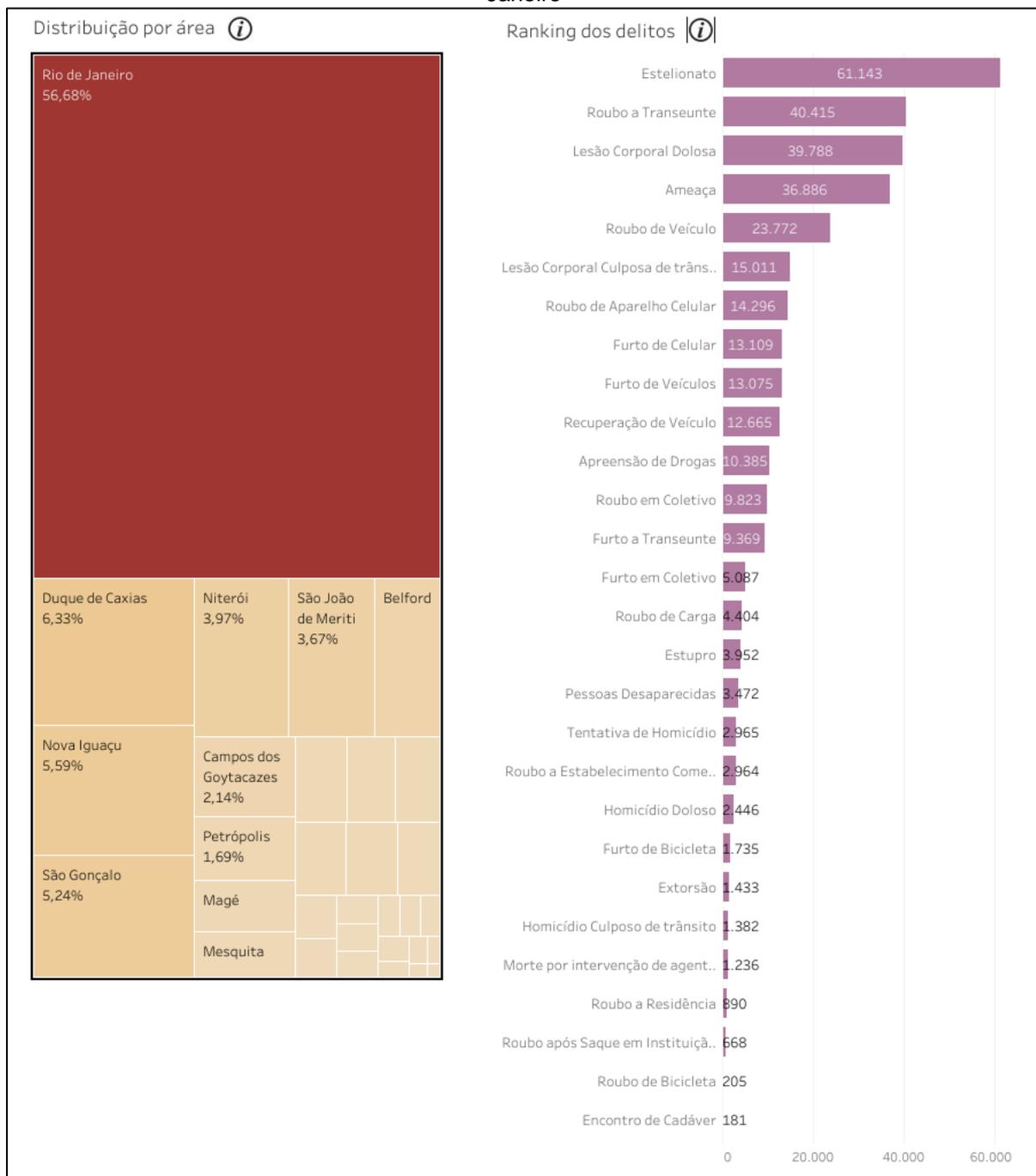
existe outro termo utilizado para se referir à outra facção criminosa - “tcp” -, comumente utilizado para se referir ao “Terceiro Comando Puro”.

A relevância dos termos que fazem referência a facções criminosas é justificada, principalmente se levarmos em conta o contexto e objetivo dos tweets das comunidades virtuais selecionadas. O “Comando Vermelho” e o “Terceiro Comando Puro” dominam diversas áreas do estado, estando presente em grande parte das comunidades, morros e favelas do Rio de Janeiro. Pela grande influência que possuem, são responsáveis por grande parte dos confrontos armados entre facções ou contra a polícia, por assaltos e violências. Outros termos ranqueados corroboram com esta visão, visto que palavras como “tiroteio”, “tiros”, “voando” (palavra utilizada para falar sobre tiroteios, como por exemplo: “a bala tá voando”), “ouvidos” (disparos de armas de fogo escutados), “disparos” estão diretamente ligados à confrontos e violência no geral.

Outro ponto que se mostra relevante é o tom de informativo, com termos fazendo referências a localidades do Rio de Janeiro. Os termos “belford_roxo”, “são_gonçalo”, “vila_kennedy”, “nova_iguaçu”, “duque_caxias”, “cpx” (sigla utilizada como abreviação de complexo) são bastante relevantes nessa base de dados, pois mostra que os tweets visam informar e alertar a população sobre as regiões que estão ocorrendo violência em geral, sobretudo confrontos armados.

Ao checar os dados de criminalidade e violência na Região Metropolitana e Norte Fluminense do Rio de Janeiro no ano de 2021, disponíveis no site do Instituto de Segurança Pública, temos o seguinte resultado (figura 3):

Figura 3 - Monitoramento de criminalidade da Região Metropolitana e Norte Fluminense do Rio de Janeiro



Fonte: Instituto de Segurança Pública. Disponível em: <http://www.ispvisualizacao.rj.gov.br/Monitoramento.html>. Acesso em: 19 jul. 2022

Observamos que a grande parte dos delitos estão concentrados no município do Rio de Janeiro, haja visto sua grande população, que concentra cerca de 39% da população do Estado. Segundo os indicadores, o município do Rio de Janeiro concentra 56,68% dos delitos do estado. Contudo, outros municípios também possuem bastante participação nesses indicadores, sendo os principais deles: Duque de Caxias, Nova Iguaçu, São Gonçalo, Niterói, São João de Meriti, Belford Roxo,

Campo dos Goytacazes, Petrópolis, Magé e Mesquita. Isso mostra que alguns dos termos mais relevantes da base de dados “Comunidades Virtuais” são justamente de áreas com maiores indicadores de criminalidade do Estado do Rio de Janeiro.

Apesar de serem necessárias análises mais técnicas, podemos entender que os dados obtidos das comunidades virtuais do Twitter indicam uma tendência do que realmente ocorre no estado do Rio de Janeiro. Apesar dos dados do Instituto de Segurança Pública não registrarem indicadores de tiroteios que ocorrem pelas cidades, que é algo bastante citado e relevante na base de comunidades virtuais, entendemos que confrontos armados também acontecem bastante nessas regiões, principalmente se levarmos em consideração o contexto histórico, social e econômico dessas regiões (áreas mais pobres e dominadas pelo tráfico e milícias armadas).

Analisando os termos com maiores td-idf das bases policiais, percebemos que apesar dos termos serem um pouco diferentes, eles possuem um significado e uma tendência similar com os termos mais frequentes (tf). As bases da Guarda Municipal e da Polícia Militar ainda estão voltadas para a utilização de termos que reforçam o próprio trabalho, com o uso de hashtags para gerar engajamento e trazer sentimento de presença para as pessoas.

Os termos mais relevantes utilizados pela Guarda Municipal: “somosaguarda”, “subgrupamento”, “gtm” (sigla para Grupamento tático móvel), “guardas”, “get” (sigla para grupamento especial de trânsito), “conteconosco”; e os termos utilizados pela Polícia Militar: “policiamentoostensivo”, “pmerjnasruas”, “percursoseguro”, “ostensivamente”, “operaçãooverão” parecem uma tentativa de reforçar o trabalho dessas duas polícias.

Ainda analisando o tf-idf das bases da Guarda Municipal e da Polícia Militar, notamos que existem palavras voltadas à prevenção (“preventivo”, “ordenamento”, “preventivamente”, “prevenção”), mantendo a mesma tendência das palavras mais frequentes (tf).

No ranking da polícia militar, podemos notar os seguintes termos: “4bpm”, “15bpm”, “18bpm”, “19bpm”, “40bpm” e “3bpm”. A sigla “bpm” é utilizada para se referir à “Batalhão”. Isso mostra que os termos mais relevantes utilizados pela página da polícia militar do Rio de Janeiro são referentes aos seus batalhões. Isso pode

significar que esses batalhões citados podem ser os mais atuantes da cidade e/ou suas áreas de atuação possuem altos índices de criminalidade. Contudo, para fazer essa afirmação primeiramente é necessário analisar dados mais amplos como: contingente dos batalhões, áreas de atuação, índices de criminalidade nas regiões dos batalhões, etc.

O tf-idf da Polícia Civil mostra termos que indicam a utilização de hashtags como: “emdefesadequemprecisar” e “vamosvirarojogo”; assim como menções a seus próprios órgãos e funcionários, como: “sepol” (sigla para Secretaria de Estado de Polícia Civil), “dhubf” (sigla para Delegacia de Homicídios da Baixada Fluminense) e “marcus” (referente à Marcus Vinícius Braga, que na data dos tweets analisados era secretário de Polícia Civil). Conforme anteriormente visto, seus termos mais relevantes ainda são referentes ao seu papel mais investigativo e executivo (no sentido de executar prisões, mandados e sentenças).

Sendo assim, percebe-se que os termos mais relevantes utilizados pelas polícias civil, militar e guarda municipal são palavras referentes ao seu trabalho em si e suas funções exercidas, além utilização de ferramentas como as hashtags para gerar engajamento. Isso mostra que a forma de comunicação dessas polícias no Twitter é muito mais estratégica e institucionalizada do que uma linguagem típica de internet. Em contrapartida, ao analisar os termos mais relevantes das comunidades virtuais, percebemos que sua linguagem é muito mais voltada para a transmissão de informações e alertas, com uma linguagem muito mais coloquial e que passa melhor a sensação de realidade dos fatos que ocorrem na cidade.

Enquanto a comunicação policial tenta passar uma imagem de presença, trabalho e se comunica de forma mais institucionalizada, as comunidades virtuais atuam como redes de alerta para a população, trazendo às redes o dia-a-dia da criminalidade e violência nas mais diversas áreas do Estado do Rio de Janeiro.

5 LIMITAÇÕES, CONCLUSÃO E TRABALHOS FUTUROS

A classificação textual é um trabalho que possui muitas particularidades e se torna ainda mais complexa quando utilizamos textos que se originam de redes sociais, especialmente o Twitter. A grande quantidade de informação que conseguimos extrair

das redes sociais somada às peculiaridades existentes na língua neste meio (jargões, abreviações, links, hashtags e inúmeras outras características) torna a análise textual mais onerosa e complexa.

Primeiramente, é difícil identificar se o tweet realmente se refere a assuntos no que diz respeito à criminalidade, violência e ocorrências. O grande volume de dados do trabalho (26.203 tweets) torna extremamente custosa a classificação manual do que realmente é relevante para o tema proposto. Assim, é necessário utilizar ferramentas e técnicas que facilitem e auxiliem na limpeza do texto, para que tweets que não possuem relação com o assunto sejam desconsiderados na reprodução dos resultados. Ainda assim, é possível que existam diversos tweets que não possuem relação com o tema tratado, mas as técnicas e ferramentas utilizadas anteriormente possibilitam a detecção dessas informações e também minimizam o impacto que esses dados causam no resultado final. Além disso, é importante citar que torna-se cada vez mais frequente a utilização do big data como mecanismo de gestão estratégica de análise. A organização desses dados torna possível interpretar e compreender diversas questões e assuntos.

Ainda que diversas conclusões possam ser tiradas deste trabalho, existem limites e pontos de atenção que precisam ser abordados. Primeiramente, a utilização de perfis que visam alertar a população pode enviesar o resultado final do trabalho, pois as páginas já existem para tratar aquele problema. A importância da utilização e aplicação de técnicas de aprendizado de máquina seria extremamente importante, pois seria possível gerar dados de todos os perfis da rede e, aplicando tais técnicas, assegurar que o contexto dos dados obtidos fosse relacionado a violência, criminalidade e ocorrências.

Sendo possível utilizar dados de perfis globais, há a possibilidade de utilizar técnicas de georreferenciamento, isto é, utilizar os dados de geolocalização disponíveis no próprio Twitter para definir melhor a área de análise de tweets, assim como gerar melhores indicadores de tempo e espaço. Conforme visto no gráfico 1, o principal meio de acesso à internet no Brasil se dá por meio de dispositivos móveis. No Twitter, caso um usuário faça uma publicação de um dispositivo móvel e autorize a rede social a utilizar informações de GPS (sistema de posicionamento global), é

possível detectar de qual local do planeta a postagem foi feita. Essa informação seria imprescindível e bastante segura para gerar dados e indicadores espaciais.

No que diz respeito às técnicas utilizadas neste trabalho, ela nos permitiu enxergar as diferenças enormes entre os perfis das polícias e das comunidades virtuais. Enquanto as comunidades virtuais se preocupam em estar sempre indicando ocorrências e localidades onde estão acontecendo crimes e violência em geral, os perfis policiais utilizam suas páginas com uma ótica completamente institucional. Isso acaba com o potencial dos perfis policiais (no que diz respeito a melhorias de comunicação entre os órgãos e a população), bem como engessa e limita a participação desses perfis na rede.

Conforme vimos anteriormente no trabalho, a mídia exerce grande influência na percepção da violência e na opinião e agenda pública. As mídias sociais também são uma forma de mídia, principalmente se levarmos em conta a velocidade na qual a informação corre por todo o mundo. Talvez esse seja um dos motivos que leva aos perfis policiais a utilizarem um tom mais ameno em suas redes, para não inflamar a população e não gerar sensações de insegurança. Contudo, podemos ver no nosso dia-a-dia que é cada vez maior a falta de sintonia entre as polícias e a sociedade. As redes sociais seriam uma ótima ferramenta para aproximar esses órgãos da população, onde poderiam utilizar estratégias de comunicação para entender melhor e se aproximar cada vez mais da sociedade, conversando em sintonia e com os mesmos objetivos.

O tf-idf permitiu analisar e categorizar os termos mais importantes de cada banco de dados analisado. Com isso, foi possível entender melhor se a comunicação feita pelas redes sociais pode ter valor para entender melhor a realidade. Foi possível fazer uma relação entre as localidades mais relevantes na comunicação das comunidades virtuais e vimos que há uma relação direta entre esses locais citados e as áreas do Estado do Rio de Janeiro onde mais ocorrem crimes e violências.

Contudo, o tf-idf possui algumas desvantagens pois esse método não entende a semântica de palavras e textos, sendo melhor adequado para documentos nível lexical. Além disso, esse método possui relação direta com o tamanho do documento e a frequência do mesmo, fazendo com que não seja adequado para documentos pequenos.

Dito isto, é importante que num trabalho futuro outras técnicas de análise textual sejam feitas nas mesmas bases de dados, como por exemplo, a técnica LSI (indexação semântica latente), que consegue analisar a ordenação de termos em um documento e a partir disso, definir as palavras e documentos mais relevantes. Desta forma, podemos comparar os resultados das duas técnicas e entender melhor qual seria a mais adequada para o trabalho proposto. Além disso, o emprego de outras análises e técnicas para identificar melhor tweets relacionados à crimes e violência pode tornar o trabalho mais robusto e seguro, o que pode permitir o uso dessas análises em setores de inteligência das polícias para construção de indicadores de criminalidade e, conseqüentemente, auxiliando na segurança pública do estado do Rio de Janeiro.

6 REFERÊNCIAS

ABREU, Mauricio de Almeida. **A EVOLUÇÃO URBANA DO RIO DE JANEIRO**. 3. ed. [S. l.: s. n.], 1997.

BENOIT, Kenneth; MUHR, David; WATANABE, Kohei. **Stopwords: Multilingual Stopword List**. [S. l.], 28 out. 2021. Disponível em: <<https://cran.r-project.org/web/packages/stopwords/index.html>>. Acesso em: 1 jul. 2022.

BRASIL. **Lei nº 7209, de 11 de julho de 1984**. Altera dispositivos do Decreto-lei nº 2.848, de 7 de dezembro de 1940 - Código Penal, e dá outras providências. [S. l.], 11 jul. 1984. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/1980-1988/l7209.htm>. Acesso em: 12 jun. 2022.

CANO, I. Registros criminais da polícia no Rio de Janeiro: problemas de validade e confiabilidade. In: FÓRUM DE DEBATES CRIMINALIDADE, VIOLÊNCIA E SEGURANÇA PÚBLICA NO BRASIL: UMA DISCUSSÃO SOBRE BASES DE DADOS E QUESTÕES METODOLÓGICAS, 1., Rio de Janeiro. **Anais...** Rio de Janeiro: Ipea; Ucam, 2000. p. 111-121.

CARVALHO, Edemir de. CIDADES BRASILEIRAS, CRESCIMENTO E DESIGUALDADE SOCIAL. **ORG & DEMO**, [s. l.], ed. 3, p. 45-53, 2002. DOI <https://doi.org/10.36311/1519-0110.2002.v3n1.439>. Disponível em: <<https://revistas.marilia.unesp.br/index.php/orgdemo/article/view/439>>. Acesso em: 18 jun. 2022.

DAHLET, V. O PROCEDER DA PESQUISA: QUAIS AS RELAÇÕES ENTRE PROBLEMÁTICA, DISSERTAÇÃO E CORPUS?. **Letras**, [S. l.], n. 21, p. 127–132, 2000. DOI: 10.5902/2176148511535. Disponível em: <<https://periodicos.ufsm.br/letras/article/view/11535>>. Acesso em: 25 jul. 2022.

DURKHEIM, Émile. **As Regras do Método Sociológico**. 9. ed. Lisboa: Editorial Presença, 2004. 167 p. Disponível em: <<https://wp.ufpel.edu.br/franciscovargas/files/2018/05/As-Regras-Do-Metodo-Sociologico-Emile-Durkheim.pdf>>. Acesso em: 22 jun. 2022.

FOGO CRUZADO RJ; DISQUE DENÚNCIA; UNIVERSIDADE FEDERAL FLUMINENSE. GENI/UFF; UNIVERSIDADE DE SÃO PAULO. NEV-USP. **Mapa dos grupos armados do Rio de Janeiro**. [S. l.], 10 2020. <Disponível em: <https://geni.uff.br/2021/03/26/mapa-dos-grupos-armados/>>. Acesso em: 14 jul. 2022.

FONTINHA, Rodrigo. **Novo Dicionário Etimológico da Língua Portuguesa**. Porto: Domingos Barreira, s/d.

GOMIS, Lorenzo. **Teoría del Periodismo**. Espanha: Ediciones Paidós Ibérica, 1991. 212 p. ISBN 84-7509-655-7. Disponível em: <<https://docs.google.com/file/d/0B2c0PCKe2W55X2dzVkhLc0cxUWM/edit?resourcekey=0-G77Kks7L1ccqQAwmNMdpRQ>>. Acesso em: 6 jul. 2022.

GOUVÊA, Maria Teresa Andrade de. **Um Modelo para Fidelização em Comunidades de Prática**. 2005. Dissertação (Mestrado em Informática) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005. Disponível em: <<http://objdig.ufrj.br/15/teses/MariaTeresaAndradedeGouvea.pdf>>. Acesso em: 24 jun. 2022.

HOHLFELDT, A. Os estudos sobre a hipótese de agendamento. **Revista FAMECOS**, v. 4, n. 7, p. 42-51, 9 abr. 2008.

IBM CLOUD EDUCATION. **Application Programming Interface (API)**. [S. l.], 19 ago. 2020. Disponível em: <<https://www.ibm.com/cloud/learn/api>>. Acesso em: 2 jul. 2022.

ISP-RJ - Instituto de Segurança Pública da Secretaria de Segurança do Estado do Rio de Janeiro. Comparativo de Áreas – Janeiro de 2021 até Dezembro de 2021. Disponível em: <<http://www.ispvisualizacao.rj.gov.br/Monitoramento.html>>. Acesso em: 3 jul. 2022.

KOHN, Karen; MORAES, Claudia Herte. O impacto das novas tecnologias na sociedade: conceitos e características da Sociedade da Informação e da Sociedade Digital Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. **CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO**, 30. Santos, 29 de agosto a 2 de setembro de 2007. Disponível em: <<http://www.intercom.org.br/papers/nacionais/2007/resumos/R1533-1.pdf>>. Acesso em: 15 jun. 2022.

Lima, Ana C. E. S.; de Castro, Leandro N. (2012). [IEEE 2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN) - Sao Carlos, Brazil (2012.11.21-2012.11.23)] **2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN)** - Automatic sentiment analysis of Twitter messages. (), 52–57. doi:10.1109/CASoN.2012.6412377. Disponível em: <https://www.researchgate.net/publication/236161884_Automatic_Sentiment_Analysis_of_Twitter_messages>. Acesso em: 28 jun. 2022.

LIMA, Roberto Kant. [ET al.] org. Reflexões sobre segurança pública e justiça criminal numa perspectiva comparada. Brasília: **Secretaria Especial de Direitos Humanos**, 2008.

LOTITO, Márcia Padilha. COMUNIDADE VIRTUAL DE APRENDIZAGEM: O AMBIENTE DO PORTAL EDUCAREDE. **Congresso Internacional de Educação a Distância**, [s. l.], 13 maio 2005. Disponível em: <<http://www.abed.org.br/congresso2005/por/pdf/104tcc1.pdf>>. Acesso em: 23 jun. 2022.

MUNIZ, J. Registros de ocorrência da PCERJ como fonte de informações criminais. In: FÓRUM DE DEBATES CRIMINALIDADE, VIOLÊNCIA E SEGURANÇA PÚBLICA NO BRASIL: UMA DISCUSSÃO SOBRE BASES DE DADOS E QUESTÕES METODOLÓGICAS, 1., 2000, Rio de Janeiro. **Anais...** Rio de Janeiro: Ipea; Ucam, 2000. p. 122-144.

MURRAY, J.; DE CASTRO CERQUEIRA, D. R.; KAHN, T. **Crime and violence in Brazil: Systematic review of time trends, prevalence rates and risk factors Aggressive Violent Behavior**, v. 18, p. 5, p. 471–483, 2013

Núcleo de Informação e Coordenação do Ponto BR (NIC.br). (2020). **Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros**: Pesquisa TIC Domicílios, ano 2019. Disponível em: <<https://cetic.br/pt/pesquisa/domicilios/analises/>>. Acesso em: 22 jun. 2022.

OLIVEIRA JÚNIOR, A. Dá para confiar nas polícias? confiança e percepção social da polícia no Brasil. **Revista Brasileira de Segurança Pública**, São Paulo, v. 5, n. 9, p. 6-23, 2011.

PINO, Angel. VIOLÊNCIA, EDUCAÇÃO E SOCIEDADE: UM OLHAR SOBRE O BRASIL CONTEMPORÂNEO. **EDUCAÇÃO & SOCIEDADE**, Campinas, v. 28, n. 100, p. 763-785, 1 out. 2007. <Disponível em: <http://www.cedes.unicamp.br>>. Acesso em: 22 jun. 2022.

R Core Team. R: **A Language and Environment for Statistical Computing**. Vienna, Austria, 2014. Disponível em: <<https://www.r-project.org/>>. Acesso em 12 dez. 2021.

ROLIM, Marcos. **A síndrome da rainha vermelha**. [S. l.]: Editora Jorge Zahar, 2006. 311 p. ISBN 85-7110-917-6. Disponível em: <https://www.academia.edu/36231412/ROLIM_Marcos_A_S%C3%ADndrome_da_Rainha_Vermelha>. Acesso em: 2 jul. 2022.

SCHLEMMER, Eliane; CARVALHO, José Oscar Fontaninide. GESTÃO DE UM CONSÓRCIO NACIONAL PARA EDUCAÇÃO A DISTÂNCIA ORGANIZADO NA FORMA DE COMUNIDADE VIRTUAL DE APRENDIZAGEM: A ESTRATÉGIA DA CVA-RICESU. **Revista Digital da CVA - Ricesu**, [s. l.], v. 3, n. 10, 11 2005. Disponível em: <<https://docplayer.com.br/58179235-Gestao-de-um-consorcio-nacional-para-educacao-a-distancia-organizado-na-forma-de-comunidade-virtual-de-aprendizagem-a-estrategia-da-cva-ricesu.html>>. Acesso em: 15 jun. 2022.

SENASP; CRISP; **DATAFOLHA**. Pesquisa Nacional de Vitimização. In: Pesquisa Nacional de Vitimização. [S. l.], 2013. Disponível em:

<https://www.crisp.ufmg.br/wpcontent/uploads/2013/10/Relat%C3%B3rio-PNV-Senasp_final.pdf>. Acesso em: 22 abr. 2022.

SILGE, Julia; ROBINSON, David. **Introduction to tidytext**. [S. l.], 9 maio 2022. Disponível em: <<https://cran.r-project.org/web/packages/tidytext/vignettes/tidytext.html>>. Acesso em: 2 jul. 2022.

Sinclair, J. 2005. "**Corpus and Text - Basic Principles**" in **Developing Linguistic Corpora: a Guide to Good Practice**, ed. M. Wynne. Oxford: Oxbow Books: 1-16. Disponível em: <http://icar.cnrs.fr/ecole_thematique/contaci/documents/Baude/wynne.pdf>. Acesso em: 25 jul. 2022.

SOARES, Andréia; BORGES, Doriam; CAMPAGNAC, Vanessa. A Pesquisa de Condições de Vida e Vitimização de 2007: notas metodológicas. In: PINTO, Andréia Soares; CAMPAGNAC, Vanessa (Org.) **Pesquisa de condições de vida e vitimização** – 2007. Rio de Janeiro: ISP, 2008, p. 10-32.

Sukanya, M.; Biruntha, S. (2012). [IEEE 2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT) - Ramanathapuram, India (2012.08.23-2012.08.25)] 2012 **IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)** - Techniques on text mining., (), 269–271. Doi:10.1109/ICACCCT.2012.6320784. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/6320784>>. Acesso em: 15 jun. 2022.

TWITTER. **API 11.0 Twitter**. 2022. <<https://developer.twitter.com/en/docs/twitter-ads-api/versioning>>. Acessado em 17 jun. 2022.

TWITTER: Veja a cronologia do Twitter. **Folha de São Paulo**, 25 abr. 2022. Disponível em: <<https://www1.folha.uol.com.br/mercado/2022/04/veja-a-cronologia-do-twitter.shtml#:~:text=A%20rede%20foi%20criada%20em,Dorsey%20em%20mar%C3%A7o%20daquele%20ano.&text=Inicialmente%2C%20o%20servi%C3%A7o%20seria%20chamado,e%20seria%20comprado%20apenas%20depois>>. Acesso em: 17 jun. 2022.

TWITTER. **Quarterly results**. [S. l.], 2022. Disponível em: <<https://investor.twitterinc.com/financial-information/quarterly-results/default.aspx>>. Acesso em: 27 jun. 2022.

VAZ, Lilian Fessier. MODERNIDADE NA CIDADE E NA HABITAÇÃO. **Revista Interfaces: Modernidade e Exclusão**, [s. l.], n. 3, 06 1997. Disponível em: <<https://revistas.ufrj.br/index.php/interfaces/article/view/32788/18409>>. Acesso em: 17 jun. 2022.

VICINITAS. [S. l.], [2017]. Disponível em: <<https://www.vicinitas.io/>>. Acesso em: 12 dez. 2021.

WENGER, Etienne. **Communities of practice: a brief introduction**. Disponível em: <<https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/11736/A%20brief%20introduction%20to%20CoP.pdf?sequence=1&isAllowed=y>>. Acesso em: 23 jun. 2022.

ZAUBERMAN, Renée. **As pesquisas de vitimização na França. In: Reflexões sobre segurança pública e justiça criminal numa perspectiva comparada**. Robert Kant de Lima... [ET al.] (org.). – Brasília: Secretaria Especial de Direitos Humanos, 2008.

ANEXOS

ANEXO A – PALAVRAS UTILIZADAS PARA BUSCA DA BASE PRIMÁRIA

Tabela 4 - Palavras chaves utilizadas na busca da primeira base de dados

<i>assalto</i>	<i>estupro</i>
<i>assaltando</i>	<i>estuprou</i>
<i>assaltaram</i>	<i>estupraram</i>
<i>assaltado</i>	<i>estuprando</i>
<i>assaltou</i>	
	<i>crime</i>
<i>furto</i>	<i>criminoso</i>
<i>furtando</i>	<i>espancou</i>
<i>furtaram</i>	<i>espanca</i>
<i>furtado</i>	<i>espancaram</i>
<i>furtou</i>	<i>espancando</i>
<i>tiro</i>	<i>agride</i>
<i>tiros</i>	<i>agressão</i>
<i>tiroteio</i>	<i>agrediu</i>
<i>atirando</i>	<i>agredindo</i>
<i>atirou</i>	<i>vândalo</i>
<i>atiraram</i>	
<i>atiram</i>	<i>assassinou</i>
	<i>assassinaram</i>
<i>matou</i>	<i>assassinato</i>
<i>mataram</i>	<i>assassinando</i>
<i>matando</i>	<i>baleados</i>
<i>matar</i>	<i>baleado</i>
<i>feminicídio</i>	
<i>sequestra</i>	<i>sequestros</i>
<i>sequestrador</i>	<i>sequestrou</i>
<i>sequestradores</i>	<i>sequestrado</i>
<i>sequestraram</i>	<i>sequestrados</i>
<i>sequestro</i>	<i>sequestrada</i>
<i>arrastão</i>	<i>sequestradas</i>

Fonte: Elaborado pelo autor

ANEXO B – LISTA DE STOPWORDS

Tabela 5 – Lista de Stopwords

leia	uma	estiverem	minha	3	não
janeiro	no	hei	numa	23	uma
día	em	há	pelos	18	os
após	os	havemos	elas	16	no
rio	da	hão	qual	2	se
vila	pelo	houve	nós	@informe_rj	na
5	ao	houvemos	lhe	padre	por
13	mas	houveram	deles	miguel	mais
x	nos	houvera	essas	10	as
14	na	houvéramos	esses	1	dos
22	ser	haja	pelas	30	como
17	as	hajamos	este	rjinformenews	mas
28	por	hajam	dele	datempodeimpedi	ao
55h	tambe	houvesse	tu	r	ele
24h	m	houvéssemo	te	crimesnews	das
25	-	s	vocês	comunidadesrj	à
19	;	houvessem	vos	rt	seu
29	:	houver	lhes	fogocruzadorj	sua
26	alem	houvermos	meus	tirosrj	ou
6	so	houverem	minhas	rjplantaio	quando
00h	ate	houverei	teu	legal_rj	muito
10h	pra	houverá	tua	videos	nos
50h	.	houveremos	teus	confirmações	já
2253	'	houverão	tuas	fotos	eu
22h)	houveria	nosso	videos	também
24	(houveríamos	nossa	avudmtuvs2	só
27	de	houveriam	nossos	r0rp00mbmh	pelo
23h	a	sou	nossas	cnaxqgrwjc	pela
17h	o	somos	dela	informenews	até
20h	que	são	delas	ottrj	isso
00h	e	éramos	esta	ott	ela
29	do	eram	estes	i	entre
09h	da	fui	estas	ii	depois
19h	em	foi	aquele	iii	sem
informa	um	fomos	aquela	iv	mesmo
av	para	fora	aquelas	^v\$	seus
@radarbrasil	não	fôramos	isto	vii	quem
@ddalertario	uma	seja	aquilo	viii	nas
lá	os	sejamos	estou	ix	me
@rjplantaio	no	sejam	está	art	esse
	se			e	

#fogocruzado	na	fosse	estamos	de	eles
15	por	fôssemos	estão	o	você
video	mais	fossem	estive	a	essa
informelegarj	as	for	estive	com	num
esquinadotrafico	dos	formos	estivemos	que	nem
40h	como	forem	estiveram	do	suas
ocorrenciarj	mas	serei	estava	um	meu
jornalextra	ao	será	estávamos	para	às
j	ele	seremos	estavam	este	fora
ddnautas	das	serão	estivera	dele	fôramos
30h	à	seria	estivéramos	tu	seja
212253	seu	seríamos	esteja	te	sejamos
helpriodjaneiro	sua	seriam	estejamos	vocês	sejam
papodefavelaofc	ou	tenho	estejam	vos	fosse
fiquei	quando	tem	estivesse	lhes	fôssemos
radardabrasil	muito	temos	estivéssemos	meus	fossem
informe_rj	nos	tém	estivessem	minhas	for
tá	já	tinha	estiver	teu	formos
santatrsemfoco	eu	tínhamos	estivermos	tua	forem
bailefunkdotrafico	també	tinham	estiverem	teus	serei
app	m	tive	hei	tuas	será
padre_miguel	só	teve	há	nosso	seremos
leisecarj	pelo	tivemos	havemos	nossa	serão
dora	pela	tiveram	hão	nossos	seria
stemfoco	até	tivera	houve	nossas	seríamos
0	isso	tivéramos	houvemos	dela	seriam
#ottrj	ela	tenha	houveram	delas	tenho
https://t.co/cnaxqgrwj	entre	tenhamos	houvera	esta	tem
c	depois	tenham	houvéramos	estes	temos
2021	sem	tivesse	haja	estas	tém
7	mesmo	tivéssemos	hajamos	aquele	tinha
é	aos	tivessem	hajam	aquele	tínhamos
@legal_rj	seus	tiver	houvesse	aqueles	tinham
4	quem	tivermos	houvéssemo	aqueles	tive
1	nas	tiverem	s	isto	teve
11	me	terei	houvessem	aquilo	tivemos
12	esse	terá	houver	estou	tiveram
20	eles	teremos	houvermos	está	tivera
30	você	terão	houverei	estamos	tivéramos
8	essa	teria	houverá	estão	tenha
21	num	teríamos	houveremos	estive	tenhamos
5	nem	teriam	houverão	estive	tenham
3	suas	"	houveria	estivemos	tivesse
2	meu	,	houveríamos	estiveram	tivéssemo
#crimesnews	às				s

@leisecarj	minha	de	houveriam	estava	tivessem
9	numa	a	sou	estávamos	tiver
@comunidadesrj	pelos	o	somos	estavam	tivermos
2020	elas	que	são	estivera	tiverem
ott-rj	qual	e	era	estivéramos	terei
#informenewsrlj	nós	do	éramos	esteja	terá
aplicativo	lhe	da	eram	estejamos	teremos
@pmerj	deles	em	fui	estejam	terão
#dora	essas	um	foi	estivesse	teria
acesse	esses	para	fomos	estivéssemos	teríamos
1177	pelas	com	foram	estivessem	teriam
				estiver	"
				estivermos	,

Fonte: Elaborado pelo autor

ANEXO C – EXEMPLOS DE DADOS DAS BASES ANALISADAS

Quadro 2 - Exemplos de dados da base Comunidades Virtuais

Tweet	Profile Name	Created At
RT @AlertaCascadura: Muitos tiros estão sendo ouvidos na região de Thomas Coelho e Eng da Rainha. Evitem a região. Clima muito tenso.	Alerta Assalto RJ	Sat Oct 24 01:47:47 +0000 2015
RT @PadreMiguelNews: Carro roubado nesta madrugada na Av. Brasil , altura de Parada de Lucas em arrastão https://t.co/RbqPKdHiQm	Alerta Assalto RJ	Tue Feb 09 21:55:03 +0000 2016
Arrastão no Túnel Santa Barbara, Catumbi.	Catumbi Alerta	Mon Apr 09 01:42:00 +0000 2018
Tiros no Morro São João, Engenho Novo. Atenção na Barão de Bom Retiro. Atividade na região.	Alerta Assalto RJ	Tue Apr 24 02:01:05 +0000 2018
OTT-RJ INFORMA : 06/07 - 02:44h Continuam os tiros em Senador Camará. Cuidado na região. #OTTRJ https://t.co/zw5jmKahgS	Onde Tem Tiroteio-RJ	Sat Jul 06 05:47:19 +0000 2019
OTT-RJ INFORMA : 06/07 - 06:30h Tiros na Vila Kennedy localidades Leão e Progresso. Cuidado na região. #OTTRJ	Onde Tem Tiroteio-RJ	Sat Jul 06 10:13:51 +0000 2019
 Piedade  Mais um flagrante de assalto, desta vez na Rua Maria Vargas, próximo ao IPEM. https://t.co/aOlyQsjSSi	 Informe Legal - Rio de Janeiro 	Wed Apr 07 16:00:27 +0000 2021
Polícia entrou em confronto na madrugada contra o CV na Vila Kennedy Leia: https://t.co/wnY6XwFS4T #crimesnews	Blog Crimes News Oficial ©	Sun Aug 01 15:55:59 +0000 2021

Fonte: Elaborado pelo autor

Quadro 3 - Exemplos de dados das bases das polícias militar, civil e guarda municipal

Text	Profile Name	Created At
Policciamento do #40BPM na antiga Estrada Rio - São Paulo com Estrada do Tingui, no bairro de Campo Grande. #ZonaOeste #PMERJ https://t.co/a79TOYeYI8	PMERJ	2021-11-29 22:45:00 UTC
Conte com nossos policiais da #UPP Tabajaras que estão posicionados na rua Real Grandeza com Aníbal Reis, em #Botafogo. #PMERJ #ZonaSul https://t.co/SCWgL62EIL	PMERJ	2021-11-29 11:30:00 UTC
Estamos atuando com policiais da #UPPBarreiraTuiuti na Rua Bela, em #SãoCristóvão. #PMERJ #ServireProteger #PolicciamentoOstensivo https://t.co/UVhIsqWQNI	PMERJ	2021-11-29 04:04:00 UTC
Equipes do #18BPM patrulham a Av. Menezes Côrtes (autoestrada #Grajaú-#Jacarepaguá). A via segue com policiamento reforçado por várias unidades e sem qualquer tipo de alteração. #PMERJ #ServireProteger #PolicciamentoOstensivo https://t.co/R107Ja9dDG	PMERJ	2021-11-28 23:57:58 UTC
#93DP e #PRF prenderam dois integrantes de uma organização criminosa que aplicava golpe em venda de tinta e enganava pessoas na Região Sul Fluminense. Eles foram capturados em Volta Redonda. https://t.co/tZtBRJS1tn	PCERJ	2021-11-29 18:35:12 UTC
#DPMA realizou operação para verificar denúncias de maus-tratos de animais domésticos, silvestres e exóticos. Na ação, sete pessoas foram detidas e conduzidas à unidade policial para esclarecimentos. Centenas de aves foram resgatadas. Saiba mais: https://t.co/NLD9toLyyW https://t.co/151FRGU7Nx	PCERJ	2021-10-05 22:00:11 UTC
#DCOD prendeu em flagrante um lutador de MMA que traficava drogas por meio de aplicativo de mensagens. Levantamento realizado pelo Setor de Inteligência da unidade apontou que ele fazia entregas dos entorpecentes na Barra da Tijuca. https://t.co/CDptTV8tUm	PCERJ	2021-08-29 14:44:21 UTC
#DCPolinter prendeu, nesta sexta-feira (06/08), um homem condenado pelo crime de roubo majorado pelo emprego de arma de fogo e concurso de pessoas. Ele participou de um roubo a um caminhão dos Correios em 2015. https://t.co/WZDQFI7tCD	PCERJ	2021-08-06 21:41:44 UTC
O Grupamento Tático Móvel (GTM) foi criado para reforçar as ações de patrulhamento realizadas pela Guarda Municipal. A unidade também atua em situações urgentes, de rápida resposta. #SomosAGuarda https://t.co/1BtRBtW2fh	GMRio	2021-11-29 17:24:01 UTC
Patrulhamento preventivo realizado pelo Grupamento de Operações com Cães da @GMRio nas proximidades da antiga estação da Leopoldina, região central da cidade. https://t.co/l3rdg8KAFg	GMRio	2021-11-26 17:14:35 UTC
Equipes da @GMRio atuando nesta sexta-feira, dia 1º, no patrulhamento preventivo da Ilha do Governador, bairro da Zona Norte da cidade. #SomosaGuarda #Conteconosco https://t.co/EYXwSOzCFm	GMRio	2021-11-22 19:07:49 UTC
Equipes da @GMRio seguem atuando na Rua Pompeu Loureiro, em Copacabana, em apoio à ocorrência de princípio de incêndio no Hospital São Lucas, nesta quarta-feira, dia 08. https://t.co/djJfgH2FA4	GMRio	2021-11-03 19:40:05 UTC

Fonte: Elaborado pelo autor