

NVivo e Twitter: da coleta de dados à exploração do dataset #coronavirusnobrasil

NVivo and Twitter: from data collection to the exploration of the #coronavirusnobrasil dataset

Sandra Depexe¹ⁱ
ORCID: <https://orcid.org/0000-0001-9536-4285>

Recebido em: 14/03/2022. Aprovado em: 20/04/2022.

Resumo

Com o objetivo de contribuir com o campo das ciências da comunicação, este texto apresenta notas de uso do software NVivo para investigação exploratória de dataset da hashtag #coronavirusnobrasil no Twitter. Aborda pesquisa qualitativa auxiliada por computador, com atenção à descrição da coleta de dados, formação de datasets e incursão às ferramentas de visualização de dados nativas do software. Articula os achados em diálogo ao fazer metodológico operacional, indicando as potencialidades e limitações do software.

Palavras-chave: Twitter; NVivo; pesquisa qualitativa; Covid-19.

Abstract

In order to contribute to the field of Communication Sciences, this article discusses the use of the NVivo software for exploratory investigation of the hashtag #coronavirusnobrasil (Corona virus in Brazil) dataset on Twitter. It debates computer-aided qualitative research, emphasizing data collection, dataset formation and the software's data visualization tools. The findings are discussed in relation to the operationalization of methodology, highlighting possibilities and limitations of the software for this purpose.

Keywords: Twitter; NVivo; qualitative research; Covid-19.

Introdução

O artigo relata experiência de uso do software de pesquisa qualitativa NVivo² para investigação exploratória de dataset formado a partir de mensagens em circulação com hashtag #coronavirusnobrasil na rede social digital Twitter. Os primeiros casos da nova síndrome respiratória causada por coronavírus, nominada como Covid-19 pela Organização Mundial da Saúde, foram relatados na China ao final do mês de dezembro de 2019. A propagação do vírus resultou em uma pandemia mundial (CHEN;

¹Universidade Federal de Santa Maria, Rio Grande do Sul, Brasil. E-mail: sandra.depexe@ufsm.br

² Software proprietário da QSR International Pty Ltd.. Informações disponíveis em: <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>

LERMAN; FERRARA, 2020; JAHANBIN; RAHMANIAN, 2020). No Brasil, em 26 de fevereiro de 2020, houve a confirmação do primeiro caso de Covid-19 e a *hashtag* #*coronavirusnobrasil* foi observada no Twitter a partir desta data³.

Embora o Twitter não apresente elevado crescimento na sua base de usuários e possua um alcance inferior ao Facebook no Brasil – são 15,5% contra 86,5% da população digital⁴ (GRUPO..., 2018) -, sua importância em ações de engajamento da audiência televisiva ou como plataforma de manifestações políticas no país é reconhecida. Assim como justificam Santaella e Lemos (2010, p. 9), acreditamos que “a peculiaridade das características do Twitter traz atributos inesperados [...] como um mirante privilegiado para pensar as questões teóricas, críticas e práticas que as redes digitais estão trazendo”.

As manifestações no Twitter sobre a pandemia de Covid-19 e a estratégia de distanciamento social nos interpelam a explorar os usos e apropriações desta rede como potência indiciária de hábitos, relações de poder e consumo midiático contemporâneo. Algumas iniciativas científicas, vinculadas à área de saúde, ciências comunicação e da informação, atentam à circulação de informações e a disseminação de desinformação (KOUZY *et al.*, 2020; LOPEZ; VASU; GALLEMORE, 2020; RECUERO, SOARES, 2021) e às conversações sobre Covid-19 no Twitter (ALSHAABI *et al.*, 2020; CHEN; LERMAN; FERRARA, 2020; JAHANBIN; RAHMANIAN, 2020). Assumimos, neste texto, o viés exploratório de pesquisa (BONIN, 2016) como procedimento essencial para construção de desenho de pesquisa empírico futuro. Por ora, a problemática guia deste texto está centrada na seguinte questão: Como o software NVivo pode contribuir com a coleta e o mapeamento de dados de um *dataset* do Twitter? A opção pelo NVivo se justifica por ser um software que não requer conhecimento em programação para efetuar a coleta de dados em redes sociais.

Compreendemos com Knafl (2018, p.2) que “a tecnologia permitiu acumular volumes de dados cada vez maiores e há um também crescente desejo associado de entendê-los”. Logo, assumimos que o modo de olhar para o empírico não é isento de um aporte teórico, metodológico e técnico (LOPES, 2010) que possibilite pensá-lo, para além de métricas (dados numéricos), como materialidade discursiva, social e cultural. Justamente, a preocupação em investigar a textualidade como constructo social põe na

³ A observação se deu durante coleta de dados para outra pesquisa, em andamento.

⁴ Estima-se que aproximadamente 61% dos domicílios no Brasil possuem acesso à internet e o país ocupa o quarto lugar mundial no número de usuários de internet (GRUPO..., 2018).

mira não apenas os estudos discursivos, em suas variadas vertentes, como também aponta para a urgência em capturar e manejar, com fins científicos, essas materialidades propagadas sob a forma de postagens e comentários nas redes (ARAÚJO, 2019; DIAS, 2012; RECUERO, 2018).

Com o objetivo de contribuir com o campo das ciências da comunicação, este texto apresenta notas de uso do software de pesquisa qualitativa NVivo como instrumental para investigação da *hashtag* #coronavirusbrasil. A ênfase, como objetivos específicos, está na descrição das rotinas operacionais, da coleta e formação de *datasets* até a visualização de dados como etapa exploratória. Conhecer as potencialidades e restrições de um software de pesquisa, como parte do plano metodológico (FLICK, 2009a, LOPES, 2010; MARTINO, 2018), permite ao pesquisador compreender limitações da investigação desde a coleta e organização dos dados. Adiante, ampliaremos a relação entre *software* e pesquisador a partir da pesquisa qualitativa auxiliada por computador - CAQDAS (*computer aided qualitative data analysis softwares*).

Pesquisa qualitativa auxiliada por computador

Após duas décadas às margens do processamento eletrônico, conforme Kelle (2008), o desenvolvimento de pesquisa qualitativa auxiliada por computador, teve início em 1980 quando pesquisadores com conhecimento e experiência em computação, passaram a desenvolver de forma independente softwares que poderiam auxiliar a análise de dados qualitativos. Teixeira (2011) relata que no Brasil a CAQDAS começou a ser utilizada na metade dos anos de 1990 e, até hoje, trava o duelo entre o fascínio e a decepção, pois exige do pesquisador tanto o domínio técnico da ferramenta quanto o domínio de suas questões de pesquisa.

A noção de que os softwares de CAQDAS não fornecem as respostas ao pesquisador é enfatizada por diversos autores (AZEEM; SALFI, 2012; FLICK, 2009a; JACKS *et al.*, 2016; KELLE, 2008; LAGE; GODOY, 2008; MOZZATO; GRZYBOSKI; TEIXEIRA, 2016; OROZCO GÓMEZ; GONZÁLEZ REYES, 2011; TEIXEIRA, 2011; TEIXEIRA; BECKER, 2001), os quais cancelam o papel cognitivo e subjetivo humano frente ao aparato informacional.

Grijó (2013) reforça a necessidade de o pesquisador investir tempo na aprendizagem do software a ser utilizado, para torná-lo potencialmente útil frente às

particularidades de cada investigação. Jacks *et al.* (2016) salientam que o uso dos softwares exige maior organização dos processos de pesquisa e pode requerer a adaptação de práticas analógicas de investigação.

Conforme Teixeira (2011, p.4), os softwares permitem “analisar todo tipo de material passível de ser transformado em texto”, tais como diários de campo, transcrições de entrevistas, grupos focais, programas rádio ou televisão, documentos diversos e material coletado na internet. A lógica de funcionamento dos programas de CAQDAS está na estruturação de um banco de dados textual, em que os registros são indexados e posteriormente recuperados. Embora a compreensão do sentido do texto não possa ser dada por um recurso computacional, as tarefas mecânicas implicadas na análise textual são otimizadas (KELLE, 2008).

O NVivo, assim como a maioria dos programas de CAQDAS, opera pelo conceito de projeto, o qual é formado por um conjunto de dados (LAGE, 2011; LAGE; GODOY, 2008). Em um projeto do NVivo⁵ há três instâncias básicas de gerenciamento e estruturação da base de dados: fontes de dados; nós e codificação (indexação); classificações e atributos.

As fontes são os materiais empíricos, em formatos compatíveis, a serem importados ao projeto, tal como: arquivos de texto, planilhas, imagens, vídeos, áudios e arquivos originários de fontes externas⁶. Em nosso caso, as fontes são os arquivos obtidos através de coletas realizadas com a extensão de navegador web NCapture, a qual é fornecida gratuitamente pela QSR International, mesma desenvolvedora do NVivo. Cabe ressaltar que nem todos os softwares de CAQDAS possuem suporte à coleta de dados, sendo este um dos diferenciais do NVivo. Com o NCapture é possível realizar captura de dados como *datasets* de redes sociais (Twitter, Facebook, YouTube) ou páginas da web como PDF. Nas próximas seções deste texto, relataremos, em detalhes, a formação e uso de *datasets* com dados extraídos do Twitter com a *hashtag* *#coronavirusbrasil*.

A indexação se refere à *codificação* destes materiais em *nós*, isto é, “a ação de relacionar passagens do texto a categorias que o pesquisador ou já desenvolveu anteriormente, ou irá desenvolver para o caso específico” (KELLE, 2008, p.397). Os *nós* representam a estrutura de codificação e são relacionados às questões de pesquisa,

⁵ As funcionalidades descritas neste texto são comuns às versões 10, 11 e 12 do software.

⁶ Como exemplos de fontes externas estão arquivos de softwares de referência, blocos de anotações, respostas de pesquisa realizada com Survey Monkey ou arquivos de coletas do NCapture.

pois “Nodes - containers of categories and coding, represent abstract ideas, concepts, people, places, processes, or any other category established documents for project⁷” (AZEEM; SALFI, 2012, p.265). Os *nós* também são úteis para agrupar resultados de consultas e auxiliar as etapas de seleção de *corpus*.

Já as *classificações* são tabelas de dados formadas por *atributos* (variáveis) que são vinculadas a *casos* relativos a um conjunto de fontes ou *nós*. As classificações de *nós* podem servir para categorizar as unidades de análise (casos) em termos de seus atributos. Logo, por exemplo, “as entrevistas realizadas com uma determinada pessoa viram um nó (ou caso). Estas entrevistas, então, podem ser classificadas segundo alguns atributos como: idade, posição política, renda, etc. Assim, todas as falas de um entrevistado específico ficam marcadas por estes atributos” (AMES, 2013, p. 235). No caso de classificações de fontes, podem organizar as variáveis relativas às fontes, como data de realização da entrevista, local, duração, etc.

A partir destas instâncias, não necessariamente de forma concomitante, podemos realizar buscas, classificações, codificações, cruzamentos de dados e criar matrizes. Além de sua função como banco de dados para pesquisa, “suas funcionalidades variam desde análises simples, como estatística descritiva e contagem de palavras até procedimentos mais complexos como elaboração de mapas de conexão e a estimação de modelos de séries temporais” (ALVES; FIGUEIREDO FILHO; HENRIQUE, 2016, p. 124). Relataremos, a seguir, experiências e possibilidades de uso do software NVivo para coletar e explorar a *hashtag* #coronavirusbrasil.

#coronavirusbrasil: coleta de dados com o NVivo

Apreendemos com Lopes (2010) que as opções metodológicas são definidas com base em quatro instâncias: *epistemológica*, a qual exerce a função de vigilância crítica da pesquisa; *teórica*, que diz respeito à teoria; *metodológica ou metódica*, que se refere à seleção de métodos de análise na pesquisa; e *operacional ou técnica*, que corresponde tanto ao domínio do repertório metodológico, às considerações sobre o uso social dos resultados da investigação, quanto à limitação imposta pela dimensão dos recursos, sejam estes materiais, financeiros, temporais ou pessoais para o desenvolvimento da pesquisa. Iremos nos ater ao último plano, tendo em vista, como já

⁷ Tradução nossa: Nós - recipientes de categorias e de codificação, representam ideias abstratas, conceitos, pessoas, lugares, processos ou qualquer outra categoria estabelecida para os documentos do projeto.

abordado, que conhecer as características um software de pesquisa, como o NVivo, permite ao pesquisador compreender e antever limitações da investigação.

Consideramos que a instância técnica “é o lugar da construção dos dados ou do objeto empírico. Compreende os procedimentos de coleta das informações e das transformações destas em dados pertinentes à problemática geral” (LOPES, 2010, p.128).Ao posicionarmos o escopo de investigação à pesquisa qualitativa, ressaltamos que seu rigor “é garantido pela atenção de quem pesquisa em conseguir informações, comparando, analisando e selecionando *ainda na fase de coleta de dados* para verificar a qualidade do que tem em mãos” (MARTINO, 2018, p. 102 grifos do autor). Essa atenção à coleta de dados, em nosso entendimento, é essencial nas investigações em meios digitais, pois como alertam Flick (2009a), Fragoso, Recuero e Amaral, (2011), as dificuldades de recorte variam em escala, heterogeneidade e dinamismo da rede. Ademais, reconhecemos com Jacks *et al.* (2017) que as mudanças na realidade sociocultural, mediadas pelas tecnologias digitais, tornam visíveis e necessários os esforços de diversas perspectivas teóricas, mesmo que espreitadas pelo risco de desatualização, em apreender e registrar processos em curso.

Como os limites do campo “não podem ser definidos com antecedência, uma vez que apenas tornam-se claros durante o estudo” (FLICK, 2009a, p. 247), a importância em considerar o aparato tecnológico como instrumento de coleta e exploração de dados, reflete no processo de (re)formulação do desenho de pesquisa. Conforme Flick (2009b, p.73 grifos do autor), um desenho de pesquisa qualitativa “deve, ao mesmo tempo, ser *sensível, flexível e ajustável* às condições em campo, e, nesse sentido, estar *aberto a novas visões* resultantes dos primeiros passos ou durante o processo de pesquisa”.Logo, a vigilância epistemológica, teórica, metodológica e operacional, debatida por Lopes (2010), carrega uma gramática, referendada por Bonin (2016), marcada por algoritmos, métricas, gráfos, dados quantificáveis e qualificáveis em volume extenso.

Para explorar a *hashtag#coronavirusbrasil* foram realizadas coletas de *tweets* por meio da ferramenta NCapture, nos dias 1º de março de 2020, 10, 20 e 25 de abril de 2020⁸. Esta atividade requer comando manual do pesquisador diretamente na plataforma web do Twitter⁹ para, então, realizar os registros. A vantagem em utilizar

⁸A coleta realizada em 1º de março de 2020 remete ao início dos casos de Covid-19 no Brasil. As demais buscaram acompanhar sistematicamente os desdobramentos depois de decretada a situação de calamidade pública em virtude da pandemia no país, em 20 de março de 2020.

⁹ Disponível em: <https://twitter.com/search-home>

essa ferramenta está no detalhamento das informações coletadas¹⁰ em *dataset*: nome de usuário; *tweet*; data e hora; tipo (se é *tweet* ou *retweet*); *retweetado* por (caso seja um RT); número de *retweets* (quantas vezes foi replicado); *hashtags*; menções (a outro usuário); nome; local; web (se o usuário indica algum site ou perfil na web); biografia; número de *tweets* do usuário; número de seguidores; número de pessoas seguindo; e coordenadas do local, os quais são organizados pelo programa em uma planilha. Cada operação de coleta gera um *dataset* em arquivo com extensão *.nvcx*¹¹, a ser posteriormente importado como fonte externa ao projeto no ambiente do NVivo.

O NCapture coleta apenas conteúdo de perfis públicos, ou seja, perfis abertos cuja disponibilidade ao acesso às postagens é irrestrita, não sendo necessário possuir uma conta na rede social para visualizar esse tipo de perfil. Entretanto, o pesquisador deve manter-se vigilante quanto ao caráter privado e pessoal que, ao menos parte, dos *tweets* coletados possuem e sua implicação ética (FLICK, 2009a). A formação de um *dataset* contará com dados de perfis institucionais, celebridades, pessoas públicas e, invariavelmente, pessoas “comuns”. Enquanto os três primeiros podem ser objeto de estudo, pois como figuras públicas “ocupam cargos ou posições que dizem respeito à vida coletiva de uma sociedade [...] bem comum e interesse público, necessitando dar transparência às suas ações e delas prestar contas à coletividade” (FRANÇA, 2014, p. 16-17), os usos das redes sociais pelos demais usuários remetem às práticas de interação midiática atravessadas por relacionamentos, conversações, afetos e, no contexto da pandemia, às formas de minimizar os impactos do isolamento e distanciamento social, tal como mostra Primo (2020). Nesta perspectiva, cabe ao pesquisador atentar às diretrizes da Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709, de 14 de agosto de 2018 modificada pela Lei nº 13.853, de 8 de julho de 2019), de forma a garantir a privacidade e o anonimato da autoria das mensagens, especialmente daquelas oriundas de pessoas não públicas.

Como já abordamos, softwares de CAQDAS operam com dados textuais (KELLE, 2008; TEIXEIRA, 2011), logo o manejo de vocábulos faz parte da vigilância epistemológica (LOPES, 2010), desde a seleção de termos de busca do Twitter e acompanha todo o trabalho de pesquisa, sendo essencial para manutenção da qualidade (MARTINO, 2018) e dos níveis de comparação pretendido (FLICK, 2009b). Como

¹⁰ Quando completados e disponibilizados pelo usuário de perfis públicos na rede.

¹¹ Formato de arquivo atribuído pelo NCapture. A leitura do arquivo ocorre apenas via importação em um projeto no NVivo, ou seja, por importação em um arquivo do tipo QSR NVivo Project (.nvp).

selecionamos para a investigação a *hashtag#coronavirusbrasil*, a qual remete ao domínio linguístico e territorial do Brasil, não recorremos a nenhum tipo de parâmetro especial de busca. Em outras experiências com o NCapture a indefinição de idioma, por exemplo, pode resultar em um denso volume de dados não condizentes a problemáticas de pesquisa, pois algumas *hashtags* ou termos chave têm alcance global. Os trabalhos de Alshaabi *et al.* (2020) e Lopez, Vasu e Gallemore (2020), com *datasets* construídos a partir de palavras-chave como “*virus*”, “*covid*”, “*coronavirus*”, “*Covid-19*”, entre outras, demonstram que a circulação do “novo coronavírus” ocorre em múltiplas linguagens no Twitter. Em ambos os trabalhos, o português ocupa o terceiro lugar no volume dos dados coletados, ficando atrás dos idiomas inglês e espanhol.

O número de mensagens capturadas em um *dataset* é variável e imprevisível, o qual reforça o entendimento de se trata de uma amostra da realidade e não a totalidade de *tweets* sobre um tema específico. Acreditamos que essa especificidade não é problemática, pois em pesquisas qualitativas o eixo norteador aponta para a qualidade do material e o encontro do ponto de saturação das informações (MARTINO, 2018). A coleta, mineração e análise de dados extraídos na internet “mostra que o domínio de *softwares* e de conhecimentos estatísticos deve estar na pauta da produção acadêmica do campo da comunicação” (PIENIZ; SILVA; MATOS, 2017, p. 39), reforçando a integração de abordagens quanti-qualitativas (FLICK, 2009b).

É possível que a criação de um *dataset* retorne, em segundos, dados em tempo real ou do passado, conforme diretrizes do software e de circulação da rede no instante de coleta. Nas experiências com a *hashtag#coronavirusbrasil* a elasticidade temporal resgatou mensagens com até oito dias anteriores à data de coleta. Entretanto, este comportamento não é padrão. A dependência de conexão, do funcionamento da plataforma web do Twitter e a restrição ao número de coletas do NCapture são algumas limitações a serem levadas em conta pelo pesquisador. Como interferência direta, que pode comprometer estratégias de pesquisa, a extensão do NVivo restringe o número de mensagens coletadas. Assim, caso haja um número razoável de *hashtags* ou palavras para busca concomitantes, talvez a extensão acuse que o limite foi excedido e não realize novas capturas por algum tempo, conforme vivenciamos em outras situações. Tais dificuldades não são exclusivas do uso do NVivo e também são mencionadas por pesquisadores que utilizam outros métodos para formação de *datasets* (RECUERO, 2018). Na Tabela 1, apresentamos o resumo quantitativo dos dados coletados.

Tabela 1 - volumetria dos dados coletados

Data de coleta	Período capturado	Tweets	Retweets	Total de comentários
01/03/2020	27/02/2020 – 01/03/2020	1.910	6.569	8.479
10/04/2020	03/04/2020 – 10/04/2020	7.426	10.152	17.578
20/04/2020	12/04/2020 – 20/04/2020	3.588	5.337	8.925
25/04/2020	17/04/2020 – 25/04/2020	2.075	2.137	4.212
Total coletado		14.999	24.195	39.194
Total excluindo duplicatas		14.346	23.454	37.800

Fonte: a autora.

Ao incorporar arquivos em um projeto do NVivo é possível combinar as fontes obtidas pelo NCapture, excluindo automaticamente as duplicatas existentes de um mesmo objeto. Adotamos esta função na importação dos quatro arquivos de *dataset* para a formação da *fonte de dados*, pois há uma sobreposição temporal entre os dias 17 e 20 de abril de 2020, a qual resultou em 1.394 registros duplicados, sendo 653 *tweets* e 741 *retweets*. Esta operação contribui para a fidelidade da investigação, especialmente ao utilizarmos demais recursos do programa que realizam, por exemplo, contagem de palavras. Assim, o corpus é formado por 37.800 mensagens oriundas de 7.426 usuários do Twitter.

A volumetria indicada na Tabela 1 é indiciária do modo como os *retweets* são importantes na estruturação da rede. Um *retweet* é uma mensagem criada por um usuário e compartilhada por outro. Em diferentes vertentes de pesquisa, essas conexões entre usuários, as quais remetem às dinâmicas de circulação e a formação de *hubs* e *clusters*, são examinadas a partir de mensagens replicadas. O índice de *retweets* foi utilizado por Kouzy *et al.* (2020) como princípio para recorte empírico, em que *tweets* com menos de cinco compartilhamentos não compuseram o corpus analítico acerca da disseminação de informações sobre Covid-19. O número de *retweets* alcançado por um *tweet* parece-nos apontar que “*retweets are driven by the content value of a tweet*”¹²(CHA *et al.*, 2010, p. 17).

Embora o arquivo do NCapture categorize as mensagens em *tweets* e *retweets*, bem como inclua o número de *retweets* de *tweets* originais, é preciso alerta de que apenas os *retweets* realizados pelo usuário através do comando “*retweetar*” serão assim identificado pelo NVivo. Se um *tweet* original é copiado e colado em um novo *tweet* e o “RT” é adicionado manualmente pelo usuário, essas mensagens serão tratadas como *tweets* em vez de *retweets*. Além disso, devido ao limite de 280 caracteres no Twitter, às

¹² Tradução nossa: *retweets* são impulsionados pelo valor do conteúdo do *tweet*.

vezes o texto no final de um *retweet* é cortado. Estas particularidades podem impactar o desenvolvimento de pesquisa e exigir estratégias complementares para a categorização ou recuperação das mensagens coletadas.

#coronavirusbrasil: visualização de dados

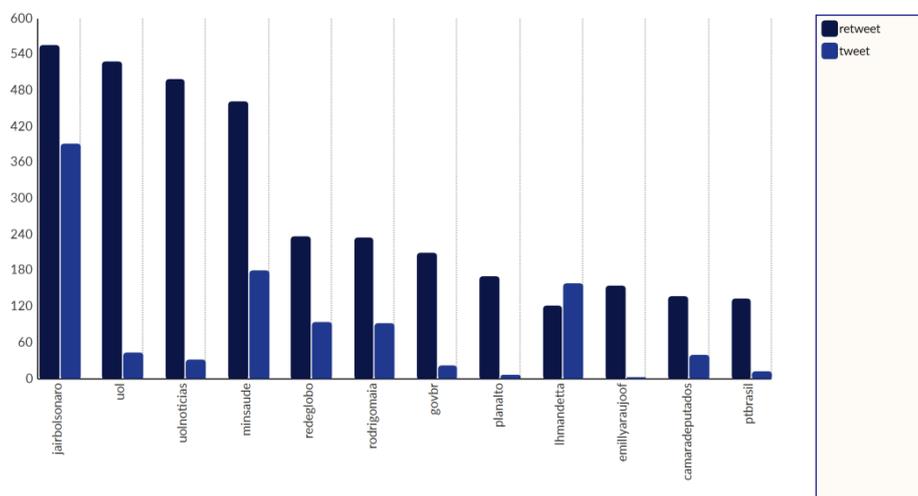
Após detalhar a operacionalização da formação do *dataset* como *fonte de dados* no NVivo, passaremos a descrever procedimentos de incursão inicial ao conteúdo das mensagens, usuários, menções e *hashtags* presentes na amostra. O exercício exploratório de um *dataset* é otimizado por funcionalidades de visualização de dados nativo do software como gráficos, nuvens de palavras, mapa de árvore, dentre outros. A nuvem de palavras é um exemplo de integração quanti-qualitativa, uma vez que sua formação é dada por uma métrica que alude à importância que um termo assume na circulação das mensagens: palavras em tamanho maior correspondem a uma maior frequência na amostra. Embora o NVivo apresente algumas opções visuais, com tipografias, cores e disposição dos textos, cabe ao pesquisador avaliar como estas podem auxiliar ou comprometer os achados, tornando mais tênue a distinção entre visualização de dados e um elemento gráfico decorativo. Ademais, os critérios de seleção como número mínimo de caracteres por palavra ou formas de agrupamento de palavras (derivadas, semelhantes ou sinônimas) conduzem a diferentes modos de representação dos dados.

Para explicitar essas relações, criamos três nuvens de palavras (Figura 1). A primeira, criada a partir das 300 palavras mais frequentes, com três ou mais letras, considerando termos exatos (Figura 1, à esquerda), enfatiza a própria *hashtag* utilizada para coletar os *tweets* e “*https*”, que juntamente a algumas sequências alfanuméricas são indicativas do volume de links compartilhado nas mensagens. Esta nuvem não estabelece uma relação hierárquica entre os demais termos, sendo pouco representativa do que esteve em evidência na circulação das mensagens. Porém, por se tratar da forma visual mais próxima ao comportamento padrão e automático do software, optamos por mantê-la para fins de comparação, reforçando o já dito: o aparato computacional não substitui a atividade do pesquisador na modelagem dos dados.

Como escopo da CAQDAS, o software apresenta uma lista de palavras impedidas, como pronomes e preposições, as quais são ignoradas na execução de alguns comandos. Na segunda nuvem (Figura 1, à direita), incluímos “*#coronavirusbrasil*” e

Assim como Kouzy *et al.*(2020), identificamos que a extratificação de mensagens é originária de indivíduos informais, perfis vinculados ao governo e à mídia. Tomamos como ponto de atenção as referências ao cenário político brasileiro, bem como a veículos de comunicação ou jornalistas dentre as menções mais frequentes na amostra. A visualização segmentada entre *tweets* e *retweets* dos perfis mencionados nas mensagens coletadas (Figura 2) permite levantarmos algumas hipóteses a serem investigadas sobre o uso da rede, dentre as quais a importância que cada um destes atores tem na circulação de informações. Embora existam limitações de formatação, não sendo possível alterar, por exemplo, esquema de cores ou alguns parâmetros da seleção de dados, os gráficos nativos do NVivo auxiliam a análise exploratória.

Figura 2 - Número de referências por menção e tipo de tweet



Fonte: a partir de dataset no software NVivo

O gráfico com os perfis mais mencionados e a percepção de que as menções ao perfil *@jairbolsonaro* são majoritárias nos comentários do Twitter coletados, coadunam com os achados da nuvem de palavras, visto que a *@* do presidente foi a única visível nas imagens e corresponde ao todo 946 menções. Cabe destacar que dentre as mensagens coletadas nenhuma é originária de *@jairbolsonaro*, ou seja, todas as vezes que o perfil de Jair Bolsonaro no Twitter é mencionado é feito por terceiros. Em uma primeira inferência, o presidente não contribui com a circulação da *hashtag* *#coronavirusbrasil*, mas é um dos personagens diretamente mais citados nela, visto que além da menção ao perfil, a palavra “*bolsonaro*” aparece 1.778 vezes e “*#bolsonaro*”, 322 vezes. O segundo perfil mais mencionado (somando *tweets* e

retweets) é o do Ministério da Saúde @minsaude com 641 citações, das quais 49 são autorreferenciais.

A partir desta pista, um segundo gráfico foi gerado, que cruzou as menções mais recorrentes com os perfis de usuários que criaram as mensagens. Como resultado, aproximadamente 48% das menções à @jairbolsonaro concentram-se em três perfis do Twitter relacionados à mídia brasileira, sendo 418 mensagens (*tweets* e *retweets*) oriundas de um veículo de jornalismo independente; 32 de uma agência de notícias especializada em *fact-checking*; e 5 de uma emissora de TV aberta. Em relação ao perfil do Ministério da Saúde, cerca de 36% das menções não escapam de autorreferencialidade do Governo Federal, uma vez que 48 são oriundas do perfil de uma emissora de TV, 132 de um programa radiofônico e 49 do próprio Ministério. Além disso, respectivamente, 15 e 8 *tweets* também são dos mesmos veículos jornalísticos que mencionam o presidente.

Conforme os achados de Kouzy *et al.* (2020), referente ao conteúdo de *tweets* em língua inglesa, “business/NGO/government, news outlets/journalists, and healthcare/public health accounts all had a lower rate of misinformation (6.1%, 18.6%, and 12.3% respectively)”¹³. E no Brasil? Estudo realizado por Recuero e Soares (2021), com dados coletados no Twitter entre 20 e 29 de março de 2020, período próximo ao utilizado neste texto, aponta “que a figura do presidente foi central para o espalhamento dos discursos desinformativos” (p.23) e jornalistas e pesquisadores produziram mensagens para desmentir informações falsas sobre a cura da doença.

Como estratégia exploratória textual, criamos outras três visualizações com as mensagens que remetem a Jair Bolsonaro e circularam sob a marcação #coronavirusbrasil conforme apresentamos anteriormente. Como parâmetros, adotamos as 50 palavras mais frequentes, termos exatos, com três ou mais letras. Embora possa incorrer sobreposição, em observação aos resultados, na Figura 3, é notável a diferenciação do contexto semântico que dá origem às nuvens. As menções diretas ao perfil @jairbolsonaro aludem a apoio ao presidente, enquanto o uso da hashtag “#bolsonaro” e da palavra “bolsonaro” parecem ser mais acionados à crítica e acontecimentos sobre a pandemia, nesta amostra. Especialmente, se atentarmos ao sentido de engajamento das *hashtags*, através da indexação das mensagens, oito *tags* são favoráveis ao presidente: #bolsonaropaidacloroquina, #bolsonaroestamoscontigo,

¹³ Tradução nossa: As empresas / ONG / governo, agências de notícias / jornalistas e contas de saúde / saúde pública tiveram uma menor taxa de desinformação (6,1%, 18,6% e 12,3%, respectivamente).

#unidosporbolsonaro, #bolsonaroestacertosim, #bolsonaroestavacerto, #bolsonarotemrazaa, #euapoiobolsonaro (Figura 3, à esquerda) e quatro, contrárias: #bolsonaroenlouqueceu, #bolsonaroacabou, #forabolsonaro, #bolsonarogenocida (Figura 3, centro e direita).

Figura 3 – Nuvens de palavras alusivas ao presidente



Fonte: a partir de dataset no software NVivo

Ainda que diferentes *hashtags* possam ser utilizadas em uma mesma mensagem, a configuração das nuvens é indiciária de um comportamento diverso. A partir da autocodificação dos comentários e de *matrizes de nós* do *dataset*, verificamos que metade das *hashtags* de apoio a Jair Bolsonaro, que aparecem nas nuvens de palavras, não obtiveram nenhum *retweet*, mesmo havendo acima de 143 *tweets* cada. No lado oposto, as *hashtags* obtiveram entre 22 e 182 *retweets*, sendo este último o volume de compartilhamentos alcançado pela *tag* com menos *tweets* (56), como mostra a Tabela 2. Como anteriormente comentamos, a importância dos *retweets* está na construção de conexões entre usuários, porém as distinções reforçam hipótese de que quando a circulação de #*coronavirusbrasil* passa ao território da polarização política, principalmente pelas *tags* em apoio ao presidente, algumas arestas não se ligam a vértices, ou seja, a rede não se forma.

Tabela 2 – Volumetria *hashtags* pró e contra Jair Bolsonaro

	Hashtag	Tweets	Retweets	Total de comentários
Apoio	# <i>bolsonaroestacertosim</i>	170	0	170
	# <i>bolsonaroestavacerto</i>	187	2	189
	# <i>bolsonaroestamoscontigo</i>	170	0	170
	# <i>bolsonaropaidacloroquina</i>	154	0	154
	# <i>bolsonarotemrazaa</i>	283	17	300

	#bolsonarotemrazaosim	262	16	278
	#euapoiobolsonaro	131	11	142
	#unidosporbolsonaro	143	0	143
Crítica	#bolsonaroacabou	124	22	146
	#bolsonaroenlouqueceu	56	182	238
	#bolsonarogenocida	191	45	236
	#forabolsonaro	315	92	407

Fonte: a partir de dataset no software NVivo

Em operações no software detectamos, por exemplo, que um único usuário¹⁴ criou 210 *tweets*, sendo 142 praticamente idênticos, os quais incluem as oito *tags* de apoio ao presidente conjuntamente. Em mensagens contrárias encontramos fenômeno semelhante, mas em número reduzido: de 52 *tweets* com mesma origem¹⁵, 25 são quase iguais, mas utilizam apenas uma das *tags* contra Jair Bolsonaro. Para além de uma pretensa conversação, estes achados conduzem a questionamentos sobre os usos do Twitter como tática para alcançar *trends topics* e assim ganhar destaque. Somado à alta frequência de postagem, a estratégia de realizar pequenas modificações em cada *tweet*, como inclusão de outras *hashtags* e algumas menções a perfis ou links variados, parecem ser utilizados para burlar os mecanismos de controle do Twitter e sugerem a circulação de uma falsa rede de apoio gerenciada por *bots*: contas automatizadas para replicar mensagens (RECUERO, SOARES, 2021). Acreditamos a Análise de Redes Sociais a partir dos dados extraídos no NVivo, com a formação de grafos, atestariam nossas hipóteses de que #coronavirusbrasil também é palco para desinformação e manipulação: nem todas *hashtags* que sobem aos “assuntos do momento” remetem às noções de cultura da participação (JENKINS, 2009) e da inteligência coletiva (LÉVY, 1998). Ademais, análise discursiva ou análise de conteúdo seriam outras vertentes possíveis para aprofundar as pistas aqui levantadas.

Por fim, ainda que os gráficos e demais ferramentas do NVivo indiquem parâmetros quantitativos da amostra, sua importância reflete na otimização da pesquisa, em especial, na incursão aos dados. Para tanto, a visualização de dados, seja em nuvens de palavras, seja em gráficos, mostra-se útil não apenas na exposição de resultados finais de pesquisa, mas “para compreender os dados e descobrir o que pode ser digno de nota ou interesse”(KNAFLIC, 2018, p. 17). Como sugere Martino (2018), dados

¹⁴ A *bio* recuperada no *dataset* é explicitamente construída em apoio a Jair Bolsonaro. O perfil encontra-se, atualmente, como banido no Twitter.

¹⁵ A *bio* recuperada no *dataset* é explicitamente construída em apoio a Luis Inácio Lula da Silva. O perfil encontra-se, atualmente, ativo no Twitter e alude a uma pessoa não pública.

numéricos, como métricas, aludem às qualidades de um material a serem investigadas qualitativamente.

Considerações finais

Compreendemos que conhecer as potencialidades e restrições de um software de pesquisa permite ao pesquisador perceber limitações da investigação, bem como ponderar se os custos de tempo de aprendizado e de investimento financeiro necessários são compatíveis com o que se pretende extrair dos dados. Neste texto buscamos apresentar notas operacionais de uso do software NVivo, o qual faz parte do escopo CAQDAS, pesquisa qualitativa auxiliada por computador, especialmente útil ao volume de dados textuais para análise.

Como vantagens para adoção do NVivo, citamos a organização do material coletado como projeto, a guisa de um banco de dados; e a formação de *dataset*, apesar desta função possuir algumas restrições, como abordadas ao longo do texto. Outro ponto positivo, mas igualmente limitado, são os modos de visualização, como gráficos e nuvens de palavras, os quais conjugam análise quantitativa à qualitativa e podem ser valiosos na exploração inicial do conjunto de dados. Como aspecto negativo, é preciso mencionar, o fato de ser um software proprietário, com valor vinculado ao dólar, que resulta em alto custo para aquisição ou *upgrade* da licença, além das exigências de um *hardware* com configuração robusta. Embora versões mais recentes do programa possuam funcionalidades afinadas à análise de redes sociais, estas não apresentam, em nosso entendimento, vantagens frente às alternativas de softwares de código aberto.

Quanto aos dados empíricos, reafirmamos que o texto assume o viés exploratório, em que exercitamos desde a coleta e organização dos dados, por meio de ferramentas de visualização de dados nativas do NVivo, sem, entretanto, alcançar todo potencial de análise qualitativa do software. Nesta perspectiva, buscamos indícios sobre usuários, menções e *hashtags* recorrentes na amostra para construção de uma problemática futura, a qual deve ser operacionalizada com a categorização e codificação de mensagens seguindo aportes teórico-metodológicos específicos para trazer luz às inquietações a respeito da construção de sentidos. Apesar deste texto assumir o caráter de notas preliminares de pesquisa, é preciso manter a vigília, visto as referências ao cenário político brasileiro. Afinal, o que se diz sobre Jair Bolsonaro no contexto de *#coronavirusnobrasil?*

Referências

ALSHAABI, T. *et al.* How the world's collective attention is being paid to a pandemic: COVID-19 related 1-gram time series for 24 languages on Twitter. p. 1–13, 2020. Disponível em: <http://arxiv.org/abs/2003.12614>. Acesso em: 26 abr. 2022.

ALVES, Dáfni; FIGUEIREDO FILHO, Dalson; HENRIQUE, Anderson. O poderoso NVivo: uma introdução a partir da análise de conteúdo. **Revista Política Hoje**, v. 24, n. 2, p. 119–134, 2016.

AMES, Valesca Daiana Both. As possibilidades de uso do software de análise qualitativa NVivo. **Sociologias Plurais**, v. 1, n. 2, p. 230–247, 2013.

ARAÚJO, Willian Fernandes. Assim falam as plataformas: uma proposta de análise da constituição de práticas e comportamentos nos textos de plataformas digitais. In: POLIVANOV, BEATRIZ *et al.* (Org.). **Fluxos em redes sociotécnicas**: das micronarrativas ao big data. São Paulo: Intercom, 2019. p. 181–205.

AZEEM, Muhammad; SALFI, Naseer Ahmad. Usage of Nvivo software for qualitative data analysis. **Academic Research International**, v. 2, n. 1, p. 262–266, 2012.

BONIN, Jiani. Questões metodológicas na construção de pesquisas sobre apropriações midiáticas. In: MOURA, CLÁUDIA PEIXOTO DE; LOPES, MARIA IMMACOLATA VASSALLO DE (Org.). **Pesquisa em comunicação**: metodologias e práticas acadêmicas. Porto Alegre: EDIPUCRS, 2016. p. 213–231.

CHA, Meeyoung *et al.* Measuring user influence in twitter: The million follower fallacy. **ICWSM 2010 - Proceedings of the 4th International AAAI Conference on Weblogs and Social Media**, p. 10–17, 2010.

CHEN, Emily; LERMAN, Kristina; FERRARA, Emilio. COVID-19: The First Public Coronavirus Twitter Dataset. p. 4–5, 2020. Disponível em: <http://arxiv.org/abs/2003.07372>. Acesso em: 26 abr. 2022.

DIAS, Cristiane. **Sujeito, sociedade e tecnologia**: a discursividade da rede (de sentidos). São Paulo: Hucitec Editora, 2012.

FLICK, Uwe. **Introdução à pesquisa qualitativa**. 3. ed. Porto Alegre: Artmed, 2009a.

FLICK, Uwe. **Desenho da pesquisa qualitativa**. Porto Alegre: Artmed, 2009b.

FRAGOSO, Suely; RECUERO, Raquel; AMARAL, Adriana. **Métodos de pesquisa para internet**. Porto Alegre: Sulina, 2011.

FRANÇA, Vera. Celebidades: identificação, idealização ou consumo? In: V. FRANÇA; J. FREIRE FILHO; L. LANA; P. SIMÕES (org). **Celebidades no século XXI**: transformações no estatuto da fama. Porto Alegre, Sulina, 2014, p. 15-36.

GRIJÓ, Wesley Pereira. Uso de software de pesquisa qualitativa na análise de dados da recepção midiática. **Entre.Meios**, v. 10, n. 10, 2013.

GRUPO DE MÍDIA DE SÃO PAULO. **Mídia Dados Brasil 2018**. Disponível em:

<http://midiadados.org.br/2018/Midia%20Dados%202018%20%28Interativo%29.pdf>. Acesso em: 30 out. 2018.

JACKS, Nilda *et al.* Um cenário de intensa mudança: 2010 a 2015. In: JACKS, Nilda *et al.* (Org.). **Meios e audiências III**: reconfigurações dos estudos de recepção e consumo midiático no Brasil. 1. ed. Porto Alegre: Sulina, 2017. p. 11–18.

JACKS, Nilda *et al.* Uso de softwares na abordagem qualitativa: a experiência da pesquisa “Jovem e Consumo Midiático em Tempos de Convergência”. **Questões Transversais**, v. 4, n. 7, p. 46–54, 2016.

JAHANBIN, Kia; RAHMANIAN, Vahid. Using twitter and web news mining to predict COVID-19 outbreak. **Asian Pacific Journal of Tropical Medicine**, n. March, p. 26–28, 2020.

JENKINS, Henry. **Cultura da convergência**. 2. ed. São Paulo: Aleph, 2009.

KELLE, Udo. Análise com auxílio de computador: codificação e indexação. In: BAUER, MARTIN W.; GASKELL, GEORGE (Org.). **Pesquisa qualitativa com texto, imagem e som: um manual prático**. 7. ed. Petrópolis: Vozes, 2008. p. 393–415.

KNAFLIC, Cole Nussbaumer. **Storytelling com dados**: um guia sobre visualização de dados para profissionais de negócios. Rio de Janeiro: Alta Books, 2018.

KOUZY, Ramez *et al.* Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. **Cureus**, v. 12, n. 3, 2020.

LAGE, Maria Campos. Utilização do software NVivo em pesquisa qualitativa : uma experiência em EaD. **ETD – Educação Temática Digital**, v. 12, p. 198–226, 2011.

LAGE, Maria Campos; GODOY, Arilda Schmidt. O uso do computador na análise de dados qualitativos: questões emergentes. **RAM. Revista de Administração Mackenzie**, v. 9, n. 4, p. 75–98, 2008.

LÉVY, Pierre. **A inteligência coletiva**: por uma antropologia no ciberespaço. São Paulo: Edições Loyola, 1998.

LOPES, Maria Immacolata Vassallo de. **Pesquisa em comunicação**. 10. ed. São Paulo: Edições Loyola, 2010.

LOPEZ, Christian E.; VASU, Malolan; GALLEMORE, Caleb. Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset. p. 1–4, 2020. Disponível em: <http://arxiv.org/abs/2003.10359>. Acesso em: 26 abr. 2022.

MARTINO, Luís Mauro Sá. **Métodos de pesquisa em comunicação**: projetos, ideias, práticas. Petrópolis: Vozes, 2018.

MOZZATO, Anelise Rebelato; GRZYBOSKI, Denize; TEIXEIRA, Alex Niche. Análises qualitativas nos estudos organizacionais: as vantagens no uso do software NVivo®. **Revista Alcance**, v. 23, n. 4, p. 578–587, 2016.

OROZCO GÓMEZ, Guillermo; GONZÁLEZ REYES, Rodrigo. **Una cortada metodológica**: abordajes cualitativos en la investigación en comunicación, medios y audiências. México: Productora de Contenidos Culturales, 2011.

PIENIZ, Mônica; SILVA, Ronei Teodoro; MATOS, Ludimila Santos. Sujeitos em trânsito na

internet. In: JACKS, Nilda *et al.* (Org.). **Meios e audiências III**: reconfigurações dos estudos de recepção e consumo midiático no Brasil. 1. ed. Porto Alegre: Sulina, 2017. p. 19–40.

PRIMO, Alex. Afetividade e relacionamentos em tempos de isolamento social: intensificação do uso de mídias sociais para interação durante a pandemia de COVID-19. **Comunicação & Inovação**, v. 21, p. 176-198, 2020.

RECUERO, Raquel. Estudando discursos em mídia social: uma proposta metodológica. In: SILVA, Tarcízio; BUCKSTEGGE, Jaqueline; ROGEDO, Pedro (Org.). **Estudando cultura e comunicação com mídias sociais**. Brasília: IBPAD, 2018. p. 13–30.

RECUERO, Raquel; SOARES, Felipe. O discurso desinformativo sobre a cura do COVID-19 no Twitter. **E-Compós**, v. 24, n.1, p. 1-29, 2021.

SANTAELLA, Lucia; LEMOS, Renata. **Redes sociais digitais**: a cognição conectiva do Twitter. 1. ed. São Paulo: Paulus, 2010.

TEIXEIRA, Alex Niche. A centralidade do pesquisador na relação com os softwares de análise qualitativa. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE SOCIOLOGIA, 15. 2011. **Anais [...]**. Curitiba,, 2011.

TEIXEIRA, Alex Niche; BECKER, Fernando. Novas possibilidades da pesquisa qualitativa via sistemas CAQDAS. **Sociologias**, v. 3, n. 5, p. 94–113, 2001.

ⁱDoutora em Comunicação. Docente do Departamento de Ciências da Comunicação e do Programa de Pós-Graduação em Comunicação da Universidade Federal de Santa Maria.