

ARQUIVOS PESSOAIS E REDES SOCIAIS: O TWITTER CONSTRUÍDO COMO DOCUMENTO HISTÓRICO

Personal archives and social networks: Twitter built as historical document

Archivos personales y redes sociales: Twitter construido como documento histórico

ALESSON RAMON ROTA^{I*}
THIAGO LIMA NICODEMO^{II**}

<http://doi.org/10.1590/S2178-149420230204>

^I Programa de Pós-Graduação em História da Universidade Estadual de Campinas – Campinas, SP, Brasil.

^{*} Doutorando em História pela Universidade Estadual de Campinas (Unicamp), com financiamento da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP); mestre em História pela Unicamp e bacharel em História pela Universidade Federal do Rio Grande (FURG), com período sanduiche na Universidade de Coimbra. Pesquisador visitante na Universidade Livre de Berlim; colaborador nos laboratórios de pesquisa CHD-Unicamp e Gumelab-FU. Vencedor, em 2016, do prêmio de monografia da Sociedade Brasileira de Teoria e História da Historiografia (SBTHH). Membro dos grupos de pesquisa Núcleo História e Linguagens Políticas: Razão, Sentimentos e Sensibilidades e Historiografias Periféricas em Perspectiva Global. Desenvolve conteúdo de divulgação científica em humanistasdigitais.com (alesson.rota@gmail.com)

 <https://orcid.org/0000-0001-9167-7903>

^{II} Universidade Estadual de Campinas – Campinas, SP, Brasil.

^{**} Coordenador do Arquivo Público do Estado de São Paulo; presidente da Comissão Estadual de Acesso à Informação (CEAI); professor de Teoria da História da Universidade Estadual de Campinas (Unicamp) e responsável pelo Centro de Humanidades Digitais do Instituto de Filosofia e Ciências Humanas (Unicamp). Doutor e mestre em História Social pela Universidade de São Paulo (USP); bacharel em História pela USP e em Direito pela Pontifícia Universidade de São Paulo (PUCSP). Autor de *Urdidura do vivido* (Edusp, 2008), *Alegoria moderna* (Editora Unifesp, 2014) e de *Uma introdução à historiografia brasileira, 1870-1970* (com Pedro A. C. dos Santos e Mateus Pereira, Editora FGV, 2018). Dedicou-se à gestão de dados governamentais, a políticas de acesso aberto à informação, à ciência e preservação digital, além de manter as linhas de pesquisa em arquivos e sobre a historiografia em escala global (tnicodem@unicamp.br)

 <https://orcid.org/0000-0002-1588-0683>

Recebido: 15/01/2023 – Aprovado: 22/03/2023

RESUMO

Este texto visa discutir as implicações do Twitter, construído para ser um documento histórico. Para isso, balizaremos os argumentos nas discussões sobre arquivos públicos e pessoais, bem como as novas demandas do universo digital, no qual entendemos que esses dados são de natureza pessoal, mas se tornam públicos no Twitter. Compreende-se que os dados cedidos à plataforma precisam ser arquivados para se configurarem como documentos históricos. Por isso, buscou-se demonstrar como é formado um arquivo do Twitter para pesquisa em história e formas de análise, construindo a evidência a partir do episódio de ataques às sedes dos Três Poderes, em 8 de janeiro de 2023.

PALAVRAS-CHAVE: arquivos pessoais, redes sociais, arquivamento, Twitter.

ABSTRACT

This text aims to discuss the implications of Twitter, constructed to be a historical document. To that end, we will base the arguments on the discussions about public and personal archives, as well as the new demands of the digital sphere, in which we understand that these data are of a personal in nature, but become public on Twitter. It is understood that the data given to the platform need to be archived to be historical documents. Thus, we sought to demonstrate how a Twitter archive is constructed for research in history and ways of analysis, by building the evidence from the episode of attacks on the headquarters of the Three Powers, on January 8, 2023.

KEYWORDS: personal archives, social networks, archiving, Twitter.

RESUMEN

Este texto tiene como objetivo discutir las implicaciones de Twitter, pensado para ser un documento histórico. Para ello, nos basaremos en las discusiones sobre archivos públicos y personales, así como en las nuevas demandas digitales, en las que entendemos que estos datos son de carácter personal, pero se hacen públicos en Twitter. Se entiende que los datos transferidos a la plataforma deben archivar como documentos históricos. Por lo tanto, buscamos demostrar cómo se constituye un archivo en Twitter para la investigación en historia y formas de análisis, construyendo evidencia a partir del episodio de los ataques a la sede de los Tres Poderes en Brasil, el 8 de enero de 2023.

PALABRAS CLAVE: archivos personales, redes sociales, archivo, Twitter.

INTRODUÇÃO

Este texto visa discutir as implicações do Twitter, construído para ser usado como um documento histórico. Segundo levantamentos do Global Overview Report, (Kemp, 2022) o Brasil é um dos principais países que mais consome conteúdo web no mundo, com mais aparelhos telefônicos que pessoas (cerca de 171,5 milhões de usuários ativos de redes sociais). No planeta, somam-se mais de quatro bilhões de usuários de internet, motivo pelo qual o crescimento sobre o uso do digital tem sido objeto de vários tipos de análises. Nas humanidades, por exemplo, podemos citar a natureza das novas mídias (Manovich, 2002), as mediações operadas pelos computadores (Telles, 2018) e os métodos de pesquisa (Nicodemo; Rota; Kisil, 2022). Visando contribuir com o debate do Twitter enquanto documento histórico, investigamos a natureza desses dados, suas potencialidades e suas restrições. Optamos por alicerçar os raciocínios no campo da história digital, por incorporar um conjunto de reflexões interdisciplinar aos tradicionais saberes da escrita da história, além de destacar o papel da prática historiadora para o entendimento da cultura eletrônica (Lucchesi; Silveira; Nicodemo, 2020)

O Twitter é uma plataforma privada que possibilita à sua comunidade de usuários enviarem e receberem atualizações. Para isso, aqueles que aderem à rede consentem em fornecer suas informações pessoais para a rede social (Twitter, 2023a). Essas informações podem ser divididas em um duplo aspecto: dados pessoais, como nome, localização, descrição do perfil, e as práticas na plataforma, relacionadas ao que publicam, com quem interagem ou como se posicionam. Acrescenta-se, ainda, dados de navegação, que não são publicizados, mas fazem parte da gestão do Twitter para funcionamento de algoritmos. Os dados pessoais são cedidos à empresa como forma de pagamento pelo uso da aplicação, os quais são capitalizados para gerar receitas, via propagandas. Ao tuitar, o usuário publica informações sobre si, sejam dados pessoais ou a narrativa que constitui a postagem. Esse conjunto de dados pode, por meio de métodos digitais, ser tomado como documento histórico.

Análises por meio do Twitter têm sido amplamente empregadas para o entendimento de dinâmicas sociais como protestos (Tufekci, 2017), desastres ambientais (Haddow; Haddow, 2013) e propagação de notícias falsas (Calvo; Aruguete, 2020). A rede é tratada como um espaço de debate público, devido à mediação que opera entre atores políticos, artísticos, empresariais, financeiros, imprensa e outros setores mais amplos da sociedade. Contudo, observando o funcionamento da plataforma e da legislação brasileira, recuperaremos a dimensão pessoal da rede social. É possível que o caráter privativo tenha menos ênfase no debate dado o diagnóstico das sociedades de massa efetuado por Hannah Arendt (2007: 68), que “destrói tanto

a esfera pública como a esfera privada”. Essa mistura entre público e privado encontra um lugar de destaque no Twitter, que se distingue de outras redes sociais por ser frequentemente utilizado por personalidades públicas para reagir aos acontecimentos ao seu redor.

Ao interrogarmos sobre a natureza do Twitter e seu processo de arquivamento, nos deparamos com os limites e transformações do arquivo contemporâneo. O que consideramos arquivos pessoais no século XXI, como anotações, fotos e mensagens, são conteúdos armazenados por grandes corporações especializadas em *big data*. Além do acesso, o digital mudou o modo com que as pessoas produzem suas subjetividades, implicando nos conteúdos pessoais publicados nas redes sociais. Na primeira parte do texto discutiremos o que é um tuíte arquivado, tencionando as distinções sobre arquivos públicos e pessoais, bem como a legislação que o regulamenta no Brasil. Em seguida, abordaremos os aspectos digitais e metodológicos. Por fim, demonstraremos alguns tipos de análises possíveis com o Twitter arquivado¹.

O TWITTER E OS ARQUIVOS

Nas últimas décadas, a estabilidade do conceito de arquivo tem sido questionada devido às viradas historiográficas vinculadas aos estudos linguísticos, à geração pós-1970, aos estudos pós-coloniais, entre outras. As antigas noções de arquivo público e pessoal entraram em pauta quando os critérios disciplinares foram problematizados. Conforme escrevem Letícia Nedel e Luciana Heymann (2018: 7),

Tais reflexões colocaram em xeque a própria estabilidade dos documentos, ao chamarem atenção para os sentidos diferenciais que lhes poderiam ser atribuídos de acordo com o tratamento técnico que recebessem, o lugar em que estivessem depositados e as ações a que fossem submetidos.

Essa desterritorialização do arquivo possibilitou novas avaliações heurísticas sobre a instituição e novas formas de documentação. Cabe lembrar, por exemplo, que arquivos pessoais podem conter informações singulares que não estão disponíveis em outras fontes, como documentos internos de empresas ou governos. Eles também podem fornecer uma visão única e não disponível em outros lugares sobre eventos históricos ou pessoas. Dependendo do conteúdo do arquivo e do contexto em que foi criado, pode fornecer perspectivas sobre história política (Abreu, 1996; Gomes; Hansen, 2016), história intelectual (Gomes, 2005), trocas globais (Iumatti; Nicodemo, 2018), relações sociais e até mesmo acerca das estruturas de poder, entre outras abordagens. Arquivos privados são, em vários casos, compostos durante a atuação de determinados personagens publicamente, sejam políticos, funcionários de Estado

ou governo, artistas, jornalistas ou empresários. Para Catherine Hobbs (2018: 260), “são os próprios filtros pessoais que revelam o caráter íntimo do criador dos arquivos”.

No cenário nacional, além dos debates sobre política de preservação, fez-se presente os critérios de acessibilidade. Passou-se a debater, nesse sentido, as práticas de acesso e a utilização desses documentos acumulados por indivíduos que podem ter valor histórico como fontes de pesquisa, sobretudo se as pessoas ocuparam posições de relevância ou tiveram contato com personagens significativos. Arquivos pessoais podem conter informações importantes, mas como eles são de propriedade privada, há uma problemática em relação ao acesso que deve ser esclarecida. Pensando no confronto entre a vida privada e os limites do acesso à informação, Costa (1998: 194) escreveu que “não se deve perder de vista o interesse público, que, especificamente no que diz respeito aos arquivos, se traduz na demanda de informações e na necessidade de difundi-las em função do exercício pleno da democracia e da pesquisa científica”.

A Lei n. 8.159 de 1991, a Lei de Arquivos (Brasil, 1991), é a principal legislação que regulamenta os arquivos públicos e privados no Brasil. Até o advento da Lei n. 12.527/2011, a Lei de Acesso à Informação (LAI) (Brasil, 2011) era também a legislação competente para tratar do acesso e sigilo a documentos públicos. A Lei de Arquivos aprofunda mecanismos constitucionais relacionados com o princípio da responsabilidade e, no limite, do Estado de direito. O arquivo emerge como a instância que permite gerir e preservar a informação, desde a sua produção, de modo que um ente público tem as condições de responder e prestar contas à sociedade por seus atos e decisões. Essa Lei também regulamenta a declaração de interesse público em coleções privadas, e o Decreto n. 4073/2002 estipula que o Conselho Nacional de Arquivos, o Conarq, deve ser a instância competente, ao menos em âmbito Federal, para propor declaração de interesse público sobre coleção privada. No entanto, deve-se notar a precariedade desse mecanismo se compararmos com o processo de patrimonialização de bens privados. Faltam ritos e instâncias que garantam a eficácia dessa declaração, não só no âmbito do Governo Federal, mas também nos estados e municípios.

A LAI representou um aprofundamento na Lei de Arquivos, considerando um momento, a década de 2010, de maior velocidade e fluxo na produção da informação e consequente aumento da demanda social por transparência. Seu contexto é emergência é o da abertura dos arquivos da ditadura e de revisão do sigilo e, não, por acaso, ela foi sincrônica à Comissão Nacional da Verdade. A LAI foi aperfeiçoada em alguns sentidos pela Lei Geral de Proteção de Dados (n. 13.709/2018) (Brasil, 2018), chamada LGPD, uma lei federal que estabelece regras para a coleta, o armazenamento, o (re)uso e a proteção de dados pessoais.

A LGPD se aplica a todas as empresas e organizações (incluindo o Poder Público) que coletam, armazenam ou usam dados pessoais de indivíduos no Brasil, independentemente de serem públicas ou privadas.

Somado ao Marco Civil da Internet (Lei n. 12965/2014), esse conjunto de leis regulamenta também as redes sociais e como elas podem ser utilizadas para pesquisa². Essas plataformas precisam seguir as regras estabelecidas pela LGPD, como obter consentimento explícito dos usuários antes de coletar, armazenar ou compartilhar seus dados, fornecer transparência sobre como os dados são usados e garantir a privacidade dos usuários. Para pesquisas baseadas nas redes, se houver necessidade de identificação das pessoas envolvidas, deve-se submeter ao conselho de ética vinculado à instituição responsável. A exceção ocorre em dados utilizados para fins estatísticos, demográficos e/ou matemáticos que não envolvam a identificação dos seres humanos envolvidos. Conforme escreve Lenora Schwaitzer (2020: 41),

a Lei não se aplica a tratamento realizado por pessoa natural se a finalidade não tiver caráter econômico ou se o tratamento possuir motivação artística ou jornalística, se possuir finalidade acadêmica — e neste caso deve assegurar os cuidados previstos para tratamento dos dados pessoais e dados sensíveis — ou se visar a segurança pública, a defesa nacional, a segurança do Estado ou tiver intuito de realizar investigação ou repressão de infração penal.

Portanto, é necessário que os dados pessoais sejam anonimizados para impossibilitar a identificação das pessoas as quais se referem.

Nesse movimento, os conteúdos extraídos do Twitter deixam de ser um acervo pessoal de mídias, textos e dados publicizados na plataforma para se transformarem em arquivos pessoais de autoria restringida, com exceção para figuras públicas (ou instituições governamentais) ou se o assunto é de interesse público. Conta nesse caso também se a conta Twitter ou o post é aberto, pois considera-se a abertura como o consentimento com a publicidade ampla daquele conteúdo. A autoria restringida deve ocorrer no momento de publicação das informações, seja enquanto metadados de uma pesquisa, seja como narrativa histórica. A questão da anonimização merece atenção especial por funcionar como uma forma de proteção ao pesquisador, principalmente caso haja dúvida se determinadas informações pessoais são realmente de interesse público.

É pertinente ressaltar que as redes sociais são também estudadas como formas de arquivamento e de conteúdo pessoal (Ellison, 2007). Ao comparar o Facebook com os antigos livros de recorte (*scrapbooks*)³, Katie Day Good (2013: 559) ressalta o princípio de

funcionamento compartilhado das duas mídias, calcado na “ contenção de — e dependência formal — de diversos fluxos de conteúdo pessoal”. No entanto, entre essas duas mídias há diferenças sobre as formas de privacidade, considerando os tipos de contas e postagens que se pode ter nas redes sociais — pública ou privada — e a distribuição de conteúdo segundo uma lógica algorítmica. Mas claro, a comparação é sadia para observar a “sobreposição duradoura entre e processos de autoarquivamento na cultura midiática” (Good, 2013: 569). De qualquer forma, as redes sociais deixam rastros de eventos e interações sociais que moldam a vida dos usuários.

A relação entre os arquivos privados e seu compartilhamento é anterior, inclusive, à internet. Ela pode ser identificada em práticas intelectuais modernas, como a compilação de citações, poemas e provérbios, que poderia ter um fim de memorização ou compartilhamento (Garvey, 2012; Katritzky, 2006). De alguma forma, tanto os livros de recortes como as redes sociais têm como fundamento o compartilhamento de experiências pessoais. A primeira delas, o SixDegrees, foi criada em 1997, permitindo que seus usuários criassem perfis e interagissem com seus amigos, estabelecendo as bases para a estrutura de funcionamento das redes sociais que viriam a seguir. Com o tempo, essas plataformas evoluíram e se diversificaram. O Friendster, lançado em 2002, e o MySpace, em 2003, introduziram uma maior personalização dos perfis, fazendo com que o compartilhamento de experiências pessoais se tornasse ainda mais significativa. Essa tendência foi reforçada pelo Facebook, em 2004, que se destacou por sua capacidade de conectar pessoas em todo o mundo. Nesse contexto, as redes sociais se tornaram de alguma forma um reflexo dos livros de recados, mas com um alcance muito maior, proporcionando um espaço onde cada usuário pudesse compartilhar suas narrativas com outras pessoas. Os momentos registrados em fotos, vídeos e textos tornaram-se recortes digitais da vida que cada um conta sobre si, assim como os fragmentos colados nas páginas dos álbuns analógicos.

É possível inferir que a maior parte do conteúdo produzido no Twitter — e em outras redes sociais — está relacionado à personalidade dos usuários, com publicações de textos, fotos ou vídeos de suas intimidades. No entanto, poucos casos ganham notoriedade e se transformam em pauta do debate público. As informações pessoais permanecem assim até que a curadoria obscura dos algoritmos as torne visíveis para muitas pessoas⁴. Na prática, o que distingue o usuário comum daquele que viraliza é o número de visualizações. Essa dinâmica é ainda mais evidente em casos de pessoas que se tornam pauta do debate público e tiveram suas redes sociais vasculhadas por outros usuários, com o objetivo de encontrar possíveis contradições ou informações que exponham a pessoa em questão. A viralização de conteúdo leva

a um escrutínio intenso da vida pessoal e das opiniões do indivíduo, gerando consequências variadas, que podem afetar reputações.

Em um estudo de caso realizado pela Microsoft a partir de entrevistas sobre os arquivos pessoais na web, as redes sociais foram consideradas menos importantes, devido ao aspecto cotidiano e efêmero, do que blogs ou armazenamentos em nuvem, caso os usuários tivessem que selecionar o que deveria ser arquivado. Ainda assim, a prática abre exceção para um possível arquivamento de redes sociais, enquanto salvaguarda, para eventos que consideram marcantes em suas trajetórias (Lindley *et al.*, 2013: 7). Chamada de autodocumentação, essa prática também tem sido pensada como um elemento da história digital. Conforme adverte Laura K. Morreale (2022), é pertinente que os(as) historiadores(as) construam padrões de arquivamento digital durante o processo de pesquisa como forma de produzir vestígios sobre a própria prática historiadora.

Pode-se perceber, portanto, que, no ambiente digital, as fronteiras entre as noções de arquivos pessoais e públicos são flexíveis. Mas, longe de ser uma exclusividade do ambiente computacional, como pensado nos debates sobre classificação de documentos, Rob Fisher (2018: 333) entende que há “zonas cinzentas entre os arquivos públicos e privados, e os arquivos pessoais e corporativos”. Seja como for, não se trata de categorias superadas, “são conceitos úteis e indispensáveis, se estiverem cientes de suas problemáticas e fronteiras” (Fisher, 2018: 333). Isso posto, os documentos digitais contêm a mesma característica dos analógicos: dependem de caso a caso a classificação, considerando o contexto de produção, quem produziu e seu conteúdo.

A SINGULARIDADE DO DIGITAL

A reflexão no campo historiográfico sobre métodos computacionais ganhou mais espaço a partir do desenvolvimento da história digital e nos debates sobre novas mídias, nos anos 1990 (Ayers, 2001). É claro que existem projetos antigos e pioneiros da metade do século XX, como Roberto Busa ou o uso da computação para calcular dados quantitativos caros à segunda e terceira geração do movimento dos Annales, além da história econômica em sua totalidade. Mesmo assim, até então, o computador foi incorporado à pesquisa sem uma reflexão mais minuciosa das modificações que ele opera nessa prática. Foi com os debates sobre mídias que se desbanalizou a mediação computacional, diante da qual Roy Rosenzweig (2011) chamou a atenção para uma tênue fronteira entre a abundância e a escassez de documentos quando falamos sobre o digital. Abundância, segundo ele, devido às inúmeras informações, publicações, possibilidades de busca no ambiente web,

mas que, em simultâneo, é efêmera, podendo ser destruída com apenas alguns cliques. Sendo assim, tanto documentos analógicos digitalizados quanto documentos nativos do digital exigirão um novo sistema de organização para o futuro. Além disso, a produção de história no ambiente digital colocou os(as) historiadores(as) diante de novos públicos, flexionando novas audiências que merecem um debate mais apurado.

Para Lev Manovich, até mesmo o processo de transformação de documentos analógicos em digitais cria outro tipo de mídia. As metamídias, como são conhecidas, apresentam estruturas internas mais flexíveis do que os documentos analógicos, possibilitando outras formas de arquivamento e pesquisa (Manovich, 2006). Evidentemente, a transformação do analógico em digital ou o trabalho exclusivamente com o nato-digital resulta em processos de mediação que modificam tanto a natureza da informação quanto a própria experiência do pesquisador (Silveira, 2018). Conforme escreve Anita Lucchesi (2014: 52), a história digital adiciona “ferramentas à oficina da história, mas os fundamentos da disciplina continuam os mesmos”. Essas ferramentas são softwares que operam mediações algorítmicas entre os dados e as formas de análises.

Os dados de redes sociais construídos como documentos históricos são de acesso consideravelmente diferentes dos acervos históricos analógicos. Conforme escreve Richard Rogers (2013: 287), plataformas de redes sociais “conotam um ‘jardim murado’, onde o acesso, publicação e direitos de conteúdo são limitados”. Uma das poucas iniciativas para armazenar todo o conteúdo publicado no Twitter, para fins de pesquisa, coube à Biblioteca do Congresso dos Estados Unidos (Library of Congress), no projeto *Twitter Archive*, descontinuado em 2017 devido à alta quantidade de dados que seriam armazenados. Este, contudo, não é um problema apenas logístico, pois o processo de arquivar depende do contexto e do critério de seleção, pressupondo, portanto, exclusão. Sendo assim, copiar a base de dados inteira do Twitter é equivalente a reproduzir a própria plataforma, logo, tamanha seria a tarefa. Atualmente, as pesquisas em redes sociais dependem da composição de acervos próprios por parte dos pesquisadores, o que tem contribuído para a virada laboratorial, fazendo com que a sala de aula, as bibliotecas, os museus, as salas de leitura e os arquivos agora funcionem como espaço de pesquisa (Pawlicka-Deger, 2020). O laboratório, quanto à sua significação, nesse sentido, refere-se mais a um conjunto de práticas do que ao espaço físico propriamente dito.

De qualquer forma, conforme escreve Jane Winters, a prática historiadora se especializou nas últimas décadas em debates sobre escassez de documentos, uso de fontes variadas e interpretação de fragmentos. A leitura de Winters reconhece as dificuldades em torno da volatilidade das páginas web enquanto documentos, mas reconhece que a fragmentação

dos dados também faz parte das fontes mais tradicionais. Além disso, como destaca Winters, da mesma forma que os bancos de dados das grandes empresas não estão disponíveis para os historiadores, parte considerável dos documentos históricos não estiveram ou ainda não estão, ou seja, o historiador já está relativamente treinado para este tipo de adversidade (Winters, 2017), motivo pelo qual a dificuldade de acesso não é um problema novo.

Nos últimos anos cresceram as iniciativas para arquivamento de dados da web, fenômeno que também se aplica ao Brasil (Rockembach, 2018). A demanda dos pesquisadores por informações referentes às redes sociais levou à criação de uma série de ferramentas não autorizadas pelas plataformas oficiais. Estas são desenvolvidas e aprimoradas pela comunidade de usuários, em plataformas especializadas em edição coletiva, o chamado Git. Criado por Linus Torvald, o mesmo compilador do sistema Linux, o Github é um sistema de controle de versões, usado principalmente no desenvolvimento de software, mas também para registrar o histórico de edições de qualquer tipo de arquivo. Chama a atenção o número de softwares desenvolvidos nos últimos vinte anos para coletar dados apenas do Twitter, passando de duzentos mil (Github, 2023). Muitos desses softwares têm vida curta, à medida que as plataformas oficiais criam sistemas para bloqueá-los, dependendo da comunidade de usuários e de novos subterfúgios para continuarem existindo.

A fim de apresentar novos elementos para esta disputa, o Twitter desenvolveu uma Interface de Programação de Aplicativos (Application Programming Interface — API) voltada para pesquisadores (Tornes, 2021). Ferramenta para coleta de dados da rede social, por meio dessa API pode-se ter acesso a informações como nome de usuário, nome verdadeiro, localização, texto postado, endereço da publicação, número de retuite da publicação, curtidas, número de respostas, número de seguidores da conta, data de criação da conta, data de criação do tuíte, se o perfil é verificado pela plataforma e o número total de seguidores e tweets. Cabe salientar que as ferramentas não oficiais de coleta de dados no Twitter, como Twint e Snsrape, trazem mais informações sobre o usuário e suas práticas, como descrição do perfil e as mídias que compartilha.

Dessa forma, pode-se dividir os dados em 1) informações sobre o usuário e 2) informações sobre o que o usuário publica na rede. Inicialmente, esses dados só poderiam ser coletados semanalmente, sem acesso ao arquivo geral de publicações, mas a partir de 2021 foi permitida a pesquisa retroativa. Contudo, em 2023 houve nova alteração, sendo permitido coletar apenas 1500 tuítes por mês gratuitamente. Desse modo, a disputa entre pesquisadores, comunidade de usuários de código aberto e as redes sociais está longe de ter um fim. No fundo, a questão colocada é sobre a soberania dos dados, sobre a capacidade

das instituições de um país ou governo controlar e garantir o uso, o acesso e a proteção de bases geradas e armazenadas. Deve-se considerar que os dados são evidência da vida, fragmentos de experiências essenciais no mundo contemporâneo. Isso contrasta com o fato de que, cada vez menos, temos o governo sobre os nossos próprios dados, sobre a nossa própria memória (Nicodemo, 2022). Seja porque os dados estão custodiados por grandes corporações em aplicativos, como as redes sociais, ou em serviços de armazenamento em nuvem. Nesse sentido, o Estado deve agir como agente mediador para que os dados sejam utilizados em sua potencialidade a fim de prover instrumentos de cidadania e democracia. Isso se apresenta como um desafio, uma grande contradição basilar entre o público e o privado. Dados pessoais podem ganhar a força e o interesse público, mas estão sob a custódia de corporações compromissadas apenas com seus próprios interesses. Isso é um desafio para a própria noção de arquivo pessoal: o que guardaremos no futuro se não temos garantia de que nada será preservado? Como será composto um arquivo pessoal do século XXI quando se pressentir o interesse em preservá-lo? Os arquivos pessoais de interesse público que conhecemos são compostos de cartas, fotos, documentos oficiais fora do lugar, notas de pesquisa etc. O arquivo pessoal do futuro, que já é presente, pode ser muito diferente, mais parecido com um conjunto de raspagens de dados, transversal em sua temática, desafiando o pessoal, desafiando o preservar. Instantâneo, instável, dessubjetivado. (Iumatti; Nicodemo, 2018: 113)

Em sociedades em que o computador é um mediador cotidiano da maioria das profissões, tornou-se prática comum criar sistemas de registro, armazenamento, softwares para dinamizar a produtividade, economizar tempo e energia. Esse fenômeno tem sido chamado de datificação, que pode ser explicado como base na transformação dos diversos aspectos da vida humana em dados, para serem tratados e transformados em informação. Isso abriu discussões sobre os limites éticos do uso de dados para o controle das populações, seja pela iniciativa pública ou privada (Magrani, 2019). As preocupações são válidas, porque “as tecnologias e seus modos de tratamento e armazenamento, bem como os fluxos de dados não beneficiam todas as populações nem enriquecem do mesmo modo todas as comunidades e localidades” (Cassino *et al*, 2021: 8). Considerando essas assimetrias digitais que podem existir nas relações público/privada, entre instituições ou Estados nacionais, a comunidade científica está se organizando por meio das práticas da ciência aberta. Trata-se de uma política que visa garantir que os conteúdos armazenados estejam a serviço da comunidade científica, mantendo sua independência teórica e confiabilidade (Cribb *et al*, 2010).

Quando pensamos em soberania sobre os dados, a primeira iniciativa a ser considerada é a criação e gestão da base pelos pesquisadores. Deve-se delimitar claramente o escopo da

base de dados, incluindo quais dados serão coletados, armazenados e utilizados; determinar quais tipos de informação serão incluídas na base de dados, bem como as definições e os formatos de dados a serem utilizados; estabelecer regras de integridade para garantir que os dados sejam precisos e consistentes, criar medidas de segurança para proteger a base de dados contra acessos não autorizados; e implementar planos de backup e recuperação para garantir que os dados possam ser recuperados em caso de falha do sistema. Além disso, é importante considerar se a base de dados será escalável para atender às necessidades crescentes de armazenamento e processamento de dados e como essa base será acessada e como as informações serão compartilhadas com outros sistemas e aplicativos. Seguindo esse plano de gestão, os pesquisadores podem compartilhar dados, métodos e resultados sem restrições indevidas. Portanto, quando armazenamos dados sobre o Twitter em um banco para pesquisa, salvaguardamos informações que podem ser sensíveis ao uso da rede social, como o apagamento, garantindo a publicidade, replicabilidade e verificabilidade dos estudos.

O TWITTER ARQUIVADO

A maioria dos serviços e aplicativos que têm como finalidade a pesquisa no Twitter são vendidos sob licença de uso e não necessariamente dialogam com uma política de ciência aberta. Alguns exemplos são as plataformas Google Analytics, Brandwatch, Sprout Social, Crimson Hexagon e Synthesio, os quais são serviços que permitem o rastreamento de tráfego proveniente de redes sociais, bem como obtêm informações demográficas sobre os visitantes, permitindo, também, acompanhar menções de palavras-chave na web, análise de dados das redes sociais para compreender o comportamento do consumidor, tendências e análise de sentimento sobre os assuntos comentados nas redes. Esses exemplos de ferramentas são de uso facilitado, porque integram a base de dados e a análise no mesmo programa, mas criam o repositório internamente nas respectivas plataformas, impossibilitando o reuso da base para outras finalidades ou em outros métodos de análise. Por isso é necessário separar os métodos de coleta de dados dos métodos de análise.

Seguindo o modelo de prática de pesquisa desenvolvido no tópico anterior, para qualquer laboratório ou grupo de pesquisa, é necessário a criação do próprio banco de dados com um plano de gestão de dados. Como dito anteriormente, o Twitter disponibilizou, a partir de 2021, uma ferramenta para a coleta de dados para a rede social, mas, no início do 2023, sua licença foi alterada (Twitter, 2023b), restringindo as formas de uso⁵. Como alternativa, temos na Figura 1 uma exemplificação de base de dados escalonável, obtida por meio da biblioteca Snsrape (2023)⁶. Esta biblioteca é um pacote Python desenvolvido para coleta de dados

em várias redes sociais, incluindo o Twitter, Facebook, Instagram e Reddi. Ela permite que os usuários colem informações como postagens, comentários, curtidas e outros dados relacionados, sem a necessidade de uma chave de API oficial. As colunas — também chamadas de entidades — representam as categorias de informações que podem ser obtidas na rede social, que são independentes do conteúdo. Se forem realizadas pesquisas com temas e recorte temporais diferentes, esses conteúdos podem ser relacionados por meio das entidades. Em outras palavras, é possível escalonar os dados, complementando com novas coletas e integrando-as à mesma base. Apesar das entidades serem previamente determinadas pela rede social, se exploradas nos limites, podem trazer contribuições para a pesquisa em história, via informações pessoais, tipos de publicação, localização, repercussão e a interação com outros usuários.

1: Data/hora	2: Tweet	4: Tipo	8: Localização do usuário	12: Seguidores	13: Segue	14: Tweets	15: Perfil verificado	16: Perfil criado	17: Retweets	18: Likes	19: Língua	20: Fonte
09/01/2023 16:04	O Sistema de Inteligê	Retweet	Ribeirão Preto, Brasil	257	412	8512	FALSO	10/11/2019 21:42	142	0	Portuguese	Twitter for Android
09/01/2023 16:04	O Sistema de Inteligê	Retweet		19	128	1121	FALSO	24/10/2022 12:00	142	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	A Rede Globo faz da s	Retweet		539	886	1265	FALSO	04/11/2022 13:05	1176	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	Denúncia: já trisham	Retweet	Vila Itoupava, Blumenau	71	375	1402	FALSO	21/03/2021 18:40	168	0	Portuguese	Twitter for Android
09/01/2023 16:04	Não Foram os Patrioti	Retweet	Vila velha ES	685	1139	40808	FALSO	07/05/2022 18:56	2853	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	Mais um da série: infi	Retweet		5	90	391	FALSO	22/09/2010 20:32	7951	0	Portuguese	Twitter for Android
09/01/2023 16:04	Era para ter sido uma	Retweet	Salvador, Brasil	3177	3282	14448	FALSO	04/06/2021 00:16	1912	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	VÁRIOS MANIFESTAN	Retweet		1543	2169	6537	FALSO	01/11/2022 03:41	2512	0	Portuguese	Twitter for Android
09/01/2023 16:04	Black Blocs infiltrados	Retweet		436	743	6534	FALSO	29/04/2022 22:09	2734	0	Portuguese	Twitter for Android
09/01/2023 16:04	A hipótese de agente	Retweet	Brasília, Brasil	19	127	472	FALSO	27/05/2009 14:14	3254	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	O Sistema de Inteligê	Retweet	São Paulo, Brasil	373	387	8000	FALSO	24/11/2022 14:04	142	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	INFILTRADOS. identi	Retweet		60	322	2213	FALSO	10/11/2022 16:20	32	0	Portuguese	Twitter for Android
09/01/2023 16:04	infiltrados e protegi	Retweet		136	202	5896	FALSO	15/11/2021 01:28	12	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	O Ministério da Justiç	Retweet	Curitiba, Brasil	49	88	2307	FALSO	08/09/2021 23:32	880	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	Muitos infiltrados, a	Retweet	BR-PE-Recife/Petrolina maritima	382	1052	10267	FALSO	05/01/2010 04:48	634	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	O sistema de inteligê	Retweet		346	641	15222	FALSO	29/04/2014 00:39	142	0	Portuguese	Twitter for Android
09/01/2023 16:04	Essa aqui gravou um v	Retweet		3	67	9280	FALSO	10/08/2012 01:12	582	0	Portuguese	Twitter Web App
09/01/2023 16:04	VAMOS AJUDAR A DIV	Retweet		3623	3468	46392	FALSO	21/02/2021 05:00	555	0	Portuguese	Twitter for Android
09/01/2023 16:04	O Sistema de Inteligê	Retweet	Brasil	407	1159	21828	FALSO	14/05/2009 11:36	142	0	Portuguese	Twitter for Android
09/01/2023 16:04	O Ministério da Justiç	Retweet	Alto Araguaia, Brasil	521	654	1768	FALSO	30/11/2022 02:31	880	0	Portuguese	Twitter for Android
09/01/2023 16:04	Mais um da série: infi	Retweet	Estados Unidos	253	347	5200	FALSO	29/07/2010 14:17	7951	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	🇺🇸🇺🇸🇺🇸🇺🇸	Retweet		1448	2888	2711	FALSO	01/10/2022 18:39	962	0	Portuguese	Twitter Web App
09/01/2023 16:04	Resumo: a quebradei	Retweet		3601	1667	569936	FALSO	18/04/2010 14:43	592	0	Portuguese	Twitter Web App
09/01/2023 16:04	🇺🇸 ai os infiltrados	Retweet	Piedade, Brasil	512	94	5189	FALSO	08/05/2011 21:08	7	0	Portuguese	Twitter for iPhone
09/01/2023 16:04	Mais um da série: infi	Retweet	BRASIL-EM ALGUM LUGAR =h	23	291	1364	FALSO	13/06/2020 19:24	7951	0	Portuguese	Twitter Web App
09/01/2023 16:04	A hipótese de agente	Retweet		185	268	5535	FALSO	12/09/2022 01:18	3254	0	Portuguese	Twitter for Android
09/01/2023 16:04	Nos grupos bolsionari	Retweet	Belo Horizonte, Brasil	220	211	3865	FALSO	18/04/2014 21:30	584	0	Portuguese	Twitter for iPhone

Fonte: Elaboração dos autores.

Figura 1 – Base de dados anonimizada construída a partir do Twitter com a biblioteca Snsrape.

Os métodos de análise da base de dados são oriundos do que tem sido popularmente chamado de ciências de dados, as quais são um conjunto de saberes interdisciplinares que envolvem computação, matemática e estatística. Estes são utilizados para construção, manipulação e visualização de dados. Geralmente os saberes das ciências de dados são operados por linguagens de programação específicas que permitem maior dinâmica no controle dos dados, tais como Python e R. Porém, com a popularização do campo e a criação de uma comunidade de usuários, tem sido possível o desenvolvimento de algumas ferramentas mais acessíveis, com interface visual ou de licença de uso livre, como CAQDAS.

CAQDAS são softwares de Análise de Dados Qualitativos Quixilado por Computador⁷, alternativas caso a prática de pesquisa não disponha de um programador para desenvolver abordagens próprias de análise por meio das bibliotecas de programação. O fato de serem programas com funções previamente construídas pode fazer com que surjam restrições

para a solução de problemas de pesquisa em história, ao comparar recortes temporais ou o cruzamento com dados externos à base de dados, seja de conhecimentos bibliográficos ou de outras fontes históricas. Mas, se utilizados em conjunto, permitem contornar as deficiências de cada programa.

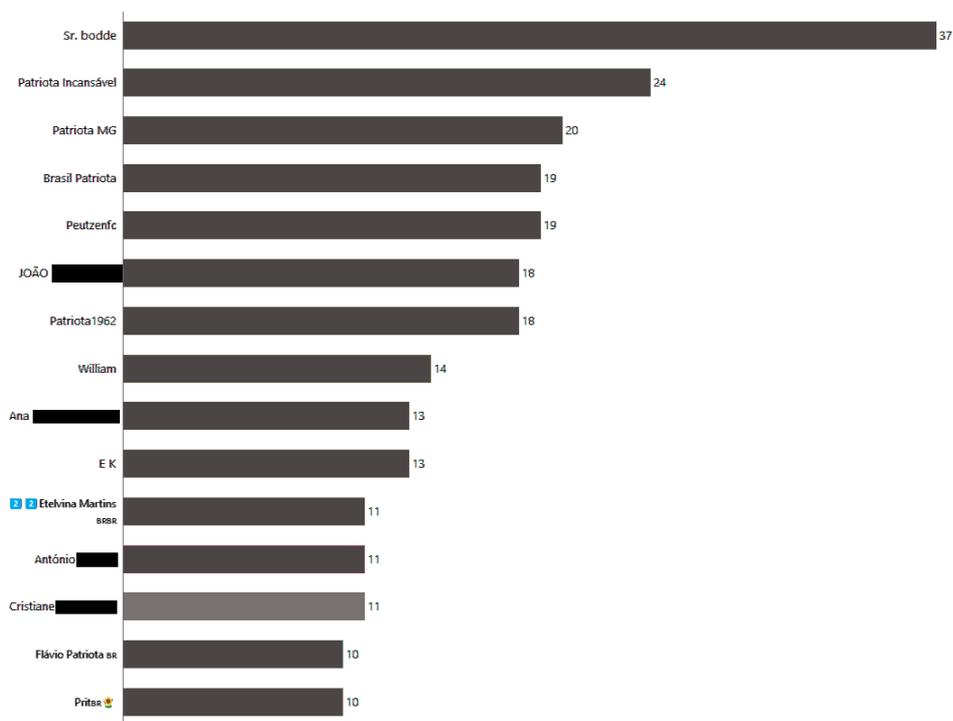
A seguir faremos uma exemplificação de como os conteúdos publicados na rede social são analisados. Para a construção das Figuras 2, 3 e 5: 1) coletamos a base por meio do Snsrscrape, 2) importamos para um CAQDAS chamado MaxQDA⁸, 3) realizamos as análises semânticas via n-gramas⁹ e 4) exportamos os resultados a uma tabela. Para a Figura 6, importamos a base extraída do Twitter, cruzamos as localidades com outra base chamada Simplemaps (2023), a fim de limpar dados imprecisos, e geramos o mapa no Microsoft Power BI¹⁰. A Figura 7 foi gerada no Gephi¹¹, que é uma aplicação especializada em grafos.

A extração dos significados são múltiplas, como: 1) a análise de sentimentos, para avaliar a opinião dos usuários sobre um determinado assunto; 2) a análise de mídia social, para verificar o que as pessoas estão falando sobre eventos, pessoas ou notícias; 3) a análise de rede e tendências para identificar relacionamentos entre usuários do Twitter e como eles estão conectados entre si; 4) a análise geográfica, para determinar a origem dos tweets e verificar como os assuntos são discutidos em diferentes áreas; e 5) a análise de influência para identificar usuários propulsores e entender como suas opiniões e mensagens afetam o público geral.

FORMAS DE ANÁLISES

Para exemplificarmos como os métodos quantitativos e qualitativos das ferramentas CAQDAS são aplicados a uma base de dados coletada a partir do Twitter, fizemos uma seleção de dados circunscritos entre os dias 8 e 9 de janeiro de 2023, quando ocorreu uma tentativa de golpe político no Brasil. Os Ataques às sedes dos Três Poderes do Brasil foram uma série de crimes praticados com o intuito de contestar os resultados das eleições de 2022, por meio da invasão e destruição do patrimônio público. Na manhã do dia 9, após o evento, mapeamos as principais palavras associadas ao termo patrimônio. Buscávamos compreender o que exatamente os usuários do Twitter discutiam sobre as destruições realizadas nos prédios dos três poderes. Logo nos primeiros exercícios heurísticos, um vocábulo chamou a atenção pela frequência de uso: liberdade. Continuando a análise explanatória, filtramos quais contas no Twitter publicaram sobre o assunto. Percebemos que a maioria das contas que postaram entre às dez e onze horas da manhã do dia 9 foram perfis identificados com o ex-presidente

da República, Jair Messias Bolsonaro (Figura 2). A hipótese inicial sugeria que seria um dos adágios utilizados para a tentativa de golpe.



Fonte: Elaboração dos autores.

Figura 2 – Gráfico construído no Excel com os perfis que mais postaram sobre liberdade. As contas pessoais foram anonimizadas.

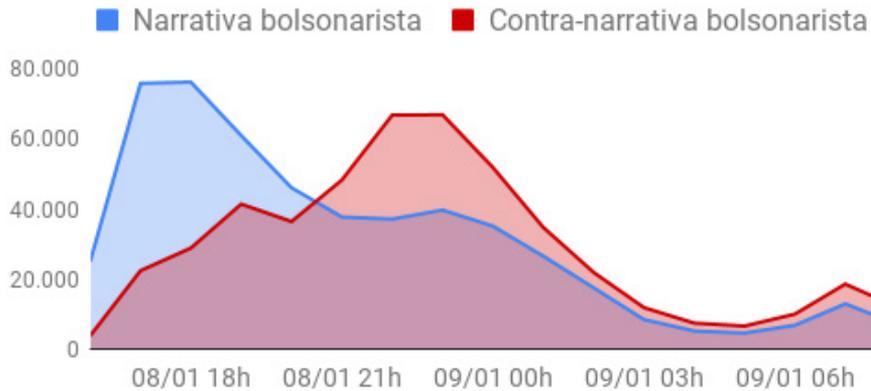
O conceito de liberdade é antigo e encontra capítulos em autores como Sócrates, Locke, Burke entre tantos outros. Ele faz parte dos pressupostos de Constituições nacionais, sendo elaborado conforme os pressupostos filosóficos, históricos e culturais de cada sociedade. Mas o conceito é, também, mobilizado no vocabulário político a partir de diversos sentidos persuasivos. Odilon Caldeira Neto, quanto a isso, observa que a palavra liberdade, adicionada ao adágio fascista “Deus, Pátria e Família”, é um dos sintomas que faz do bolsonarismo um fenômeno político único na história do Brasil, “como uma forma de transformação absoluta da sociedade por meio de um processo de perseguição de inimigos e dos considerados diferentes” (Caldeira Neto, 2023). Neto (2022) tem argumentado que o bolsonarismo faz parte de movimentos neofascistas, diferentes do fascismo histórico, e que tem a internet como meio de propagação global. Ao analisarmos as construções semânticas mais associadas ao termo liberdade chegamos ao resultado presente na Figura 3.

Ranking	Combinação de palavras	Frequência	Porcentagem
1	desceu chama-se bezerros	8960	11,15
1	gado desceu chama-se	8960	11,15
3	nossa sagrada liberdade	8655	10,77
4	poste uma foto	3872	4,82
5	marcha pela liberdade	3692	4,59
6	fazendo ato antidemocrático	2269	2,82
6	liberdade fazendo ato	2269	2,82
6	pq paga imposto	2269	2,82
6	vidraça destruindo vidraça	2269	2,82
10	mil dólares escondidos	2138	2,66
11	mil reais numa	1962	2,44
11	reais numa mala	1962	2,44
13	uma lema vazio	1233	1,53
14	investiga transações financeiras	1149	1,43
15	transações financeiras suspeitas	1148	1,43
16	ministério subordinado à	1147	1,43

Fonte: Elaboração dos autores.

Figura 3 – Frases mais frequentes nos tuítes que mencionaram a palavra liberdade. Os cálculos de n-gramas foram realizados no MaxQDA e exportamos para uma tabela.

Percebemos que as principais frases articuladas conjuntamente com o termo liberdade denotam uma disputa semântica em relação ao termo. Nos casos de “nossa sagrada liberdade”, “marcha pela liberdade” e “liberdade fazendo ato”, nota-se a tentativa de provocar uma instabilidade semântica ao conceito, ao sugerir que o termo é irrestrito. Nega-se o paradoxo do conceito, no qual o limite da liberdade individual é dada pela existência do outro. Todavia, se por um lado é possível visualizar a defesa contundente da liberdade irrestrita, de outro, as críticas ao ato antidemocrático ocupam as primeiras posições. Essa disputa com leve derrocada do bolsonarismo naquele momento também foi notada pelo analista de dados Pedro Barciela (Figura 4)¹².



Fonte: Pedro Barciela, Twitter.

Figura 4 – A mobilização causada pelas primeiras imagens da invasão à Praça dos Três Poderes repercutidas nas redes sociais bolsonaristas foram sufocadas pela crítica à destruição.

No entanto, percebemos outro movimento paralelo ao uso do vocábulo liberdade: uma tentativa de associar os atos golpistas a possíveis “infiltrados”. Ao seguir essa pista, percebemos como a narrativa bolsonarista tentou transferir a responsabilidade pelos atos golpistas ao chamado campo da esquerda. Seriam os infiltrados os responsáveis pela destruição do patrimônio público, conforme a Figura 5.

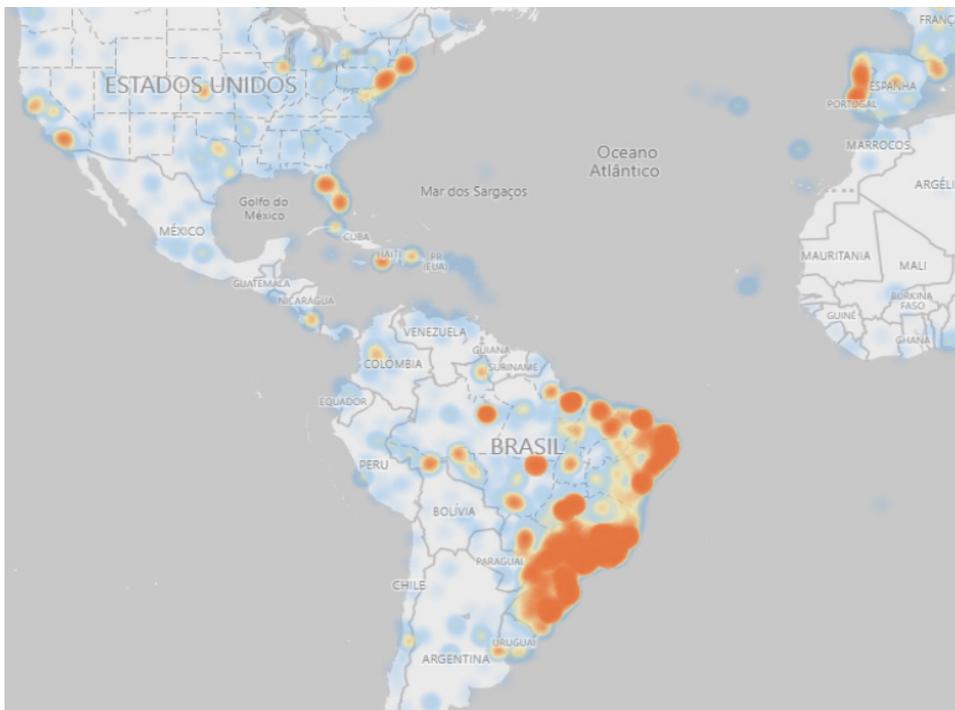
Ranking	Combinação de palavras	Frequência	Porcentagem
1	petistas infiltrados estavam	369	1,40
2	ontem serão responsabilizados	238	0,90
2	vandalismo ontem serão	238	0,90
4	culpa nos infiltrados	173	0,65
5	policiais dando suporte	172	0,65
6	inicio ninguém queria	149	0,56
7	estavam infiltrados organizando	147	0,56
7	existem vídeos comprovando	147	0,56
7	pt estavam infiltrados	147	0,56
10	caminho estava livre	126	0,48

Fonte: Elaboração dos autores.

Figura 5 – Frases mais frequentes nos tuítes que mencionaram a palavra infiltrados. Os cálculos de n-gramas foram realizados no MaxQDA e exportamos para uma tabela.

Nas Figuras 3 e 5 é possível observar mais detalhadamente como ocorre o processamento de linguagem natural por meio da computação. O que estamos fazendo é calcular quais

as frases mais frequentes associadas a uma determinada palavra. Esse exercício é chamado de n-grama, em que o valor de n determina o tamanho da sequência extraída. Considerando que trabalhamos com uma base de dados reduzida, circunscrita apenas entre os dias 8 e 9 de janeiro de 2023 e mapeamos vocábulos específicos, no qual somam-se pouco mais de cem mil publicações, é pouco provável que existam frases mais extensas e iguais. O raciocínio depende da possibilidade de uma frase existir no mundo e que mais de um usuário tenha utilizado a mesma sequência de caracteres para descrever o fenômeno. No caso das Figuras 3 e 5, a sequência mais extensa em carácter quantitativo e qualitativo obtida foi uma tri-grama, isto é, uma combinação de três palavras. Enquanto para “liberdade” as combinações mais frequentes chegam a ocupar 11,15% do total da base de dados, nos “infiltrados” temos pouco mais de 1%. Essas porcentagens permitem mensurar o nível de padronização das estruturas semânticas para a descrição dos fenômenos.

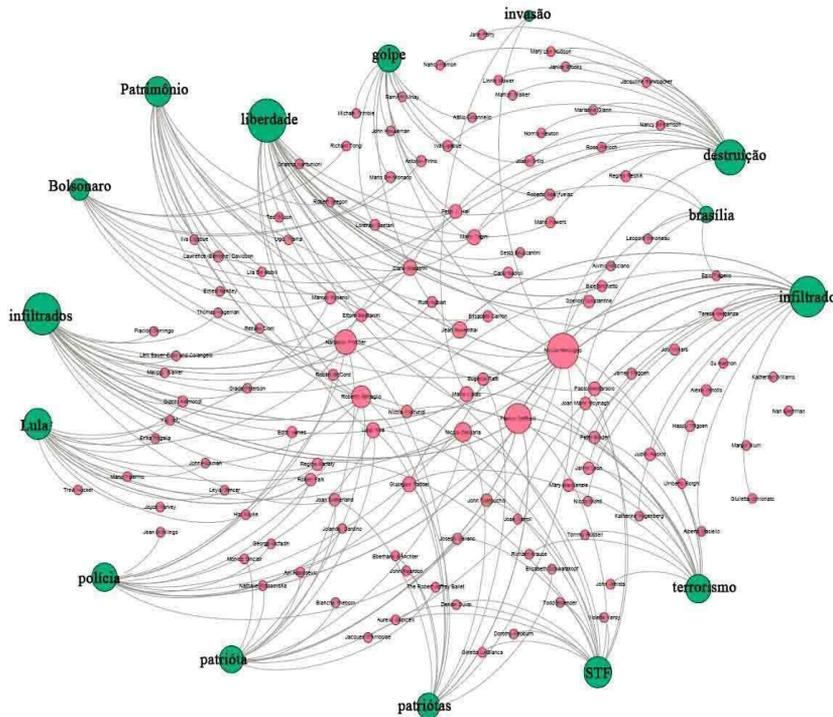


Fonte: Elaboração dos autores.

Figura 6 – Gráfico construído no Microsoft Power BI após o tratamento dos dados sobre localidade com a base Simplemaps (2023). As cores quentes representam maior concentração de publicações.

Na Figura 6, ao analisar as cidades dos tuítes sobre liberdades, observamos publicações que se estendem por todo território nacional e até no exterior. Aqui, nos deparamos com

uma das principais características da rede social, a repercussão global, sem limitações de fronteiras físicas ou geográficas, mesmo que a coleta de dados tenha privilegiado somente a língua portuguesa. Se por um lado o Twitter se caracteriza por sua natureza pessoal e imediata, em que usuários compartilham pensamentos, emoções e experiências em tempo real, por outro, é justamente essa característica da plataforma que permite a propagação e discussão de eventos globais. Quando ocorrem eventos de grande repercussão, seja uma catástrofe natural, um acontecimento político ou um marco histórico, o Twitter se transforma em um centro debates, no qual opiniões e perspectivas pessoais são convertidas em debate público.



Fonte: Elaboração dos autores.

Figura 7 – Grafo de correlacionamento de assuntos com autores. Composto no Gephi.

Na Figura 7 observamos a disposição das principais contas que comentaram sobre a questão dos “infiltrados”. É possível medir a importância dos nós em uma rede segundo o grau de centralidade de um assunto, identificar e mensurar os agrupamentos de nós que têm mais conexões entre si do que com o restante da rede e analisar o fluxo de informação. Essas análises ainda podem ser estendidas em recortes temporais, observando a variação da rede ao longo do tempo. A análise de rede aqui não se limita a um grupo de pessoas

ou assuntos, e sim pode ser aplicada a um único indivíduo, de modo a pôr em perspectiva a sua atuação numa rede social. É preciso observar, ainda, que um acervo digital composto em uma base de dados organizada pode ter seus índices catalográficos dinâmicos. Basta o pesquisador redefinir os parâmetros de busca e visualização, podendo ser ordenado por pessoas específicas, grupos de pessoas, temas, recortes temporais e tantos outros. No exemplo específico sobre os ataques às sedes dos Três poderes prezamos pela organização em torno dos temas “liberdade” e “infiltrados”, mas poderíamos rastrear pessoas específicas e a partir delas redefinir o centro da pesquisa.

CONSIDERAÇÕES FINAIS

Nesta reflexão, buscamos estabelecer uma conexão entre os conceitos tradicionais do debate arquivístico e os desafios do arquivamento digital, tendo como foco o Twitter enquanto um documento histórico. Embora elementos dos tradicionais arquivos pessoais possam ser identificados nas redes sociais, é preciso reconhecer também mudanças. O acesso às informações pessoais é condicionado à própria estrutura de funcionamento da rede social, implicando na padronização de dados e das subjetividades envolvidas no processo. Esse condicionamento nos coloca em dúvida a possibilidade de um arquivo pessoal digital, como no analógico, mas permite, ao menos, produzir evidências históricas que serão melhor compreendidas no futuro. Nesse sentido, é papel do Estado garantir tanto a privacidade quanto a salvaguarda dos dados pessoais. Para o desenvolvimento do experimento foi fundamental aplicar métodos digitais de coleta e arquivamento de dados, os quais possibilitam transformar o conteúdo de uma rede social em material analisável para estudos históricos. A seleção e a preservação do conteúdo das redes sociais em um arquivo pesquisável e escalonável dependem de práticas ligadas à ciência aberta, como a gestão dos próprios dados. Para desenvolver uma abordagem computacional voltada à pesquisa histórica, é imprescindível combinar conhecimentos interdisciplinares próprios da história digital, que intercala os saberes historiográficos tradicionais com métodos computacionais e mídias. Lançando mão desses raciocínios, compreendemos que o Twitter é composto majoritariamente por dados pessoais de múltiplos significados que são publicizados, mas dependem de fatores algorítmicos para serem visualizados e realmente tornarem-se públicos. Essas postagens aparentemente isoladas na rede podem oferecer aberturas para pensar a vida cotidiana, as opiniões e as emoções das pessoas em uma perspectiva histórica que ainda precisa ser melhor desbravada. Explorar esse tipo de conteúdo pode enriquecer nosso entendimento sobre a sociedade e as interações humanas em um contexto digital, em que a prática de pesquisa em história estará cada vez mais inserida.

NOTAS

1 As bases de dados coletadas para o experimento estão disponíveis no Repositório de Dados de Pesquisa da Unicamp, disponível em: <https://doi.org/10.25824/redu/SPBXDM>.

2 Para os casos mais abrangentes de navegação existe a Lei n. 12.965/2014, chamada também de Marco Civil da Internet.

3 *Scrapbooks* são coleções de memórias pessoais e lembranças frequentemente criadas por indivíduos como uma maneira de preservar sua história pessoal. Sua origem é moderna, tendo se popularizado nos séculos XIX e XX.

4 No início de 2023 foi liberado parte do conjunto de algoritmos que compõe a recomendação de conteúdo aos usuários do Twitter. Trata-se de um evento novo para os padrões das redes sociais, que mantêm em segredo seu funcionamento, e por isso é necessária precaução com a iniciativa. Através do código disponibilizado é possível ter noção sobre os pesos diferentes dado para cada tipo de ação do usuário ao deslizar as páginas e interagir com os botões, mas ainda não é possível determinar exatamente como funciona a rede.

5 Atualmente é possível coletar apenas 1,5 mil tweets por mês gratuitamente.

6 Em programação, uma biblioteca é um conjunto de funções, classes e outros recursos que fornecem funcionalidades prontas e testadas para tarefas específicas.

7 Traduzimos o termo “Computer-assisted” para “auxiliada por computador” visando um vocabulário que representa melhor a relação entre humanidades e computação.

8 O Maxqda é uma ferramenta de análise quantitativa e qualitativa de dados criada para auxiliar pesquisadores a organizar, analisar e visualizar dados de pesquisa qualitativa, como entrevistas, questionários abertos, transcrições de áudio e vídeo, texto e imagens. Alternativas a ela são Atlas.Ti e Nvivo, também adquiridas sob licença de uso. Essas três ferramentas foram projetadas para abarcar inúmeros tipos de pesquisa em humanidades, conservando uma interface amigável de uso. Suas funções permitem ler, rastrear, comparar, quantificar dados e criar categorias de análise para diversos tipos de mídias. Há opções de livre acesso, mas que não contam com o mesmo número de recursos e atualizações frequentes, como RQDA, TAMS Analyzer, AQDAS, CATMA, Transana, OpenVV e Weft QDA.

9 N-gramas são uma técnica de processamento de linguagem natural para análise de texto que consiste em extrair sequências contínuas de n tokens (palavras, letras, fonemas etc) de um texto. Exemplificamos no próximo tópico.

10 O Power BI é uma ferramenta de visualização de dados desenvolvida pela Microsoft que permite aos usuários criar painéis e relatórios interativos a partir de uma variedade de fontes

de dados, incluindo bancos de dados relacionais, planilhas, serviços em nuvem e outras fontes. São recursos avançados de análise de dados, por meio de gráficos, tabelas e mapas. A ferramenta pode ser gratuita ou paga, dependendo dos recursos utilizados. Alternativas a ela são Tableau, Google Data Studio e LibreOffice.

11 O Gephi é uma ferramenta de software livre para análise e visualização de redes complexas, tais como redes sociais, biológicas e de transporte. Ele permite importar dados de várias fontes e formatos, analisar e manipular os dados para criar visualizações interativas e personalizadas que ajudam a identificar padrões e estruturas internas nas redes. Uma alternativa a ela é o Iramuteq, também com licença livre.

12 Apesar de Pedro Barciela não demonstrar os critérios utilizados em sua análise, é um dos poucos canais que monitoram em tempo real eventos políticos brasileiros com dados do Twitter.

REFERÊNCIAS

ABREU, R. *A fabricação do imortal: memória, história e estratégias de consagração no Brasil*. Rio de Janeiro: Rocco, 1996.

ARENDT, H. *A condição humana*. Rio de Janeiro: Forense Universitária, 2007.

AYERS, E. L. The Pasts and Futures of Digital History. *History News*, Richmond, v. 56, n. 4, p.5-9, 2001.

BACH, M. Arquivamento da web: estudos de caso internacionais e o caso brasileiro. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 16, n. 1, p. 2-24, 2018.

BRASIL. Lei n. 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei n. 8.112, de 11 de dezembro de 1990; revoga a lei n. 11.111, de 5 de maio de 2005, e dispositivos da Lei n. 8.159, de 8 de janeiro de 1991; e dá outras providências. Brasília, DF: *Diário Oficial*, 2011.

BRASIL. Lei n. 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: *Diário Oficial*, 2018.

BRASIL. Lei n. 8.159, de 8 de janeiro de 1991. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. Brasília, DF: *Diário Oficial*, 1991.

CALDEIRA NETO, O. "Brasil é um laboratório da extrema direita global". [Entrevista cedida a] Laís Modelli. *DW*, São Paulo, 11 de janeiro de 2023. Disponível em: <https://www.dw.com/pt-br/brasil-%C3%A9-um-laborat%C3%B3rio-da-extrema-direita-global/a-64351759>. Acesso em: 13 jan. 2023.

CALDEIRA NETO, O. Neo-Fascism in Brazil: From the Local to the Global? *Esboços*, Florianópolis, v. 29, n. 52, p. 579-598, 2022.

CALVO, E.; ARUGUETE, N. Fake news, trolls y otros encantos: cómo funcionan (para bien y para mal) las redes sociales. Buenos Aires: Siglo XXI, 2020.

CASSINO, J. F. et al. *Colonialismo de dados: como opera a trincheira algorítmica na guerra neoliberal*. São Paulo: Fundação Perseu Abramo, 2022.

- CRIBB, J. *et al.* *Open Science: Sharing Knowledge in the Global Century*. Clayton: Csiro, 2010.
- ELLISON, N. B. Social Network Sites: Definition, History, Scholarship. *Journal of Computer-Mediated Communication*, Hoboken, v. 13, n. 1, p. 210-230, 2007.
- FISHER, R. Por uma teoria dos arquivos privados: revendo os escritos fundadores de Jenkinson e Schellenberg. In: NEDEL, L.; HEYMANN, L. Q. (org.). *Pensar os arquivos: uma antologia*. Rio de Janeiro: Editora FGV, 2018. p. 329-359.
- GARVEY, E. G. *Writing with Scissors: American Scrapbooks from the Civil War to the Harlem Renaissance*. Oxford: Oxford University Press, 2012.
- GITHUB. Search. Busca realizada: Twitter, 2023. Disponível em: <https://github.com/>. Acesso em: 2 set 2023.
- GOMES, A. C.; HANSEN, P. S. *Intelectuais mediadores: práticas culturais e ação política*. São Paulo: José Olympio, 2016.
- GOMES, A. C. *Em família: a correspondência de Oliveira Lima e Gilberto Freyre*. Campinas: Cecult, 2005.
- GOOD, K. D. From Scrapbook to Facebook: A History of Personal Media Assemblage and Archives. *New Media and Society*, Londres, v. 15, n. 4, p. 557-573, 2013.
- HADDOW, G.; HADDOW, K. S. *Disaster Communications in a Changing Media World*. Waltham: Butterworth-Heinemann, 2013.
- HOBBS, C. O caráter dos arquivos pessoais: reflexões sobre o valor dos documentos de indivíduos. In: NEDEL, L.; HEYMANN, L. Q. (org.). *Pensar os arquivos: uma antologia*. Rio de Janeiro: Editora FGV, 2018. p. 261-274.
- IUMATTI, P. T.; NICODEMO, T. L. Arquivos pessoais e a escrita da história no Brasil: um balanço crítico. *Revista Brasileira de História*, Marília, v. 38, n. 78, p. 97-120, 2018.
- KATRITZKY, M. A. *The Art of Commedia: A Study in the Commedia dell'Arte 1560-1620 with Special Reference to the Visual Records*. Amsterdam: Rodopi, 2006.
- KEMP, S. Digital 2022: Global Overview Report. *Datareportal*, Singapore, 26 jan. 2022. Disponível em: <https://datareportal.com/reports/digital-2022-global-overview-report>. Acesso em: 12 dez. 2022.
- LINDLEY, S. E. *et al.* Rethinking the web as a personal archive. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 22., 2013, Rio de Janeiro. *Anais [...]*. Rio de Janeiro: WWW, 2013. p. 1-11.
- LUCCHESI, A. Conversas na antessala da academia: o presente, a oralidade e a história pública digital. *História Oral*, Niterói, v. 17, n. 1, p. 39-69, 2014.
- LUCCHESI, A.; SILVEIRA, P. T.; NICODEMO, T. L. Nunca fomos tão úteis. *Esboços*, Florianópolis, v. 27, n. 45, p. 161-169, 2020.
- MAGRANI, E. *Entre dados e robôs: ética e privacidade na era da hiperconectividade*. Porto Alegre: Arquipélago, 2019.
- MANOVICH, L. *The Language of New Media*. Cambridge, MA: The MIT Press, 2002.
- MANOVICH, L. Visualização de dados como uma nova abstração e anti-sublime. In: LEÃO, L. (org.). *Derivas: cartografias do ciberespaço*. São Paulo: Sesc, 2006. p. 149-162.
- MORREALE, L.K. History as Antidote: The Argument for Documentation in Digital History. *History and Theory*, Middletown, v. 61, n. 4, p.64-76, 2022.

- NEDEL, L.; HEYMANN, L. Q. *Pensar os arquivos: uma antologia*. Rio de Janeiro: Editora FGV, 2018.
- NICODEMO, T. Manifesto pela soberania de dados. *Blog Thiago Nicodemo*, Brasil, 15 nov. 2022. Disponível em: <https://thiagonicodemo.substack.com/p/manifesto-pela-soberania-de-dados?r=1t33kd&utm>. Acesso em: 29 dez. 2022.
- NICODEMO, T. ROTA; A.; KISIL, I. M. Introdução. In: NICODEMO, T.; ROTA; A.; KISIL, I. M. (org.). *Caminhos da história digital no Brasil*. Vitória: Mil Fontes, 2022.
- PAWLICKA-DEGER, U. The Laboratory Turn: Exploring Discourses, Landscapes, and Models of Humanities Labs. *DHQ*, Boston, v. 14, n. 3, p. 1-22, 2020.
- ROCKEMBACH, M. Arquivamento da web: estudos de caso internacionais e o caso brasileiro. *Revista Digital de Biblioteconomia e Ciência da Informação*. Campinas, v. 16, n. 1, p. 7-24, 2017. DOI: 10.20396/rdbci.v16i1.8648747. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8648747>. Acesso em: 28 abr. 2023.
- ROGERS, R. *Digital methods*. Cambridge, MA: The MIT Press, 2013.
- ROSENZWEIG, R. *Clio Wired: The Future of the Past in the Digital Age*. New York: Columbia University Press, 2011.
- SCHWAITZER, L. B. LGPD e acervos históricos: impactos e perspectivas. *Archeion Online*, João Pessoa, v. 8, n. 2, p. 36-51, 2020.
- SILVEIRA, P. T. *História, técnica e novas mídias: reflexões sobre a história na era digital*. 2018. Tese (Doutorado em História) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.
- SIMPLEMAPS. Cidades do Mundo, 2023. Disponível em <https://simplemaps.com/data/world-cities>. Acesso em: 31 mar. 2023.
- SNSCRAPE. Documentação, 2023. Disponível em: <https://github.com/JustAnotherArchivist/snscrape>. Acesso em: 31 mar. 2023.
- TORNES, A. Um novo passo para o futuro da pesquisa acadêmica com a API do Twitter. *Blog [do] Twitter*, Brasil, 26 jan. 2021. Disponível em: https://blog.twitter.com/pt_br/topics/product/2019/-um-novo-passo-para-o-futuro-da-pesquisa-academica-com-a-api-do-. Acesso em: 2 maio 2022.
- TUFEKCI, Z. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. New Haven: Yale University Press, 2017.
- TWITTER. Sobre tweets públicos e protegidos. Central de Ajuda, 2023a. Disponível em: <https://help.twitter.com/pt/safety-and-security/public-and-protected-tweets>. Acesso em: 2 jan. 2023.
- TWITTER. Twitter API v2, 2023b. Disponível em: <https://developer.twitter.com/en/docs/twitter-api>. Acesso em: 31 mar. 2023.
- WINTERS, J. Coda: Web archives for humanities research – some reflections. In: BRÜGGER, N. (Org.). *The Web was History*. Londres: UCL Press, 2017. p. 238-248.