

Articulação entre métricas e dados textuais como experimentação metodológica para os estudos em circulação

Articulación entre métricas y datos textuales como experimentación metodológica para los estudios en circulación

Articulation between metrics and textual data as methodological experimentation for circulation studies

DOI: <https://doi.org/10.1590/1809-58442024114pt>

Luan Moraes Romero

<https://orcid.org/0000-0003-4495-6672>

Viviane Borelli

<https://orcid.org/0000-0003-0643-2173>

Universidade Federal de Santa Maria, Programa de Pós Graduação em Comunicação. Santa Maria, RS - Brasil

Resumo

O objetivo do artigo é experimentar possibilidades de articulação entre métricas e dados textuais nas pesquisas em circulação. Trata-se de um recorte de pesquisa mais ampla sobre os circuitos constituídos em torno das três primeiras reportagens da série Vaza Jato publicada pelo The Intercept Brasil, em 9 de junho de 2020. Para nortear a discussão, fazemos uma breve menção a alguns estudos sobre os processos de circulação e discutimos a abordagem teórica de Braga (2017) sobre a noção de episódios comunicacionais. Para mostrar as possibilidades de integração entre as métricas e a perspectiva inferencial, referida pelo autor, apresentamos o episódio nomeado como “Cara de Pau”. Foi por meio da observação de pistas deixadas pelos rastros digitais, que conseguimos fazer uma articulação entre os dados quantitativos gerados por softwares e os textos.

Palavras-Chave: Circulação; Métrica; Dados textuais; Episódios comunicacionais; Vaza-Jato.

Resumen

El objetivo del artículo es experimentar las posibilidades de articulación entre métricas y datos textuales en la investigación en circulación. Este es un extracto de una investigación más amplia sobre los circuitos formados alrededor de los primeros tres reportajes de la serie Vaza Jato publicados por The Intercept Brasil, el 9 de junio de 2020. Para orientar la discusión, mencionamos brevemente algunos estudios sobre procesos de circulación y discutimos el enfoque teórico de Braga (2017) sobre la noción de episodios comunicacionales. Para mostrar las posibilidades de integración entre la métrica y la perspectiva inferencial, a la que se refiere el autor, presentamos el episodio denominado “Cara de pau”. Fue a través de la observación de las pistas dejadas por las huellas digitales que logramos establecer una conexión entre los datos cuantitativos generados por el software y los textos.

Palabras clave: Circulación; Métrica; Datos textuales; Episodios de comunicación; Vaza-jato.

Abstract

This work's objective was to experiment the possibilities of articulation between metrics and textual data in researches focused in circulation of meanings. This is an excerpt from a broader research on the circuits formed around the first three reports of the Vaza Jato series published by The Intercept Brasil, on June 9, 2020. To guide the discussion, we briefly mention some studies on circulation processes and we discuss Braga's (2017) theoretical approach on the notion of communicational episodes. To show the possibilities of integration between metrics and the inferential perspective, as referred by the author, we present the episode named “Shameless”. It was through the observation of clues left by the digital tracks that we managed to make a connection between the quantitative data generated by software and the texts.

Keywords: Circulation; Metrics; Textual data; Communication episodes; Vaza-Jato.

Introdução

Em março de 2021, o ministro do Supremo Tribunal Federal (STF) Edson Fachin anulou todas as condenações impostas ao ex-presidente Luis Inácio Lula da Silva pela Justiça Federal do Paraná que tinham relação com as investigações da Operação Lava Jato¹. O The Intercept Brasil (TIB) aponta em *newsletter*, enviada para seus leitores em 8 de março de 2021, que o juiz reconhece algo que o procurador Deltan Dallagnol sabia, e que a agência de notícias revelou em 2019. Parte do texto refere: “Curitiba não era vara competente para o processo do triplex, já que essa vara só julgava denúncias relacionadas à Petrobras. Só o fato de esse processo estar sendo tramitado no Paraná constitui uma fraude processual evidente”.

Essa constatação faz parte de uma das três primeiras reportagens da série Vaza Jato publicadas pelo TIB, em 9 de junho de 2020 e que foram objeto de investigação em dissertação de mestrado que analisou como ocorreu a circulação de sentidos em plataformas digitais (VAN DIJCK, POELL & DE WAAL, 2018). Essa pesquisa de escopo maior identificou e descreveu a constituição de um circuito comunicacional (BRAGA, 2017) em torno dessas reportagens, que integram plataformas distintas utilizadas pelo TIB, como o site, a *fanpage* e o canal do Youtube. A partir da perspectiva desenvolvida por Braga (2017), cada um desses pontos nodais evidencia episódios comunicacionais que foram analisados de forma mais detalhada em investigação de mestrado (ROMERO, 2021).

Partindo desse contexto mais amplo, elegemos como foco para esse artigo² a integração entre perspectivas metodológicas que consideram métricas geradas com o auxílio de softwares, como o **Iramuteq**, com uma perspectiva inferencial sobre a comunicação (BRAGA, 2017). Assim, a proposta se baseia em táticas analíticas, que envolvem revisitar os textos e ponderar os sinais revelados pelas métricas, com o intuito de realizar inferências que permitam extrair conhecimento relevante para abordar a questão de pesquisa. No nosso estudo, consideramos o termo “textos” para referir tanto o que foi produzido pelas reportagens publicadas pelo The Intercept Brasil quanto o que foi construído pelos distintos atores sociais nos espaços de comentários do site, da Fanpage e do canal do Youtube³.

Para isso, partiremos de uma breve reflexão sobre algumas perspectivas relacionadas aos estudos em circulação no campo da Comunicação, que desencadeiam abordagens metodológicas diversas, para depois podermos avançar no delineamento de uma concepção de integração de diferentes formas de análise. Por meio dessa experimentação metodológica, pretendemos contribuir para o enriquecimento tanto teórico quanto metodológico das pesquisas em circulação.

Em um segundo momento, para corroborar tal percepção, apresentamos a processualidade de um episódio comunicacional, denominado por nós como “Cara de Pau”, que tem materialização no vídeo publicado no canal do Youtube intitulado “Sérgio Moro diz que não é juiz investigador”⁴ no dia 9 de junho de 2019. A partir da eleição, descrição e análise desse episódio comunicacional, evidenciaremos a importância da integração entre analíticas com base em métricas aplicadas a dados textuais com a tática metodológica de retorno aos textos.

Estudos em circulação e abordagem metodológica

Ao iniciarmos a reflexão teórica é importante referir brevemente algumas contribuições sobre como a instância de circulação tem emergido como uma visada epistemológica para os estudos em Comunicação. As investigações que têm como base a articulação teórica relacionada à circulação se pautam por uma diversidade de abordagens teóricas. Segundo Grohmann (2019), seria possível abarcar de maneira tentativa, perspectivas que se filiam a pelo menos três horizontes teóricos: aquelas que dialogam com análises sociodiscursivas, outras que

1 Informações disponíveis em: <<https://g1.globo.com/politica/noticia/2021/03/08/entenda-a-decisao-de-fachin-que-anulou-as-condenacoes-de-lula-e-o-que-acontece-agora.ghhtml>> . Acessado em 21 mar 2021. No dia 23 de março de 2020, o STF também considerou que o então juiz Sérgio Moro fora parcial nos julgamentos do ex-presidente Luiz Inácio Lula da Silva na Operação Lava Jato. In: <https://noticias.uol.com.br/ultimas-noticias/bbc/2021/03/23/com-voto-de-kassio-nunes-stf-forma-maioria-para-rejeitar-recurso-de-lula-contra-moro.htm>. Acessado em 30 mar 2021.

2 Consideramos que todos que se inscrevem nessas diferentes plataformas são participantes desses circuitos comunicacionais. Como o foco do artigo não é uma discussão conceitual de termos utilizados na área de Comunicação como “ator, sujeito, leitor, produtor, coprodutor, comentarista, interagente, prosumer, receptor, usuário”, entre outros, vamos denominá-los atores sociais ou participantes, em função da proposição teórica de Braga (2017) e Verón (2004, 2013) e também para termos uma leitura mais fluída.

3 Tais reflexões foram baseadas nas investigações do sociólogo alemão Niklas Luhmann, especialmente na obra LUHMANN, Niklas. Social Systems. Stanford. Stanford University Press, [1984] 1995.

4 Não cabe aqui uma discussão mais profunda das obras dos autores, mas sinteticamente, destacamos que Sebeok e Umiker-Sebeok (2004) problematizam o método adotado por Charles Peirce e por Sherlock Holmes, que baseiam-se no olhar dos pequenos fatos e das especificidades. Para eles, só depois de colher uma série de indícios é que será possível fazer inferências. Já Ginzburg (1989, p. 177) compreende que o paradigma indiciário ou semiótico acabou penetrando variados âmbitos do conhecimento “modelando profundamente as ciências humanas. Minúsculas particularidades paleográficas foram empregadas como pistas que permitiam reconstruir trocas e transformações culturais”.

se pautam a refletir a partir de uma perspectiva culturalista em sincronia com os estudos culturais, e ainda uma abordagem que busque alinhar as reflexões sobre a circulação do capital com a circulação de sentidos.

Dentre elas, nos filiamos às investigações de cunho semiodiscursivo que buscam desenvolver uma perspectiva crítica sobre a apropriação de dados coletados, com olhar atento aos rastros deixados pelos distintos atores sociais nas mais diversas plataformas digitais. Dialogamos, assim, com as proposições de GINDIN e BUSSO (2018) acerca da semiodata e a necessidade de crítica e de uma abordagem interpretativa e crítica. Compartilhamos de tais preocupações, de que não deve haver um “encantamento dos dados” ou mera descrição, visto que tais signos devem ser analisados, bem como é necessário fazer uma reflexão crítica de tais processos de produção de sentidos (Borelli, no prelo).

Assim, concebemos, como Fausto Neto (2018), que as investigações em circulação por uma abordagem semiodiscursiva perpassam diferentes momentos e angulações possíveis, quando consideramos o contexto latino-americano de pesquisas em Comunicação. Para o autor, tais indagações emergem na obra de Eliseo Verón em distintos momentos e abordam a circulação enquanto um espaço de defasagem discursiva (VERÓN, 2004). Como destaca Fausto Neto (2018, p. 15), a circulação vai além de uma “zona de passagem” e representa muito mais que um “elo intermediário”.

Ao longo de sua produção acadêmica, o semiólogo argentino atualizou suas concepções diante da diversidade das realidades comunicacionais que observava. No seu último livro publicado, Verón (2013) refere que não seria mais possível considerar a circulação somente como um lugar de defasagem, um lugar de passagem entre produção e reconhecimento, já que observava lógicas de acoplamentos estruturais⁵ que complexificam essas relações.

Nos apoiamos na concepção de Braga (2017) que elabora uma teoria intermediária sobre os episódios, dispositivos de interação e circuitos comunicacionais. Consideramos que uma das contribuições dessa articulação teórica é a possibilidade de utilizar metodologias a partir de um paradigma indiciário, proposto por Ginzburg (1989) e discutido por Sebeok e Umiker-Sebeok (2004)⁶. Ambos problematizam que para a construção do conhecimento é necessário seguir vestígios, sinais e revelações que, muitas vezes, se manifestam nos pormenores e nos detalhes e que acabam passando despercebidos. No caso da nossa proposta, as métricas geradas pelo Iramuteq demandaram um olhar mais atento aos dados textuais, pois foi preciso tensionar o que fora dito pelos distintos atores sociais em relação aos dados quantitativos. Dessa maneira, foi a partir da observação de pistas e de rastros que conseguimos fazer inferências.

Os autores inspiram Braga (2008, 2017) a refletir sobre os estudos em Comunicação por meio de uma perspectiva inferencial. Para o autor, é a inferência que “permite atualizar o código, exercê-lo para além de um processo mecânico de codificação/decodificação, que não faria mais que transmitir informação” (BRAGA, 2017, p.31). Com base nessa perspectiva, Braga (2017) concebe que a comunicação se dá através de processos inferenciais entre atores sociais que compartilham códigos em comum. Dessa maneira, para o autor, a formação de dispositivos interacionais se dá a partir de uma reiteração estratégica de tentativas de compreensão mútua no tecido social. Os episódios comunicacionais seriam a ordem da materialidade e da singularidade das trocas. Braga et al (2017) utiliza como exemplo as aulas: se fossem sobre um determinado assunto seriam algo da dimensão do dispositivo, já a materialidade do encontro do professor com seus alunos em uma única aula seria seu episódio. Dessa maneira, para ele, o circuito comunicacional seria constituído a partir da articulação entre diferentes dispositivos que se entrelaçam.

Seguindo tais compreensões, Borelli & Kroth (2020) desenvolveram pesquisa em que discutem as reconfigurações do dispositivo radiofônico diante da emergência da circulação e da complexificação do processo de mediação da sociedade. Já Frigo, Romero e Borelli (2019) articularam a noção de dispositivo interacional (BRAGA et al, 2017) com a de plataforma (VAN DIJCK, POELL & DE WAAL, 2018), evidenciando a necessidade de haver códigos e inferências para a constituição de um arranjo disposicional de interação (BRAGA, 2017).

É importante referir que Van Dijck, Poell & De Waal (2018) concebem que a plataforma da sociedade tem passado por três processos: a) datificação, b) commodificação e c) seleção. Para os autores, tais processos afetam o setor de notícias, que passam a depender de empresas de tecnologia infra estruturais, tais como Facebook e Google. Além disso, ocorrem mudanças de ordem editorial, visto que muitas organizações jornalísticas e seus editores passam a tomar decisões importantes e rever estratégias de distribuição com base em dados produzidos pelas plataformas digitais.

5 Tradução nossa para “betweenness centrality”. Recuero (2017) traduz como “grau de intermediação”, na obra RECUERO, R. Introdução à Análise de Redes Sociais Online. 1. ed. Salvador: EDUFBA, 2017. v. 1. 101p, por lidar com redes em que os nós são perfis sociais, assim compreendemos que ambas expressões são adequadas.

6 Para destacar os textos utilizamos o duplo recurso do negrito e do sublinhado para os vocábulos selecionados, e somente negrito para guiar a síntese da relevância do destaque.

Assim, é importante termos a compreensão sobre o funcionamento de tais processos no contexto do desenvolvimento de plataformas digitais, pois nas investigações científicas que se debruçam sobre rastros digitais se faz necessário problematizar a contextualização dos dados obtidos (BREITER e HEPP, 2018), como é o processo de captura, tratamento e análise dos dados.

Entretanto, diante da impossibilidade de ampliar tais questões sobre o funcionamento das plataformas, a plataformização e as distintas repercussões na sociedade e nos modos de interação e de comunicação (reconhecemos sua importância, mas não são centrais nessa reflexão), passamos a especificar como construímos nosso percurso metodológico com base nos gráficos e métricas geradas pelo software livre de análise lexicométrica Iramuteq.

É importante considerar que a aplicação de métodos quantitativos podem ajudar a identificar padrões e outras formas observáveis que não são possíveis por outras metodologias. Dessa forma, entendemos que a utilização de softwares para coleta de dados pode se constituir numa potência para problematizar tais processos de produção de sentidos em circulação. Entretanto, a perspectiva quantitativa é tensionada com um olhar qualitativo, sem pretensão de colocar essas duas abordagens em contraposição, mas sim, refletir sobre suas especificidades e transversalidades.

Integramos a análise estatística, a nuvem de palavras e os gráficos de árvore máxima da análise de similitude (DEGENE e VERGES, 1973) que leva em consideração a métrica de centralidade de intermediação (NEWMAN e GIRVAN, 2004; BRANDES, 2001) aplicada a dados textuais. Cada uma delas contribuiu como uma forma de olhar nossos dados empíricos e, em conjunto, de forma integrada, nos deram subsídios para voltar aos textos produzidos e observar os sentidos em contexto, assim como a processualidade do episódio comunicacional.

O processo de extração e mineração de dados é intrincado e abarca diversas etapas. A análise estatística das palavras empregadas em um texto específico foi realizada tanto para fazer uma depuração inicial dos dados textuais, conduzida manualmente, visando mitigar disparidades entre expressões comuns e nomes próprios. Tal procedimento é crucial para aprimorar as visualizações, como tem sido problematizado em outras reflexões sobre potencialidades, limitações e possibilidades na utilização de softwares para análises e visualização de dados na área da Comunicação (WOBETO, ROMERO E BORELLI, 2023).

Adicionalmente, desenvolvemos métricas como a porcentagem de ocorrência de palavras analisadas em relação ao total de ocorrências, a porcentagem de frequência de uma palavra específica em relação ao total de ocorrências e a média de ocorrências por corpus. Com a elaboração das listas de palavras, é viável retornar aos textos originais com base em uma palavra escolhida pelo pesquisador, assim como por meio da seleção de um grupo de palavras. Esta abordagem revelou-se significativa para retornar aos textos originais a partir da seleção de vocábulos, fornecendo uma noção da relevância de determinada palavra ou conjunto em relação às demais empregadas em um dado contexto.

Ademais, quando é feita uma nuvem de palavras, se parte de uma noção frequencial dos usos de determinado termo, em relação ao conjunto total de palavras em um texto. É um gráfico visual que pode ajudar em determinados casos a perceber quais são as temáticas de maior frequência em um texto, ou conjunto de textos. Em nossa pesquisa, as inferências feitas a partir dela, se dão em articulação conjunta com os gráficos de árvore máxima.

A análise de similitude é resultado da incorporação dos estudos sobre a teoria dos grafos na matemática pelos estudos linguísticos (MARCHAND; RATINAUD, 2012). O Iramuteq, ao realizar esse tipo de análise, gera um gráfico de uma rede formada, tendo como nós as palavras integrantes do texto, e atribui pesos às arestas que os conectam pela medida de co-ocorrência. Assim, de acordo com a proximidade e a frequência de uma palavra com as outras em determinados textos, organiza as arestas. Ou seja, com maior frequência e maior proximidade, a conexão entre elas é maior.

Há diversas formas de se observar e construir esse gráfico, contudo, utilizamos os algoritmos que formam a árvore máxima dessas relações entre as palavras em um determinado texto que levam em consideração a métrica de centralidade de intermediação⁷ (NEWMAN e GIRVAN, 2004; BRANDES, 2001). Antes de aprofundarmos as noções que envolvem tal métrica, é importante compreender que uma árvore máxima é uma forma de simplificação extrema, fazendo emergir de maneira visualmente maior, os nós que representam máximos, ou seja, que são articulados com maior frequência e em maior proximidade com outros nós (DEGENE e VERGÈS, 1973).

Dessa maneira, a métrica utilizada se baseia na intensidade com que determinado nó (em nosso contexto, palavras) fica entre outros dois nós (NEWMAN e GIRVAN, 2004; BRANDES, 2001). Posto de maneira contextual, na nossa investigação, quanto mais uma palavra é utilizada em determinado texto em conjunto com outras palavras ela assume uma importância local, mesmo que em termos frequenciais com relação ao todo da rede, ela tenha pouca

⁷ Uma primeira versão do artigo foi apresentada ao Grupo de Trabalho Recepção, Circulação e Usos Sociais das Mídias do XXX Encontro Anual da Compós, Pontifícia Universidade Católica de São Paulo, São Paulo - SP, 27 a 30 de julho de 2021. Após discussão e revisão teórico-metodológica, o caso a ser detalhado nesta reflexão foi mudado.

intensidade. A partir dessa métrica, é possível destacar agrupamentos de palavras que possuem afinidade em seus usos recorrentes em determinado texto, organizando-as em classes.

Assim, os gráficos de árvore máxima foram gerados pelo Iramuteq e exportados para serem tratados no software de visualização de dados Gephi. Esse tratamento foi feito para destacar a métrica de centralidade de intermediação, apresentando a intensidade de maneira gradativa crescente - da cor azul a vermelha - nos gráficos que compõem essa investigação. Ademais, foi efetuado o cálculo da modularidade da cada uma das árvores, para perceber se a formação dos agrupamentos de palavras é minimamente confiável, como destacam Newman e Girvan (2004).

Apresentamos a seguir a aplicação da analítica proposta, e destacamos alguns procedimentos efetuados. Dessa maneira, é necessário dizer que destacamos, apenas as classes morfológicas e: adjetivo, substantivos, substantivos complementares, verbos, verbos suplementares e formas não reconhecidas. Consideramos que nossa escolha se baseia em nossos objetivos de pesquisa que buscam compreender como os sentidos circulam, assim como Camargo e Justo (2013) corroboram a noção de tais classes mostrarem palavras com sentidos pleno. Concebemos que mesmo se determinado vocábulo seja utilizado em sua polissemia, com o movimento de volta aos textos poderemos avaliar essa questão.

Os gráficos de árvore máxima foram feitos no software Iramuteq, depois transferidos para o Gephi, no qual aplicamos os algoritmos de visualização “não sobreposição”. Efetuamos o cálculo da modularidade e destacamos o tamanho das palavras de acordo com sua frequência. As cores das arestas mostram a intensidade da métrica centralidade de intermediação indo do azul ao vermelho de maneira gradativa. Já a intensidade de coocorrência entre as palavras é mostrado de acordo com o calibre das arestas, ou seja, fios mais grossos mostram intensidade alta.

Além disso, trazemos no âmbito da apresentação dos dados das nuvens de palavras os percentuais de frequência relativa dos termos em relação ao total de ocorrências no *corpus*. Tais porcentagens foram feitas de maneira artesanal com base na tabela de frequência absoluta exportada do software. Já quando apresentamos as árvores, nos referimos à métrica de centralidade de intermediação.

Com relação à exposição dos agrupamentos de palavras, iremos expandir com exemplos as que tiverem como primeiro vocábulo um substantivo ou adjetivo, em conformidade com nossa opção metodológica por buscar as que possuem sentido pleno. Exibimos os agrupamentos indo do que apresenta a palavra de maior centralidade de intermediação até o menor.

Também para fins de sistematização dos dados, serão eleitos alguns comentários a título de exemplos e que estão etiquetados de acordo com o episódio selecionado: (“**exemplo de comentários** selecionados”⁸ com a sigla C1 e assim por diante). Além disso, com o tratamento feito com os dados textuais, se perderam os sinais de pontuação e foi mantida a grafia apresentada em tais comentários. Neste processo de mineração e de tratamento dos dados, foi necessário, ainda, alterar manualmente termos que continham caracteres não reconhecidos pelo Iramuteq, como o hífen e foi preciso utilizar termos em caixa baixa, como nomes próprios (Brasil, Lava Jato, Lula, Moro, Glenn Greenwald, entre outros). Dessa forma, a escrita que compõe o corpus foi mantida, visto que é preciso saber como a linguagem dos softwares opera.

A partir da construção desses gráficos e das análises geradas, é possível fazer inferências no âmbito da pesquisa sobre as processualidades observadas no episódio comunicacional eleito aqui para análise. Para isso, é necessário que retornemos aos textos originalmente produzidos (pelo TIB e pelos leitores) para que possamos compreender os sentidos evocados pelas palavras em seus contextos específicos. A seguir detalhamos o percurso que envolve os dados coletados, a análise estatística, a geração da nuvem de palavras e a análise de similitude para o conjunto de comentários do episódio comunicacional.

O episódio comunicacional “Cara de pau”

Passamos a apresentar o episódio comunicacional “Cara de pau” que tem relação com o vídeo postado no canal do Youtube do TIB, intitulado “Sérgio Moro diz que não é juiz investigador”. O vídeo possui 58 segundos de duração e mostra uma fala pública do ex-juiz Sérgio Moro quando declara: “eu não tenho estratégia de investigação nenhuma”. A fala é contraposta com o texto que segue junto na descrição do vídeo, em que a agência de notícias declara: “nós mostramos que ele sugeriu a Deltan Dallagnol que trocasse a ordem de fases da Lava Jato”. Inspirados em Braga (2017), consideramos que a noção de episódio comunicacional abarca a junção do exposto pelo TIB em conjunto com os comentários encontrados.

8 In: <https://www.youtube.com/watch?v=3AutRXEAEKg> . Acessado em 21 mar 2021.

Na classe 3 { “bandido” (2006), “ministro” (375), “defender” (375) }, com 69 comentários, há xingamentos a Sérgio Moro (“esse país é uma piada o lixo do ministro da justiça é um **bandido** dos mais inescrupulosos” C 12), como também há noção de que mesmo diante do apresentado pela série, Lula ainda continua tendo culpa (“o teu deus lulinha foi condenado em todas as instâncias jurídicas _moro foi só o primeiro até os ministros do stf nomeados pelo próprio o condenaram pare de defender **bandido** rapaz” C 13). Como também surgem comentários com marcas de argumentação jurídica (“no processo penal da atual constituição o juiz não pode ter nenhum papel ativo no processo defendem um **bandido** contra o outro mas o primeiro ainda é **bandido** esses cidadãos de bem são ridículos” C 14).

Na classe 1 { “brasileiro”, (929), “boca” (188), “deixar” (0) }, com 45 comentários, há comentários que mencionam o jornalista Glenn Greenwald, um dos autores da série, de maneira desrespeitosa¹¹ (“esse greenwald tem que chupar uma [...] bem grande para ocupar a bôca com outra [...] e não aquela minhoca do marido dele nem esse besteriol sonstituiu família com o seu macho gay **brasileiro** para simplesmente fazer weakleaks no brasil” C 15), outros que pedem para que ele divulgue o material recebido como é costume do site Wikileaks (“o glenn_greenwald deveria fazer como o wikileaks colocar todos os audios na mídia para avaliação e perícia inclusive glenn_greenwald deveria ser justo e grampear os companheiros que roubaram tudo de todo o povo **brasileiro** desde 2003 inclusive de seus filhos seja justo glenn_greenwald” C 16). Além disso, há um pedido de informação jurídica especializada para poder entender melhor a situação (“alguém sabe me dizer qual é o sistema do código processo penal **brasileiro** e se pudesse explicar esse sistema em poucas palavras baseado em algum artigo obs não estou entendendo nada do que está sendo dito” C 17).

Na classe 8 { “cara” (927), “falar” (375), “peixe” (0) }, com 81 comentários, há xingamentos a Sérgio Moro (“**cara** de pau” C 18), como também a expressão “cara” é utilizada para interação com outros usuários (“é pra falar de investigação de partido já esqueceu que a pf estava na sede do partido do seu presidente e o psl está sendo investigado também risos_ha **cara** você acredita mesmo no que está escrevendo mas é só um fanático que se acha despertado” C 19).

Na classe 11 { “país” (930), “colocar” (0), “mundo” (0) }, com 55 comentários, emergem comentários que falam sobre como o país parece ser (“bem passiva aquela liberação de áudio pro jn de uma presidente em exercício com um ex presidente nomeado para o ministério após a própria determinação do fim da escuta e divulgado em tempo record em **país** sério isso teria consequências chocantes” C 20) e (“num **país** sério esse green sei la o que ja estava vendo o sol nascer quadrado isso sim” C21) ou estar (“por isso esse **país** está arruinado pessoas que defendem bandidos total inversão de valores o pt fez bem seu trabalho e destruiu a moral de uma parte da população” C 22).

Os parabéns não aparecem com destaque, quando olhamos para as métricas utilizadas em nossa investigação, mas aparecem no *corpus* em 43 comentários. Há congratulações ao trabalho do TIB (“a máscara desse farsante caiu **parabéns** pelo trabalho glenn_greenwald e equipe the _tib” C 23) e parabenizações ao ex-juiz Sérgio Moro (“**parabéns** _moro colocou na cadeia políticos e empresários corruptos e poderosos num país que só prendia pobres” C 24).

Assim, após apresentação das classes emergentes, ao retomar a discussão sobre a processualidade do episódio interacional podemos perceber como já apresentado no episódio anterior, que há compartilhamento de códigos comuns entre o TIB e os comentadores, assim como há um processo inferencial das trocas interacionais (BRAGA, 2017). É possível notar que a postagem do vídeo no canal do Youtube tem como objetivo apresentar mais uma evidência a série de reportagens, pois ele aparece no corpo da terceira reportagem publicada no site.

Assim, retomando a noção dos fluxos comunicacionais, um de orientação TIB-leitores e outro leitores-leitores, é possível notar que no primeiro fluxo, há um destaque para a confluência de posições com relação às críticas feitas ao ex-juiz Sérgio Moro, único personagem presente no vídeo, que é acusado de traidor da pátria e xingado em muitos dos comentários.

Com relação ao outro eixo, notamos que há marcas de negociação de sentidos com relação ao que foi exposto pela agência de notícias (“se o **_moro** estivesse conversando com o lula vocês viriam cheios de mimimi e você ganha o que pra ser gado do mito” C 25). Como também surgem marcas do estabelecimento de um conflito aberto, quando se contrapõem as noções de quem são os bandidos (“o brasil cambada de vagabundo que defende bolsonaro e _moro eles roubaram milhares de reais e vocês fica lambendo o cu destes **bandidos**” C 26) (“risos_ha olha só quem fala o cara que quer o **bandido** solto e o juiz preso você é patético cara” C 27).

Além disso, notamos que as trocas com marcas de argumentação jurídica foram as que levaram a uma busca por negociar os sentidos em fluxos adiante, como por exemplo os comentários que emergiram na classe 7. É possível também encontrar marcas que ressaltam o espaço dos comentários serem considerados uma ambiência. (“ia passando e resolvi dá uma paradinha e deixar recadinho o **bandido** de nove dedos está em cana lá em curitiba” C 28).

11 Optamos por omitir expressões obscenas.

Assim, levando em consideração tais aspectos é possível perceber que a lógica da plataforma pode ser considerada importante para compreensão sobre a processualidade do episódio interacional, como também o contexto político.

Inferimos que esse episódio é marcado por distintos conflitos, tanto na disputa por sentidos entre o que as reportagens evidenciam, quanto entre o que os leitores dizem nos espaços de comentários. Tais constatações apontam para dinâmicas complexas da circulação, em que os participantes produzem múltiplos sentidos por meio de processos difusos e em fluxos adiante (BRAGA, 2017).

Depois do detalhamento do percurso teórico e metodológico, no qual propusemos a articulação entre as métricas geradas por meio da utilização do software Iramuteq e dos gráficos construídos com auxílio do Gephi e o tensionamento com os dados textuais, passamos a elencar algumas notas a título de fechamento de nossa proposta.

Considerações finais

Compreendemos que a abordagem inferencial integrada ao uso de gráficos e métricas contribui para aprofundar a reflexão em torno das processualidades que incidem sobre a circulação de sentidos. Assim, levando em conta que a metodologia a ser empregada depende de diversos fatores, como a problemática envolvida, a jornada do pesquisador e a sua filiação teórica, consideramos que os encaminhamentos adotados em nossa investigação podem contribuir para outras pesquisas sobre a circulação de sentidos. Não é raro vermos pesquisas em Comunicação que se limitam em apresentar os dados quantitativos, sem os problematizá-los. Nossa experimentação aqui detalhada foi justamente fazer um esforço teórico e metodológico para que os dados não ficassem soltos, mas que pudessem nos dar pistas para um olhar mais atento.

Com relação ao encadeamento de nossa investigação, consideramos que partir de um olhar baseado no uso gráfico e numérico das porcentagens de frequência relativa das nuvens de palavras, para as árvores máximas, com as centralidades de intermediação (NEWMAN e GIRVAN, 2004) e a modularidade, pode ser mais uma abordagem metodológica a ser empreendida em investigações em circulação. Em nossa pesquisa, consideramos que as nuvens de palavras contribuem para se ter uma noção visual sobre quais seriam os principais assuntos ou temas presentes nos comentários. Contudo, a integração com a visualização das árvores máximas e suas métricas se revelou importante para podermos efetuar um aprofundamento sobre os sentidos presentes no episódio comunicacional. Aqui é importante destacar a transição efetuada dos gráficos do Iramuteq para o Gephi, pois é no último que se pode aumentar ou diminuir a proporção tanto das árvores quanto dos nós, além de efetuar o cálculo das métricas.

Além das inferências sobre a forma visual, a exploração numérica dos gráficos pode nortear a direção analítica, que em nosso caso, se preferiu expandir o olhar sobre os agrupamentos encontrados através da palavra com maior centralidade de intermediação (NEWMAN e GIRVAN, 2004). Contudo, é possível, se for de interesse de outros pesquisadores, efetuar outros percursos, a depender da questão norteadora da pesquisa e interesses de investigação. Optamos por dar destaque para aquilo que emergiu em excesso, ou seja, buscamos olhar para as recorrências com maior intensidade.

Por fim, nossa pesquisa se soma a outras da área que tateiam o universo complexo da circulação e suas interpenetrações (FAUSTO NETO, 2018) e fluxos contínuos e adiante (BRAGA, 2017). Ainda, de forma tentativa, buscamos olhar para o episódio comunicacional instigados pelas investigações sobre a circulação de sentidos por uma abordagem crítica (GROHMANN, 2019). Intencionamos também contribuir para discussões iniciadas por Borelli e Dias (2018) na visada por indicar possíveis avanços e limitações metodológicas em pesquisas em circulação.

Nossa experimentação metodológica que busca integrar diferentes táticas analíticas exige cuidados na elaboração da argumentação, pois há indícios que se reforçam mutuamente, assim como outros que parecem indicar compreensões divergentes. No episódio comunicacional apresentado, a discussão emerge quando olhamos para a centralidade da menção na árvore máxima com dois galhos antagônicos vinculados a Sérgio Moro que ora se mescla a noção de bandido, traidor da pátria, ora se mescla a imagem de juiz exemplar.

Com relação aos desafios enfrentados no desenrolar da pesquisa, é importante destacar que no processo de mineração dos dados se perde algumas sutilezas dos comentários, como o uso de emojis e outras imagens. Outro ponto desafiador é com relação à limpeza dos dados textuais, que foi efetuada de maneira artesanal. Sabemos que é possível elaborar algoritmos para que se automatize essa questão, assim é importante a aproximação com pesquisadores de outras áreas do saber para que a partir de parcerias isso possa se tornar viável.

Buscamos integrar noções quantitativas com as qualitativas. Ou seja, detalhar o contexto para os dados textuais por uma perspectiva inferencial foi a forma encontrada nesta pesquisa para dar conta desta problemática. Tal encaminhamento se revelou promissor, principalmente, quando se olha para as menções ao personagem do vídeo, mostrando tanto o descontentamento, quanto a idolatria, revelando uma riqueza de detalhes e nuances, que integrada

a análise estatística dá maior robustez a pesquisa.

Referências

- BORELLI, Viviane V. (2023 no prelo). A problematização teórica-metodológica nas pesquisas de circulação e plataforma. In: J. Ferreira, Jairo,; A. M. C. M. Silveira, V. Borelli ; Dalmolin, Aline ; Rosa, Ana Paula. Löfgren, Isabel (Orgs.) . **IA, algoritmos e plataformas: Questões e hipóteses na perspectiva da midiatização**. 1. ed. Santa Maria (RS): FACOS-UFSM. v. 1.
- BORELLI, V; DIAS, M. S. M. Circulação discursiva: desafios metodológicos para compreender as interações entre jornais e leitores. **Rizoma**, [S.L.], v. 6, n. 2, p. 95-109, 7 jul. 2018. APESC - Associação Pro-Ensino em Santa Cruz do Sul. <http://dx.doi.org/10.17058/rzm.v6i2.12602>.
- BORELLI, VIVIANE; KROTH, M. E. . Circulação e construção de circuitos: a dinâmica do rádio. **E-COMPÓS** (BRASÍLIA), v. 23, p. 1-25, 2020.
- BRAGA, José Luiz. Dispositivos Interacionais; Circuitos de Comunicação. In: BRAGA, José Luiz; CALAZANS, Regina; RABELO, Leon et al. **Matrizes interacionais: a comunicação constrói a sociedade**. Campina Grande: EDUEPB, 2017. P.71-64.
- BRAGA, José Luiz. Comunicação, disciplina indiciária. **Matrizes**, v. 1, n. 2, p. 73-88, abr. 2008. Disponível em: <www.matrizes.usp.br/ojs/index.php/matrizes/article/download/46/28>. Acesso em: 6 mar. 2016.
- BRANDES, U. A faster algorithm for betweenness centrality. **The Journal Of Mathematical Sociology**, [S.L.], v. 25, n. 2, p. 163-177, jun. 2001. DOI: <https://doi.org/10.1080/0022250X.2001.9990249>
- BREITER, A.; HEPP, A. The Complexity of Datafication: putting digital traces in context. In: HEPP, A.; BREITER, A.; HASEBRINK, U. (org.). **Communicative Figurations**. Londres: Palgrave, 2018. p. 387-405.
- CAMARGO, B. V.; JUSTO, A. M.. IRAMUTEQ: um software gratuito para análise de dados textuais. **Temas em Psicologia**, Ribeirão Preto, v. 21, n. 2, p. 513-518, dez. 2013. Disponível em <http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X2013000200016&lng=pt&nrm=iso>. Acesso em: 11 nov 2020.
- DEGENNE, A. e VERGÈS, P. Introduction à l'analyse de similitude. **Revue Française de Sociologie**, v. 14, n. 4, p. 471, Out 1973. Disponível em: <<https://www.jstor.org/stable/3320247?origin=crossref>>. Acesso em: 24 maio 2020.
- FAUSTO NETO, A. Circulação: trajetos conceituais. **Rizoma**, [s.l.], v. 6, n. 2, p.08-40, 7 jul. 2018. APESC - Associação Pró-Ensino em Santa Cruz do Sul. <http://dx.doi.org/10.17058/rzm.v6i2.13004>.
- FRIGO, D.; ROMERO, L. M.; BORELLI, V. Plataformas, dispositivos interacionais e circulação: mapeamento do episódio “Vaza Jato”. **Anais de VIII Colóquio Semiótica das Mídias**, p. 17, 2019.
- GINDIN, Irene Lis; BUSSO, Mariana Patricia. Investigaciones en comunicación en tiempos de big data: sobre metodologías y temporalidades en el abordage de redes sociais. In: adComunica. **Revista de Estrategias, Tendencias e Innovación en Comunicación**, n. 15. Castellón: Asociación para el Desarrollo de la Comunicación adComunica y Universitat Jaume I, 2018, p. 25 - 43.
- GROHMANN, Rafael. Os rastros digitais na circulação de sentidos: pela desnaturalização e contextualização de dados na pesquisa em comunicação. **Galáxia (São Paulo)**, São Paulo , n. 42, p. 150-163, Dec. 2019. Disponível em <<http://dx.doi.org/10.1590/1982-25532019340183>>. Acesso em 20 Jan. 2021.
- GINZBURG, Carlo. **Mitos, emblemas, sinais**. São Paulo: Companhia das Letras, 1989.
- MARCHAND, P.; P. RATINAUD. (2012). L'analyse de similitude appliquee aux corpus textuels: les primaires socialistes pour l'election présidentielle française. Em: **Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles**. JADT 2012. (687–699). Apresentado em 11eme Journées internationales d'Analyse statistique des Données Textuelles. JADT 2012., Liège, Belgique
- NEWMAN, M. E. J.; GIRVAN, M.. Finding and evaluating community structure in networks. **Physical Review e**, [S.L.], v. 69, n. 2, p. 69-74, 26 fev. 2004. Mensal. American Physical Society (APS). DOI: <http://dx.doi.org/10.1103/physreve.69.026113>.
- ROMERO, L. M. **Série Vaza Jato: Análise do circuito comunicacional formado em torno das três primeiras reportagens**. 2021. 102 f. Dissertação (Mestrado) - Programa de Pós-graduação em Comunicação, Universidade Federal de Santa Maria, Santa Maria, 2021.
- WOBETO, Samara Letícia; ROMERO; Luan Moraes e BORELLI, Viviane. A utilização de softwares para análise e visualização de dados nas pesquisas em Comunicação. **Anais da ABCIBER XVI - SIMPÓSIO NACIONAL DA ABCIBER 2023**. ISSN : 2175-2389. In: <https://abciber.org.br/simposios/index.php/abciber/abciber16/paper/view/2071>. Consulta em 3 abr 24.

VAN DIJCK, J; POELL, T.; DE WAAL, M. **The Platform Society**. Public Values in a Connective World. Nova York: Oxford University Press 2018.

VERÓN, E. **Fragmentos de um tecido**. São Leopoldo: Unisinos, 2004.

VERÓN, E. **La semiosis social**, 2. Ideas, momentos, interpretantes. Buenos Aires, Paidós Planeta, 2013.

Sobre os autores

Luan Moares Romero

luan_155@hotmail.com

Jornalista pela Universidade Federal de Santa Maria (UFSM) Mestre e doutorando pelo Programa de Pós-Graduação em Comunicação da Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil.

Viviane Borelli

viviane.borelli@ufsm.br

Docente do Programa de Pós-Graduação em Comunicação e do Departamento de Comunicação da Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil. Líder do grupo de pesquisa “Circulação midiática e estratégias comunicacionais” (Cimid). Doutora em Ciências da Comunicação pela Unisinos.

Recebido em: 26/11/2023
Aprovado em: 24/08/2024

Editoras Responsáveis:
Marialva Barbosa e Sonia Virgínia Moreira



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC- BY). Os autores retêm todos os direitos autorais, transferindo para a Intercom: Revista Brasileira de Ciências da Comunicação o direito de realizar a publicação original e mantê-la sempre atualizada.