

ELEIÇÕES, *FAKE NEWS* E OS TRIBUNAIS

METODOLOGIA

ELEIÇÕES, *FAKE NEWS* E OS TRIBUNAIS

DESINFORMAÇÃO ONLINE NAS ELEIÇÕES BRASILEIRAS

RELATÓRIO DE METODOLOGIA DE PESQUISA 2021

Coordenação acadêmica:

Alexandre Pacheco da Silva
Marina Feferbaum
Victor Nóbrega Luccas

Coordenação executiva:

Rodrigo Moura Karolczak
Carlos Augusto Liguori Filho
João Pedro Favaretto Salvador

Pesquisadores principais:

Luiz Fernando Antonelli Galati
Jeniffer Martins da Silva
Guilherme Kenzo dos Santos
Theófilo Miguel de Aquino

Pesquisadores estagiários:

Ana Luiza Vidotti
Camila Merino Moya Leiva
Giovanna Brügger

Pesquisador colaborador:

Ezequiel Fajreldines dos Santos

Projeto Gráfico:

Tatiane Guimarães

Contato: cepi.direitosp@fgv.br



* Qualquer parte desta publicação pode ser reproduzida, desde que citada a fonte.

.....
KAROLCZAK, R. M. et al. **Eleições, Fake News e os Tribunais**: relatório de metodologia. São Paulo: CEPI FGV Direito SP, 2021. Disponível em: <https://fgv.academia.edu/fgvcepi>.
.....

SUMÁRIO

INTRODUÇÃO	7
OBJETIVOS	11
REVISÃO BIBLIOGRÁFICA	13
DELIMITAÇÃO DO UNIVERSO DE PESQUISA	17
Esforço preliminar para a construção do universo	18
Levantamentos no Processo Judicial Eletrônico	20
Coleta de decisões judiciais	22
FILTRAGEM E FORMAÇÃO DA AMOSTRA DE PESQUISA	27
Rodadas de filtragem automatizada	32
Filtragem manual e análise de relevância	33
ESTRUTURAÇÃO DOS DADOS	37
A Planilha de Partes	37
A Planilha de Processos	45
As Planilhas de URLs e de Precedentes	47
REFERÊNCIAS	49
ANEXOS	53
ANEXO I - Dispositivos em Processos	54

Lista de Figuras

Figura 1: Exemplo de documento disponível no portal do TSE	19
Figura 2: Exemplo de documento disponível no PJe	21
Figura 3: Excertos da tabela de filtragem da amostra	34

Lista de Tabelas

Tabela 1: Conjuntos de Termos de Busca para Seleção de Processos Judiciais	30
Tabela 2: Variáveis Descritivas do Repositório de dados eleitorais do TSE (Aplicáveis quando a parte é Candidato)	38
Tabela 3: Variáveis Descritivas das Partes e Outras	40
Tabela 4. Categorização das partes	42
Tabela 5: Codificação de variáveis de resultados de processo	477

INTRODUÇÃO

O projeto “Eleições, Fake News e os Tribunais: desinformação online nas eleições brasileiras” teve como principal objetivo a construção e análise descritiva de um banco de dados de decisões da justiça eleitoral que tratam de *fake news* e desinformação online nas eleições gerais de 2018 e nas eleições municipais de 2020¹. A pesquisa, portanto, foi realizada em duas fases. A primeira fase foi realizada entre 2018 e 2020, e consistiu na coleta e análise de processos referentes às eleições gerais de 2018. A segunda fase foi realizada entre 2020 e 2021, e consistiu na coleta e análise de processos referentes às eleições municipais de 2020. Como ambas as fases seguiram metodologias semelhantes, condensamos os passos seguidos em um único relatório, apontando, sempre que necessário, as distinções e adaptações que foram necessárias em razão do aperfeiçoamento do método.

Assim, este relatório expõe as decisões metodológicas tomadas para alcançar esses objetivos, descrevendo o passo a passo da amostragem e da codificação de variáveis, ilustrando essas atividades com parte dos achados descritivos. Ao final da leitura, esperamos que quem leia o relatório seja capaz de replicar e utilizar nosso trabalho e nossas ferramentas, verificando a validade dos resultados e produzindo, a partir deles, mais dados e análises essenciais para a compreensão do processo eleitoral brasileiro.

A pesquisa teve como motivação o reconhecimento de que processos eleitorais e referendos ao redor do mundo têm sido palco de desinformação e notícias fraudulentas online, fenômeno popularmente conhecido como *fake news*, o que demanda atenção especial. Os processos de desinformação online contemporâneos puderam ser observados, entre outros casos, nas eleições gerais do Japão em 2014 (Schäfer, Evert, e Heinrich, 2017); nas eleições presidenciais dos Estados Unidos da América (Allcott, Gentzkow, 2017) e no referendo sobre o *Brexit* no Reino Unido (Bastos, Mercea, 2019), ambos em 2016. Em 2017, as eleições presidenciais da França também passaram por casos de *fake news* (Ferrara, 2017). Mais recentemente, as eleições gerais na Índia em 2019 foram marcadas por um elevado volume de desinformação online (Ponniah, 2019), ainda que não tenhamos encontrado estudos acadêmicos sobre o tema. No Brasil,

¹ Os bancos de dados de 2018 e 2020 podem ser baixados, respectivamente, nos links <https://hdl.handle.net/10438/29935> e <https://hdl.handle.net/10438/31157>.

diversos indicadores sugerem que o assunto está sendo percebido cada vez mais pelo público. O *Digital News Report 2018*, da Reuters Institute (Newman et al. 2018) revelou que 85% dos brasileiros dentro de sua amostra² dizem que estão "muito ou extremamente preocupados a respeito do que é real e o que é falso quanto a notícias na internet". É a maior proporção entre os 37 países estudados no relatório.

Ademais, há uma significativa diferença entre o consumo de mídias sociais³ e a confiança depositada nestes meios de comunicação no Brasil. Brasileiros estão entre os mais ávidos usuários de mídias sociais no mundo. Baseado nos resultados do *Digital News Report 2018*⁴, ao menos 52% dos usuários acessam a rede social Facebook com frequência, enquanto 48% frequentemente utilizam do aplicativo de mensagens WhatsApp. Ao menos 66% utilizam mídias sociais como fontes de notícias. No entanto, apenas 32% confiam em notícias compartilhadas em mídias sociais, enquanto canais de notícias familiares são confiados por 58% da população amostrada.

No entanto, um passo anterior à mensuração e descrição do fenômeno da desinformação é a sua correta conceituação. Assim, é fundamental observar os termos correntes de forma pormenorizada. Ainda que de maior popularidade por seu uso político recorrente, a expressão *fake news* é imprecisa para descrever o fenômeno que despontou nas eleições mencionadas, o que impeliu estudiosos do tema a buscarem outros conceitos que delimitam de forma mais clara o mesmo fenômeno. Conforme aponta Rais (2018, p. 106-107), a tradução literal do termo — notícias falsas — não auxilia na decisão do que deve ou não ser regulado, na medida em que a falsidade do conteúdo por si só não interessa ao Direito. Para o autor, os elementos relevantes ao Direito são, na verdade, o dano ou o potencial danoso desse tipo de disseminação de informação. Nesse sentido, variantes do termo “notícias fraudulentas” são preferíveis à tradução literal do termo.

Em crítica de sentido semelhante, o *High Level Expert Group on Fake News and Online Disinformation* (HLEG) da União Europeia (em português, Grupo Independente de Alto Nível sobre Notícias Falsas e Desinformação Online) recomendou o abandono do termo *fake news* por seu uso enganador em disputas políticas e por sua limitada

² De acordo com o relatório, foram entrevistadas 2007 pessoas, sendo os resultados mais representativos da população urbana do que rural. (Newman et al., 2018, p. 6)

³ Utilizamos mídias sociais como sinônimo de redes sociais.

⁴ Idem, pp. 116-117.

abrangência⁵. Para o grupo, a noção de desinformação é mais interessante por abstrair mais significado do que a mera ideia de notícia falsa, incluindo "informações falsas, inexatas ou deturpadas concebidas, apresentadas e promovidas para obter lucro ou para causar um prejuízo público intencional,"⁶ não exclusivamente no âmbito eleitoral.

No Judiciário brasileiro, em resposta ao debate crescente sobre o papel das *fake news* nas eleições, o Tribunal Superior Eleitoral (TSE) editou a Resolução nº 23.551/17, que não contém expressamente o termo "*fake news*," mas que cria mecanismos para a remoção de conteúdo online "sabidamente inverídico" em manifestações espontâneas na Internet no contexto de campanhas eleitorais⁷. Estes mecanismos possuem escopo amplo o suficiente para abranger conteúdo compartilhado por usuários comuns, sem qualquer tipo de ligação oficial a candidatos e partidos — algo com precedentes na legislação eleitoral brasileira apenas no contexto da divulgação de fatos sabidamente inverídicos na propaganda eleitoral⁸.

Para que se entendam os impactos regulatórios do TSE nesta resolução, é preciso ter em mente duas classes de conteúdo online no contexto do Direito Eleitoral elaboradas por Rais et al. (2018, p. 47-48): conteúdo orgânico e a propaganda eleitoral. A primeira é a livre manifestação de usuários na Internet, caracterizada pelo "exercício do direito de livre manifestação do pensamento". A segunda é o tipo de conteúdo propagandístico "gerado ou editado por (i) candidatos, partidos políticos ou coligações; ou (ii) qualquer pessoa natural" tradicionalmente regulamentado pelas normas eleitorais.

A resolução, por meio do art. 22, §1º, inova ao permitir que um conteúdo orgânico online possa ser removido caso haja divulgação de fatos sabidamente inverídicos. Ela reconhece, no entanto, que a manifestação espontânea de indivíduos na esfera online

⁵ High Level Group on Fake News and Online Disinformation (2018) A multi-dimensional approach to disinformation. Luxembourg: Publications Office of the European Union. Disponível em: <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271>. Acesso em: 29 de Maio de 2019.

⁶ Idem, p. 11. Do original em inglês: "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit".

⁷ Art. 22, Resolução nº 23.551/17: "É permitida a propaganda eleitoral na internet a partir do dia 16 de agosto do ano da eleição (Lei nº 9.504/1997, art. 57-A). § 1º A livre manifestação do pensamento do eleitor identificado ou identificável na internet somente é passível de limitação quando ocorrer ofensa à honra de terceiros ou divulgação de fatos sabidamente inverídicos. § 2º O disposto no § 1º se aplica, inclusive, às manifestações ocorridas antes da data prevista no caput, ainda que delas conste mensagem de apoio ou crítica a partido político ou a candidato, próprias do debate político e democrático".

⁸ Art. 323, Código Eleitoral: "Divulgar, na propaganda, fatos que sabe inverídicos, em relação a partidos ou candidatos e capazes de exercerem influência perante o eleitorado: Pena - detenção de dois meses a um ano, ou pagamento de 120 a 150 dias-multa" (grifos nossos).

seria uma classe de conteúdo expressamente distinta da propaganda eleitoral de acordo com o art. 23, §6º. Normativamente, haveria consequências diferentes para cada classe de conteúdo e caberia à Justiça Eleitoral “avaliar o conteúdo de forma mais profunda” para determinar a aplicabilidade de regras eleitorais específicas (Rais et. al, 2018, p. 48).

Estudos demonstram que as eleições de 2018 foram marcadas pela ação de robôs sociais — difusores automatizados de mensagens e interações — e por casos de desinformação com alcance considerável (Ruedinger, 2018). Outros estudiosos apontam para o surgimento de uma mídia “hiper-partidária” no Brasil que floresce em um ambiente político polarizado e de diálogo pobre entre os lados opostos do espectro político. Esta mídia seria responsável pela produção de conteúdo enviesado e opinativo que, apesar de eventualmente propagar informações falsas, seria melhor caracterizada por matérias de “informação de combate,” isto é, grosso modo:

[U]m recorte conveniente do noticiário do dia, uma notícia com uma manchete sensacionalista, um fato retirado do seu contexto, um exagero ou uma especulação apresentada como fato — ocasionalmente, pode até mesmo ser uma mentira (Ribeiro, Ortellado, 2018, p. 73).

Ambas as perspectivas demonstram a relevância e o emergente interesse em pesquisas sobre desinformação online no Brasil, ainda que a produção acadêmica no tema seja incipiente. Enquanto a literatura brasileira contribui para a identificação do problema e seu atual tratamento jurídico, a revisão bibliográfica abaixo explora a produção acadêmica internacional, inserindo a pesquisa em um debate mais amplo na literatura.

Nas seções seguintes trataremos de: (i) descrição dos objetivos de pesquisa; (ii) revisão bibliográfica; (iii) processos de coleta de dados; (iv) processos de refinamento e filtragem de amostra; e (v) descrição de dados.

1. OBJETIVOS

Com esse cenário em mente, percebemos ser necessária a compreensão de como os tribunais eleitorais brasileiros lidaram com as problemáticas acima descritas no decorrer das eleições. Para atingir este fim, esta pesquisa objetivou compilar e analisar decisões judiciais da Justiça Eleitoral de todos os estados e do Tribunal Superior Eleitoral que tratam de questões associadas às *fake news*, à desinformação e à produção e divulgação de fatos sabidamente inverídicos na Internet.

Como foi descrito acima, a expressão *fake news* é a mais popular e, ao mesmo tempo, a de significado menos claro no debate contemporâneo. Tanto a dificuldade de definir o termo quanto seu uso político recorrente justificam um olhar mais profundo para sua apropriação pelos tribunais eleitorais brasileiros.

A noção de desinformação, por outro lado, abarca o espectro de condutas que envolvem a produção e disseminação intencional de informações falsas, imprecisas ou enganadoras. Trata-se de um conceito mais amplo, mas, ao mesmo tempo, melhor definido do que aquele normalmente abrangido pela literalidade do termo *fake news*, que se refere especificamente a notícias falsas.

Por fim, a expressão “fatos sabidamente inverídicos” se tornou comum no debate jurídico e vocabulário do Judiciário brasileiro por sua adoção pelo Tribunal Superior Eleitoral (TSE), na Resolução nº 23.551/17, como forma de lidar com as recentes demandas de atuação sobre os processos de desinformação online. Esses três referenciais de linguagem serão fundamentais para definir nosso universo de pesquisa, classificar decisões judiciais e compará-las.

Nesse sentido, os objetivos deste projeto consistem em duas vertentes: de um lado, a formação de um banco de dados organizado e estruturado que contenha decisões judiciais relevantes de forma acessível, além de dados descritivos a respeito de seus relativos processos — como partes, classes judiciais e movimentações. Por outro, buscamos contribuir para a produção acadêmica internacional sobre *fake news* e desinformação online por meio do uso de dados inovadores no caso brasileiro. Conforme Tucker et al. (2018, p. 54), importantes lacunas na literatura internacional sobre *fake news* são justamente: (i) a avaliação do impacto de novas leis para coibir a distribuição de desinformação online, e (ii) a observação de como processos de desinformação ocorrem em países diferentes. Este esforço de pesquisa busca contribuir nestas duas frentes.

Para isso, desenvolvemos ferramentas digitais automatizadas de coleta, descrição e filtragem de decisões a partir de bancos de dados públicos da Justiça Eleitoral. Também revisamos a amostra manualmente para garantir que seja pertinente e contenha apenas decisões relevantes.

Para orientar a construção da mencionada base de dados de processos relevantes para a temática, levantamos algumas questões fundadas em revisão bibliográfica, cujas respostas pressupõem a coleta e análise de documentos processuais organizados.

De um ponto de vista qualitativo, a base de dados procura fornecer documentos e dados descritivos que permitam a avaliação de duas perguntas de pesquisa: (i) quais são os significados de *fake news* utilizados nos tribunais para fundamentar decisões? E (ii) quais os principais fundamentos jurídicos que situam o debate sobre desinformação e informações inverídicas no Brasil? Já pela perspectiva quantitativa, os dados coletados devem viabilizar pesquisas que busquem responder ao menos duas perguntas: quais são as covariantes significativas que explicam (iii) o volume de casos e (iv) o resultado dos processos judiciais que tratam de conteúdo online com base em alegações de *fake news*, desinformação e fatos sabidamente inverídicos? Tais questionamentos orientaram os critérios de filtragem e seleção de documentos e codificação de variáveis que serão descritos nas próximas seções.

2. REVISÃO BIBLIOGRÁFICA

Na literatura acadêmica internacional sobre *fake news*, há um corpo de produção significativo que se utiliza de dados de análise de rede no Twitter para observar como redes de desinformação se comportam e quais atores contribuem para sua distribuição (Bovet, Makse, 2019; Brummette et al., 2018; Jang et al., 2018; Shao et al., 2017; Shao et al., 2018). Metodologicamente, estes estudos levantam perguntas que caracterizam as redes e buscam elementos de causalidade entre as interações de agentes. Enquanto alguns exploram como diferentes grupos utilizam-se do termo *fake news* (Brummette et al., 2018), outros comparam a difusão de conteúdo produzido por *fact-checkers*⁹ com a difusão de notícias falsas (Shao et al., 2018). A utilização do Twitter como rede social de escolha se dá por sua API¹⁰ menos restritiva em comparação ao Facebook, o que permite coleta automatizada de tweets dentro de alguns parâmetros de razoabilidade definidos pela própria plataforma.

Outra parte da literatura buscou compreender *fake news* e desinformação ao compilar sites notórios por divulgar notícias fraudulentas e observar interações nas redes sociais que os mencionam, como páginas no Facebook e contas no Twitter (Allcott, Gentzkow, 2017; Allcott et al., 2018; Guo, Vargo, 2018). Essa perspectiva metodológica permite comparações de acesso com sites de mídias tradicionais, compartilhamento de agenda de sites de notícias falsas com sites de notícias hiper-partidários, e proporções de acesso entre diferentes plataformas virtuais. No entanto, estes estudos dependem de instituições que compilam sites de notícias falsas para formar seus bancos de dados, o que pode ser mais difícil para estudos fora dos Estados Unidos da América e Europa.

No conjunto, esses estudos demonstram que o objeto das *fake news* e da desinformação é complexo e difuso nas redes sociais, tornando sua observação

⁹ *Fact-checkers* são agências de checagem de notícias especializadas em investigar a veracidade de fatos narrados e providenciar informações corretas quando necessário. Alguns exemplos de agências brasileiras incluem: Agência Lupa, da Revista Piauí <<https://piaui.folha.uol.com.br/lupa/>>, Aos Fatos, uma organização independente multidisciplinar <<https://aosfatos.org/aos-fatos-e-noticia/>>, Pública, uma agência de investigação sem fins lucrativos <<https://apublica.org/checagem/>>, e Boatos, uma organização de checagem de boatos diversos <<https://www.boatos.org>>.

¹⁰ *Application Programing Interface* é um software que permite a interface de uma aplicação com softwares de terceiros. No contexto de pesquisas de análise de rede no Twitter, a API da plataforma virtual permite que desenvolvedores utilizem seus próprios softwares para a coleta sistemática de dados dentro dos parâmetros e limites determinados pela rede social.

sistemática difícil. Por um lado, as análises de rede no Twitter conseguem capturar a emergência de conteúdos de desinformação e o comportamento de redes de usuários difundindo-os e interagindo entre si. No entanto, a reprodução desta perspectiva sofre da parcialidade da utilização do Twitter por brasileiros que, de acordo com o *Digital News Report 2018* (Newman et al. 2018), apresentam adesão de apenas 26% da população amostrada e utilização para notícias numa proporção de 14%. Quando comparado ao Facebook e WhatsApp, que possuem adesão de 75% e 83% e utilização para notícias de 52% e 48%, respectivamente, é importante ponderar se amostras no Twitter são representativas do universo total de desinformação online no Brasil.

Por outro lado, metodologias que identificam e compilam fontes de desinformação para então observar sua difusão em mídias sociais apresentam dados mais bem distribuídos entre diferentes plataformas virtuais, mas não necessariamente acompanham a velocidade de emergência de novas fontes de desinformação. Além disso, também enfrentam constante inutilização e abandono de websites antes relevantes para disseminar notícias fraudulentas. Nesse sentido, estudos que se utilizam dessa metodologia encontram um objeto rico, porém de difícil acompanhamento. Essa abordagem se torna mais viável de reprodução na medida em que mais instituições brasileiras acompanham e compilam informações sobre sites que propagam desinformação.

Assim, uma contribuição importante para essa literatura é a observação do fenômeno no Brasil sob um ponto de partida distinto para a coleta de dados. A utilização de dados de tribunais permite, ainda que de uma maneira indireta, flexibilidade de observação entre redes sociais e mídias jornalísticas digitais como meios de difusão de *fake news*, além de apresentar dados novos que complementam os estudos já realizados. É evidente que o recorte proposto não permite comparações diretas com análises de rede ou com a observação de acesso à sites, mas é possível observar o reflexo dessas atividades nos tribunais de uma maneira representativa. Tratando-se de desinformação na esfera política e eleitoral, é esperado que partidos e candidatos estejam particularmente interessados em defender suas imagens e levar aos tribunais casos relacionados ao tema. Por outro lado, este recorte não permite inferir sobre desinformação online para além da realidade eleitoral. Em resumo, apresentamos a oportunidade de observar a judicialização de casos de desinformação online na esfera eleitoral.

Além disso, estudos emergentes sobre os significados e usos da expressão *fake news* apresentam questões relevantes para a avaliação qualitativa de decisões judiciais sobre o tema (Farkas, Schou, 2018; Hirst, 2017; Jensen, 2018; Levinger, 2018). Farkas e Schou (2018, p. 303-307) resumem a questão ao contextualizar os múltiplos significados de *fake news* como uma disputa política. De acordo com os autores, há ao menos três interpretações sobre *fake news* no contexto norte-americano: (i) desinformação em meio digital apontada por acadêmicos, jornalistas e comentaristas que criticam a estrutura econômica da Internet — tornando rentável a utilização de anúncios em sites de volume intenso de acesso — como principal causa do fenômeno; (ii) *fake news* como uma maneira de deslegitimar e criticar opositores políticos, seja por sugerir que o lado oposto do espectro político é mais suscetível a conteúdos de desinformação, seja por acusar opositores de serem fontes de casos de desinformação; e, por fim, (iii) *fake news* como crítica à mídia tradicional na tentativa de deslegitimar seu trabalho jornalístico. Ainda que paralelos não sejam exatos para o uso da expressão no Brasil, é interessante observar se partidos e candidatos num contexto eleitoral se apropriam do termo *fake news* para atacar opositores ou mídias críticas.

Também é importante observar como os tribunais articulam noções de *fake news*, isto é, se o debate nas cortes é sobre um levantamento fiel dos dados e da verdade, ou sobre as garantias e os limites do trabalho jornalístico crítico, ou ainda sobre o uso político do termo para atacar opositores.

Outra literatura periféricamente interessante para a pesquisa se refere a *judicial politics* na América Latina (Iaryczower et al., 2002; Kapiszewski, Taylor, 2008; Pérez-Liñán et al., 2006; Taylor, 2006) devido a uma perspectiva mais ampla que trata, entre outras questões, de fenômenos e crises políticas que impactam o sistema judicial e criam novas demandas nos tribunais. Kapiszewski e Taylor (2008, p. 741-742) resumem alguns temas da literatura e, pertinente para esta pesquisa, apontam que a produção acadêmica contemporânea na área se debruça sobre as perspectivas de tribunais e juízes sobre assuntos políticos. Questionam, também, o papel dos tribunais e os impactos do exercício de suas atividades sobre os resultados de políticas públicas.

Tendo em vista os impactos eleitorais relevantes, a criação de novas normas para lidar com a questão da desinformação online, e a natureza política do objeto, argumentamos que o design de pesquisa proposto aqui se enquadra nesta literatura de *judicial politics*. Também é interessante notar que este conjunto de obras contribui para o refinamento metodológico de análises quantitativas de decisões judiciais por meio da

utilização de modelos estatísticos e variáveis de controle pertinentes, assim como outros estudos de metodologia semelhante que tratam de temas fora do escopo proposto (Carruba et al., 2008; Deakin, 2018; Gabel et al., 2012; Hall, Windett, 2013; Smith, Todd, 2015). Outro aspecto importante da literatura se dá no levantamento de designs de pesquisa interessantes. Kapiszewski e Taylor (2008) citam, por exemplo, que a maioria das obras encontradas no estudo observavam tribunais superiores, o que torna a comparação de diferentes instâncias em nível estadual e federal mais atrativa devido a sua escassez na literatura.

A pesquisa pretende se inserir dentro desse conjunto bibliográfico por, de um lado, se valer de ferramentas e metodologias já desenvolvidas, e, por outro, tentar suprir algumas lacunas e perguntas em aberto que identificamos neste levantamento por meio de uma perspectiva brasileira.

3. DELIMITAÇÃO DO UNIVERSO DE PESQUISA

Dado que esta pesquisa buscou analisar decisões judiciais que aplicam a legislação eleitoral, os esforços de coleta centraram-se exclusivamente nos Tribunais Regionais Eleitorais e no Tribunal Superior Eleitoral, os quais compõem a Justiça Eleitoral.

Esta delimitação significa que os órgãos do Poder Judiciário pertencentes à Justiça Estadual, Federal, Trabalhista e Militar não foram observados. Assim, apenas uma fração do total de documentos produzido pelo Poder Judiciário foi coletada: são ajuizados aproximadamente 29 milhões de novos processos por ano, entre os quais menos de 1 milhão pertencem à Justiça Eleitoral¹¹. Com isso, e levando em conta a liberdade que juízes possuem para aceitar demandas e fundamentar as suas decisões, é possível que existam decisões fora da Justiça Eleitoral que sejam pertinentes para esta pesquisa. Entretanto, dado o elevadíssimo número de documentos, a diversidade nas fontes e a baixa padronização, avaliar este corpo de decisões é inviável.

Ademais, para evitar um excesso na complexidade da coleta de decisões, esta pesquisa optou por deixar de coletar decisões do Supremo Tribunal Federal, ainda que esse tribunal tenha competência para julgar recursos de decisões da Justiça Eleitoral¹².

Dentro do conjunto de processos nos Tribunais Regionais Eleitorais e Tribunal Superior Eleitoral, o universo de pesquisa se centra em três elementos: recorte temporal, tipo de documento observado, e plataforma utilizada como fonte de dados.

Com o objetivo de analisar decisões em torno das eleições de 2018, escolhemos ações ajuizadas entre dezembro de 2017 e março de 2019. Por um lado, dezembro de 2017 permite observar o regimento da Resolução nº 23.551¹³ de 18 de dezembro de 2017, que dispõe sobre propaganda eleitoral, utilização e geração do horário gratuito e condutas ilícitas em campanha eleitoral nas eleições de 2018. Em especial, a resolução introduz o Art. 22, §1º, que limita a livre manifestação do pensamento do eleitor na

¹¹ Conforme relatórios do Justiça em Números, disponíveis em: <<http://www.cnj.jus.br/pesquisas-judiciarias/justicaemnumeros/2016-10-21-13-13-04/pj-justica-em-numeros>>. Acesso em: 01 de abril de 2019.

¹² Art. 120, §3º c/c art. 102, inciso III, Constituição Federal.

¹³ Resolução disponível em: <<http://www.tse.jus.br/legislacao-tse/res/2017/RES235512017.html>>. Acesso em: 13 de março de 2019.

Internet nos casos de ofensa à honra de terceiros ou divulgação de fatos sabidamente inverídicos. Por outro lado, março de 2019 permite a inclusão de processos relativos às eleições de 2018 em sua totalidade, levando-se em conta uma expectativa razoável de que estes processos tenham sido julgados e que decisões tenham surgido em até cinco meses depois do evento¹⁴. Essa fase da pesquisa foi realizada no decorrer de 2019 e 2020.

No que diz respeito à segunda fase da pesquisa, em que coletamos e analisamos decisões das eleições de 2020, foram coletadas as decisões publicadas entre janeiro de 2019 e janeiro de 2021. O recorte temporal da fase da pesquisa realizada no decorrer de 2021 foi definido a partir dos resultados da primeira fase, que revelaram que esse período de 12 meses compreende a gigantesca maioria dos processos relevantes sobre o tema durante o período eleitoral.

Os documentos observados são as decisões judiciais da Justiça Eleitoral que tratam de *fake news* e desinformação online, sendo as especificidades da metodologia descritas em sequência. Ao coletar decisões de todas as instâncias e de todos os estados brasileiros, é possível realizar estudos comparativos e quantitativos. Da mesma forma, decisões judiciais frequentemente trazem comentários de juízes que são informativos para a análise qualitativa do tema, ainda que não sejam necessariamente relevantes para o mérito da decisão (e. g. comentários circunstanciais sobre *fake news* e desinformação em geral). Informações adicionais às decisões judiciais — como partes, data de distribuição e dados geográficos, também são úteis para análises quantitativas e qualitativas, e por isso foram coletadas.

3.1. Esforço preliminar para a construção do universo

De uma forma geral, pesquisas jurisprudenciais na área do Direito tomam como um ponto de partida comum as ferramentas digitais de pesquisa disponibilizadas pelos tribunais. No contexto da Justiça Eleitoral, isso significa lançar mão do portal do TSE, que conta

¹⁴ A escolha temporal da amostra apresenta processos que, devido ao fim do pleito de 2018, chegaram ao fim ou foram extintos sem resolução de mérito devido a perda de objeto (Art. 485, inciso IV, CPC). A data também é reflexo do cronograma de pesquisa com base em experiências anteriores de pesquisa em Tribunais Eleitorais, principalmente com o projeto "Justiça Eleitoral e Conteúdo Digital nas Eleições de 2014," conduzida pelo Centro de Ensino e Pesquisa em Inovação nos anos 2015 e 2016. Para mais informações, cf. <https://direitosp.fgv.br/sites/direitosp.fgv.br/files/arquivos/pesquisa_justica_eleitoral_fgv_2.pdf>. Acesso em: 22 de abril de 2019.

com uma página dedicada a decisões monocráticas do próprio TSE¹⁵ e outra para pesquisas de decisões colegiadas do TSE, além de decisões monocráticas e colegiadas dos tribunais regionais¹⁶.

Figura 1: Exemplo de documento disponível no portal do TSE

Documento 1:

0601851-89.2018.6.00.0000

AIJE - Ação de Investigação Judicial Eleitoral nº 060185189 - BRASÍLIA - DF

Acórdão de 13/12/2018

Relator(a) Min. Jorge Mussi

Publicação:

DJE - Diário de justiça eletrônico, Tomo 48, Data 12/03/2019

Ementa:

AÇÃO DE INVESTIGAÇÃO JUDICIAL ELEITORAL. ELEIÇÕES DE 2018. PRESIDENTE E VICE-PRESIDENTE DA REPÚBLICA. PRELIMINARES. INÉPCIA DA PETIÇÃO INICIAL. ILEGITIMIDADE PASSIVA. REJEIÇÃO. DEPOIMENTO PESSOAL. MEIO DE PROVA. FALTA DE PREVISÃO LEGAL. CONSENTIMENTO DA PARTE. POSSIBILIDADE. PRESCINDIBILIDADE NO CASO CONCRETO. ALEGAÇÃO. ABUSO DO PODER ECONÔMICO. ELEMENTOS. CARACTERIZAÇÃO. USO. RECURSOS PÚBLICOS OU PRIVADOS. GRAVIDADE. DESEQUILÍBRIO DO PLEITO. PROPAGANDA ELEITORAL NEGATIVA. APRESENTAÇÕES ARTÍSTICAS. CANDIDATO. PRESERVAÇÃO DA IGUALDADE DE CONDIÇÕES NA DISPUTA. CONFIGURAÇÃO. ATO ABUSIVO. EXIGÊNCIA. PROVA SEGURA. AUSÊNCIA DE DEMONSTRAÇÃO. IMPROCEDÊNCIA.

Fonte: Ferramenta de pesquisa de jurisprudência do TSE, disponível em: <<http://www.tse.jus.br/jurisprudencia/decisoes/jurisprudencia>>. Este resultado é encontrado buscando pelo número CNJ 0601851-89.2018.6.00.0000.

Desse modo, a primeira fase da pesquisa compreendeu também um esforço de testes metodológicos que permitissem a identificação da melhor forma de coleta. O esforço inicial consistiu na coleta de todos os resultados possíveis desses mecanismos. Para isso, utilizamos as ferramentas inserindo apenas o critério de data, ou seja, sem especificar outros elementos que pudessem filtrar resultados. Como as ferramentas do site limitam-se a exibir 1.000 resultados, realizamos consultas com intervalos de um dia para cada tribunal, iniciando em 01/01/2017. No total, foram realizadas aproximadamente 21 mil consultas.

O resultado das consultas consiste em informações sobre ações julgadas na data especificada. As informações variam bastante, quase sempre se referindo a dados do processo, como número, data, julgadores, a ementa do julgamento e um link para o

¹⁵ Disponível em: <<http://www.tse.jus.br/jurisprudencia/decisoes/monocraticas-do-tse>>. Acesso em: 01 de abril de 2019.

¹⁶ Disponível em: <<http://www.tse.jus.br/jurisprudencia/decisoes/jurisprudencia>>. Acesso em: 01 de abril de 2019.

arquivo da decisão. Estes dados são organizados de forma estruturada para o olhar humano, ou seja, um observador consegue identificar de forma objetiva a qual categoria cada dado pertence.

Entretanto, a disposição deles — na maior parte inseridos como um bloco de texto — exigiu que fossem elaborados padrões para a sua extração automatizada. Esses padrões foram criados com algum sucesso, de modo que os dados foram inseridos em uma planilha gerencial. Além disso, todos os arquivos relacionados — cerca de 200.000 — foram baixados e organizados em pastas por Tribunal.

Apesar do volume, o material coletado por esse método não se mostrou satisfatório. Os dados se revelaram insuficientes para análises precisas, e os documentos acessados, onde se encontra o texto integral das decisões, estavam em formatos cuja extração do texto era difícil. Esses problemas tornavam mais complexa a tarefa de identificar quando decisões faziam parte de um mesmo processo. Por fim, diversas decisões que podem ser substanciais, tais como despachos e decisões interlocutórias, ficaram de fora.

Embora essa coleta não tenha resultado em um material útil para a pesquisa, ela permitiu um primeiro contato com o material empírico. Da observação desse universo, foi possível observar que o número identificador padronizado dos processos que tramitam na plataforma Processo Judicial Eletrônico inicia, invariavelmente, com 0600001. Conforme é relatado em sequência, essa constatação foi importante para viabilizar a coleta definitiva por meio dessa plataforma.

3.2. Levantamentos no Processo Judicial Eletrônico

Para superar essas dificuldades, os esforços foram voltados para a plataforma Processo Judicial Eletrônico (PJe) da Justiça Eleitoral¹⁷. Aqui, o resultado da pesquisa é a informação processual, que contém dados sobre o andamento do processo e os documentos produzidos pelos juízes. Os processos disponibilizados não são o conjunto total de processos na Justiça Eleitoral, mas apenas algumas classes. A Portaria nº 1.143/2016¹⁸ especifica quais classes devem ser incluídas a partir de 20 de dezembro de

¹⁷ Consulta pública disponível em: <<https://pje.tse.jus.br:8443/pje-web/ConsultaPublica/listView.seam>>. Acesso em: 13 de março de 2019.

¹⁸ Portaria disponível em: <<http://www.tse.jus.br/legislacao-tse/prt/2016/PRT11432016.html>>. Acesso em: 13 de março de 2019.

2016¹⁹, sendo que alguns tribunais avançaram além desses requisitos. As classes de utilização obrigatória incluem (i) Petição e (ii) Representação, as mais comuns para processos envolvendo conteúdo online apontado como falso, de forma que é plausível crer que uma proporção suficientemente representativa dos processos que buscamos estão contidos nesta base de dados, justificando sua utilização para a extração dos documentos.

Os Tribunais Regionais Eleitorais, frise-se, possuem competência originária em julgar todas as representações e pedidos de direito de resposta envolvendo as eleições federais e estaduais conforme disposto no artigo 3º da Resolução 23.547 de 18 de dezembro de 2017²⁰. Ainda, essa mesma Resolução atribuiu competência originária ao Tribunal Superior Eleitoral para julgar representações e pedidos de direito de resposta envolvendo as eleições presidenciais.

Figura 2: Exemplo de documento disponível no PJe

Número Processo	Data da Distribuição	Classe Judicial	Assunto
0601851-89.2018.6.00.0000	26/10/2018	AÇÃO DE INVESTIGAÇÃO JUDICIAL ELEITORAL (11527)	DIREITO ELEITORAL - Eleições - Cargos - Cargo - Presidente da República DIREITO ELEITORAL - Eleições - Cargos - Cargo - Vice-Presidente da República DIREITO ELEITORAL (11428) - Eleições (11583) - Transgressões Eleitorais (11716) - Abuso (11717) - Abuso - De Poder Econômico
Jurisdição	Órgão Julgador Colegiado	Órgão Julgador	Eleição
TSE	Colegiado do Tribunal Superior Eleitoral	Corregedor Geral Eleitoral Ministro Jorge Mussi	2018 - Eleições Gerais
Município/UF			
BRASÍLIA / DISTRITO FEDERAL			

Fonte: Ferramenta de consulta de processos eletrônicos do TSE, disponível em: <<https://pje.tse.jus.br:8443/pje-web/ConsultaPublica/listView.seam>>. Este resultado é encontrado buscando pelo CNJ 0601851-89.2018.6.00.0000.

Nas informações processuais, os dados dos processos no PJe são organizados em campos (e não em um bloco de texto), o que torna a sua coleta mais fácil e segura. Os

¹⁹ Atualmente, há 18 classes processuais de utilização obrigatória do PJe. São elas: Ação de Impugnação de Mandato Eletivo; Ação de Investigação Judicial Eleitoral; Ação Rescisória; Conflito de Competência; Consulta classe exclusiva do TSE; Criação de Zona Eleitoral ou Remanejamento; Exceção; Instrução; Lista Tríplice classe exclusiva do TSE; Petição; Prestação de Contas; Propaganda Partidária; Reclamação; Recurso Contra Expedição de Diploma; Registro de Partido Político; Representação; Suspensão de Segurança; e Processo Administrativo.

²⁰ “Art.3º. As representações poderão ser feitas por qualquer partido político, coligação, candidato ou pelo Ministério Público e deverão dirigir-se (Lei nº 9.504/1997, art. 96, caput, incisos II e III): I – ao Tribunal Superior Eleitoral, na eleição presidencial; II – aos tribunais regionais eleitorais, nas eleições federais, estaduais e distritais”. Disponível em: <<http://www.justicaeleitoral.jus.br/arquivos/tse-representacoes-reclamacoes-e-pedidos-de-resposta-eleicoes-2018>>. Acesso em: 17 de junho de 2019.

documentos produzidos por juízes estão em linguagem HTML (*Hypertext Markup Language*), que é de fácil leitura automatizada. Além disso, a informação reúne todos os documentos, o que elimina o risco de considerar duplamente um caso se houver mais de uma decisão pertinente. Por fim, esses documentos não se limitam a decisões finais, contendo também despachos e decisões interlocutórias, o que enriquece o universo de pesquisa.

Cumprido destacar que o PJe possui outras informações que não foram colhidas nesse momento, tais como o acesso integral aos autos. Ocorre que este acesso depende de identificação no sistema por meio de login utilizando certificado digital, o que poderia levar a um bloqueio de acessos dado o volume de consultas. Além disso, a informação contida em outros elementos de um processo que não os documentos produzidos por juízes seria de difícil processamento em razão da falta de padronização. Outra limitação evidente do sistema é a ausência de informações a respeito de processos sigilosos, o que foi constatado no decorrer da pesquisa. Assim, esta pesquisa focou nos documentos cujo acesso e processamento são possíveis dado o contexto e recursos disponibilizados.

3.3. Coleta de decisões judiciais

Cada tribunal possui o seu próprio ponto de acesso ao PJe, ou seja, cada Tribunal Regional Eleitoral e o Tribunal Superior Eleitoral contam com um site próprio para o sistema. Porém, os sistemas são padronizados, existindo uma pequena diferença apenas no Tribunal Superior Eleitoral na forma como o campo do número do processo é preenchido.

Os campos para pesquisa são bastante limitados, podendo-se buscar processos por nome e CNPJ/CPF da parte, advogado das partes e número do processo.

O método utilizado foi a busca pelo número do processo que foi uniformizado²¹ em razão da Resolução N° 65 de 16/12/2008 do CNJ. De um modo geral, o padrão da numeração segue o formato NNNNNNN-DD.AAAA.J.TR.OOOO.

- NNNNNNN: Número sequencial do processo. Em tese, o primeiro processo do ano recebe o número 0000001, o segundo recebe 0000002, e assim por diante.

²¹ Disponível em: <<http://www.cnj.jus.br/atos-normativos?documento=119>>. Acesso em: 01 de abril de 2019.

Observando os resultados do esforço preliminar de pesquisa, percebemos que processos com distribuição física (fora do PJe) começam pela entrada 00000001, ao passo que processos com distribuição digital iniciam pela sequência 06000001. Isso quer dizer que número do primeiro processo distribuído pelo PJe em um determinado ano começa com 06000001, o segundo com 06000002, e assim por diante.

- DD: Dígito verificador, calculado com base nos outros dígitos do número do processo, conforme especificado pela regulação.
- AAAA: Ano de ajuizamento do processo. Não respeita a data de última distribuição disponível na seção "Dados do Processo" de um processo no PJe e provavelmente depende da data de ajuizamento ou da data de primeira distribuição em casos de redistribuição. Pelo observado, esse parâmetro é alterado após a volta do recesso — por volta de 20 de janeiro. Isto é, processos de última distribuição nos primeiros dias de um ano poderão ter o ano anterior em seu número CNJ. Como exemplo, temos o processo número 0601271-05.2018.6.02.0000 que foi redistribuído em 28/01/2019.
- J: Identificador do órgão ou segmento. No caso da Justiça Eleitoral, o parâmetro recebe o número 6.
- TR: Identifica o tribunal dentro do segmento. No caso do Tribunal Superior Eleitoral, este parâmetro é 00. Para os Tribunais Regionais Eleitorais, ele vai de 01 a 27.
- OOOO: Unidade de origem do processo. No contexto do PJe na justiça eleitoral, todos são 0000.

Estabelecendo um ponto de partida, todos esses parâmetros podem ser montados. Tomando como exemplo o Tribunal Regional Eleitoral do Rio Grande do Sul (cujo identificador TR é 21) e o ano de 2017, temos que a parte final do número dos processos será 2017.6.21.0000 (equivalente aos parâmetros AAAA.J.TR.OOOO). Os cinco primeiros processos receberam na sua parte inicial os números 06000001, 06000002, 06000003, 06000004 e 06000005. Combinando as partes iniciais com a final, é possível calcular o dígito verificador. O resultado é o seguinte conjunto: 06000001-49.2017.6.21.0000, 06000002-34.2017.6.21.0000, 06000003-19.2017.6.21.0000, 06000004-04.2017.6.21.0000, 06000005-86.2017.6.21.0000.

Em outras palavras, acrescentando a parte inicial do número do processo e calculando o dígito verificador, foram estabelecidos todos os números de processo possíveis para um determinado tribunal em um determinado ano. Formada essa lista, o programa de *scraping*²² desenhado pela equipe técnica iniciou as consultas.

Há dois resultados possíveis para uma consulta por número de processo. O primeiro é que a consulta retorne um processo judicial. Nessa situação, o programa explora a informação processual e os documentos, armazenando este conjunto de informações num banco de dados estruturado.

Também é possível que o sistema retorne a mensagem “Sua pesquisa não encontrou nenhum processo disponível”. Considerando que o número do processo segue uma ordem sequencial, isso pode ocorrer porque, embora exista um processo, ele não está disponível para a consulta pública. Um exemplo disso é o processo número 0600034-55.2017.6.24.0000, em tramitação no Tribunal Regional Eleitoral de Santa Catarina, que consta como sigiloso no Diário Oficial²³. Pode-se cogitar também que a distribuição desse processo tenha sido cancelada por alguma razão interna do tribunal.

Por outro lado, também é possível que a consulta não retorne resultados porque a sequência de processos para aquele ano chegou ao final. Em relação ao Tribunal Regional de Roraima, tem-se que o último processo distribuído considerando o ano de 2017 foi de número 0600091-03.2017.6.23.0000 (ajuizado em 15/12/2017). Assim, os números de processo que seguem a sequência (0600092-85.2017.6.23.0000, 0600093-70.2017.6.23.0000) não retornam resultados. Nessa situação, para encontrar o próximo processo em ordem cronológica, é necessário alterar o parâmetro do ano e reiniciar o parâmetro sequencial para 0600001. Desse modo, o processo seguinte ao de número 0600091-03.2017.6.23.0000 foi o 0600001-58.2018.6.23.0000, ajuizado em 17/01/2018.

Nota-se, também, que essa metodologia presume a correta numeração dos processos. É possível que determinados tribunais ou varas da justiça eleitoral adotem sistemas diferentes de numeração, o que faria com que processos numerados de forma distinta não fossem coletados. Nas duas fases da pesquisa, contudo, notamos que, quando ocorre, a quantidade de processos numerados de forma diversa é ínfima se

²² *Scraping*, ou raspagem de dados, é um método automatizado de download sistemático de informações disponíveis em páginas web.

²³ Disponível em: <<https://www.tre-sc.jus.br/djesc-consulta/downloadPdf/diario-2017-125.pdf?ido=79812>>. Acesso em: 01 de abril de 2019.

comparada ao universo total e, normalmente, está concentrada em varas de cidades pequenas com poucos processos.

Assim, consideradas essas limitações, esse funcionamento foi utilizado para determinar o ponto final de cada rotina de consulta. Ainda no exemplo do Tribunal Regional de Roraima, iniciamos a consulta pelo processo de número 0600001 em 2017 e carregamos a sequência até encontrar 100 resultados consecutivos que não retornavam processos. Constatando a sequência de consultas que não retornavam processos, o programa entende que deve seguir para o ano seguinte, reiniciando o número sequencial. Seria possível, para garantir a coleta de mais processos, aumentar essa quantidade de resultados consecutivos para 150, 200, ou até 1000 processos. Contudo, entendemos que o acréscimo de poder computacional necessário (e, consequentemente, o tempo de pesquisa acrescido) não justificaria os poucos processos a mais que seriam obtidos.

Em um primeiro momento, portanto, os pesquisadores buscaram compilar todas as decisões judiciais e informações processuais disponibilizadas no PJe entre 2017 e março de 2019 para, em seguida, prosseguir com a formação e filtragem da amostra de pesquisa.

Ao final dos procedimentos da primeira fase da pesquisa, referente às eleições gerais de 2018, o banco de dados compilado reuniu informações de 95.684 processos judiciais²⁴. Avaliando os resultados, entendemos que os únicos processos que não foram coletados foram aqueles que, inseridos no PJe, não foram disponibilizados para consulta pública. O número desses processos pode ser de até 825, podendo ser menor em razão de eventuais distribuições canceladas.

²⁴ Nossos dados são de difícil comparação com os dados do Justiça em Números do Conselho Nacional de Justiça por alguns motivos. Primeiramente, a última edição do Justiça em Números de 2018 cobre o ano de 2017 e não apresenta dados de 2018 ou 2019. Segundo, mesmo para o ano de 2017, não há uma variável que agregue todos os processos de 2017 presentes especificamente no Processo Judicial Eletrônico. Ainda, buscamos coletar processos de todas as classes judiciais disponíveis no Processo Judicial Eletrônico, o que difere da metodologia de contagem de casos do Justiça em Números. Por fim, vale ressaltar que casos relacionados às eleições de 2016 estão fora de escopo desta pesquisa. Essas ressalvas como um todo razoavelmente explicam discrepâncias. Com isso em mente, de acordo com Justiça em Números 2018 (Conselho Nacional de Justiça, 2018), a Justiça Eleitoral registrou 27.683 casos novos entre os Tribunais Regionais Eleitorais e 3.370 casos novos no Tribunal Superior Eleitoral em 2017. Nossa coleta agregou 5.430 casos em TREs e 4.343 casos no TSE em 2017. Para o ano de 2018, foram coletados 72.987 casos entre TREs e 1.997 casos no TSE. Por fim, de janeiro até março de 2019 foram coletados 4.100 processos nos TREs e 123 processos no TSE.

No caso da segunda fase da pesquisa, o banco de dados compilado reuniu informações de 31.725 processos. Acreditamos que o número reduzido de processos coletados na segunda fase se justifica, pois uma parte maior dos conflitos eleitorais nas eleições municipais é solucionada ainda na primeira instância, em cartórios eleitorais. Somente parte deles sobem aos Tribunais Regionais em sede de recurso ou Mandado de Segurança, o que resulta em menos processos obtidos mediante consulta a esses Tribunais. Diferentemente, nas eleições gerais, os TREs tem competência originária para processos que tratam, por exemplo, de disputas entre governadores.

4. FILTRAGEM E FORMAÇÃO DA AMOSTRA DE PESQUISA

A seleção das decisões judiciais que compõem a amostra a ser analisada pelos pesquisadores se deu prioritariamente mediante um processo automatizado de busca de palavras-chave agrupadas em conjuntos. O sistema buscador de palavras analisou cada um dos documentos HTML dos processos no PJe (compilados conforme descrito na seção anterior), verificando a presença simultânea de ao menos um dos elementos de cada um de dois conjuntos de palavras distintos.

Esses dois conjuntos de palavras foram definidos de forma a assegurar que o buscador de palavras obtivesse o maior número possível de documentos relacionados ao objeto da pesquisa. Nesse sentido, os pesquisadores desenvolveram o buscador de palavras e o alimentaram com palavras-chave com a finalidade de compilar decisões judiciais julgadas por completo entre dezembro de 2017 e março de 2019 que tratam de conteúdo apontado como inverídico, falso, fraudulento ou enganoso publicado por meios digitais.

O Conjunto 1 diz respeito, portanto, a palavras-chave relativas ao fenômeno das *fake news* e desinformação. Mais especificamente, selecionamos expressões comuns no debate contemporâneo sobre o tema mencionadas pela literatura especializada (“*fake news*,” “*fake*,” “desinformação,” “fato sabidamente inverídico”). Além disso, diante do entendimento de que a noção de desinformação abarca diversos tipos de informação inverídica ou fraudulenta, combinamos diferentes formas de conteúdo (“fato,” “notícia,” “matéria,” “conteúdo,” “informação”) com expressões adjetivas que denotam inveracidade ou fraudulência (“inverídico,” “falso,” “fraudulento,” “enganoso”). Por fim, incluímos também variantes das expressões “sabidamente inverídico(a)” e “fidedignidade da informação” (somente na base das eleições de 2020) para capturar a linguagem normativa relevante.

O Conjunto 2 identifica a natureza digital do objeto de pesquisa, contendo, portanto, expressões relacionadas a aspectos técnicos da Internet, expressões comuns do jargão de usuários, e menções às principais plataformas de redes sociais utilizadas no Brasil.

Importa destacar que, inicialmente, desenvolvemos um terceiro conjunto de palavras-chave para realizar um recorte temporal a partir de Dezembro de 2017, justificado na seção 3, e necessário devido à raspagem que incluiu o ano de 2017 em sua totalidade. No entanto, este terceiro conjunto não se provou suficiente para capturar todos os processos de 2018, dado que nem sempre há indicações textuais de data e ano nos documentos disponíveis. Ainda, consideramos utilizar para esse fim um dado específico presente nos processos do PJe que identifica o período eleitoral referente. Esse campo, porém, também apresenta inconsistências e erros de preenchimento. Optamos, portanto, por realizar a filtragem temporal manualmente²⁵.

Em síntese, quando realizada uma seleção de decisões, um documento só seria coletado pelo buscador de palavras quando contivesse, simultaneamente, pelo menos uma palavra-chave do Conjunto 1 e ao menos uma palavra-chave do Conjunto 2. Nesse sentido, optamos por uma amostragem automatizada que tivesse parâmetros mínimos de coleta para evitar erros de não-coleta de casos relevantes, com a contrapartida de incluir mais casos de falsos positivos para revisão manual.

A título de exemplo, seria selecionado um documento que contivesse as palavras-chave “fake news,” contida no Conjunto 1 e “internet,” contida no Conjunto 2. No entanto, não seria selecionado um documento que somente contivesse as palavras-chave “fake” e “desinformação,” ambas contidas no Conjunto 1, diante da ausência de palavras-chave do Conjunto 2 e, conseqüentemente, da qualificação do conteúdo em questão como publicado em meios digitais.

Para garantir a completude da pesquisa, a busca pelas expressões contidas nos conjuntos se deu por meio de *regex*. Em Ciência da Computação, *regex* (*regular expressions* ou, em português, expressões regulares) são padrões utilizados para encontrar expressões em um texto. Com eles, podemos encontrar diferentes variações de uma determinada expressão-base sem ter que recorrer a várias buscas especializadas. Por exemplo, se desejarmos encontrar em um texto todas as ocorrências das expressões “O carro A quebrou” e “O carro B quebrou”, não precisamos escrever dois trechos distintos de códigos de programação que busquem por cada uma dessas expressões: basta escrevermos um só trecho de código que utilize para a busca o padrão (a expressão regular) “O carro [A|B] quebrou”. Nesse caso, portanto, o trecho “[A|B]”, específico da

²⁵ As *regular expressions* utilizadas para buscar datas foram as seguintes: “(elei.{8}2018)”, “2018”, “(dezembro.{8}2017)” e “12.{5}2017”.

linguagem utilizada pelas *regex*, indica ao programa que ele deve capturar expressões cujo caractere que se encontra após a palavra “carro” e antes da palavra “quebrou” seja A ou B – ou seja, o operador de linguagem “|” pode ser traduzido como “ou”. Se quiséssemos que a expressão “O carro C quebrou” também fosse encontrada, bastaria que substituíssemos o trecho [A|B] da *regex* descrita pelo trecho [A|B|C].

Uma expressão regular tem, é claro, limitações. Assim, o padrão “O carro [A|B] quebrou” é capaz de captar as expressões “O carro A quebrou” e “O carro B quebrou”, mas não é capaz de captar as expressões “Quebraram o carro A” ou “Quebraram o carro B”. Para que possamos captar todas essas quatro sentenças em um só trecho de código, precisaremos utilizar o operador “|” para aglutinar duas expressões regulares em uma só: “O carro [A|B] quebrou|Quebraram o carro [A|B]”. A complexidade das *regex* aumenta, portanto, quando queremos expandir sua abrangência e sua precisão simultaneamente.

Nossas formulações de *regex* serviram para incluir a coleta de expressões que estão gramaticalmente erradas; flutuações de caracteres e acentuação – como espaços duplos, ausência de espaços, barras, pontos e hífen entre palavras; e palavras dentro de uma expressão, como, por exemplo, as variantes “fato sabidamente inverídico”, “fato que se sabe inverídico” e “fato manifestamente inverídico”, que, na linguagem de programação Python, podem ser encontradas por meio da *regex* “\W(fato.{,20}inver.dic.*?)\W”. Os resultados de palavras de *regex* foram curados de forma que resultados indesejados foram excluídos do sistema de busca. Notadamente, o *regex* “\W(post.*?)\W” foi o único com exclusão de mais de 3 termos devido a resultados associados às palavras “postal”, “postergar”, “posto” e “postular”.

Quando selecionados e coletados pelo buscador de palavras, os documentos foram armazenados em pastas identificando seus respectivos processos. Cada pasta, portanto, centraliza todos os documentos encontrados que fazem parte de um mesmo processo.

A seleção final das expressões regulares Python que orientaram o buscador de palavras, ao fim das rodadas de filtragem e revisão dos falsos positivos, foi a seguinte:

Tabela 1: Conjuntos de Termos de Busca para Seleção de Processos Judiciais*

C1. Fake News	C2. Internet e Meio Digital	
(fake\s?.*)\W (desinform.*)\W \W(conte.do.{20}enganos.*)\W \W(conte.do.{20}fals.*)\W \W(conte.do.{20}fraudulent.*)\W \W(conte.do.{20}inver.dic.*)\W \W(fato.{20}fals.*)\W \W(fato.{20}inver.dic.*)\W \W(informa.{20}enganos.*)\W \W(informa.{20}fals.*)\W \W(informa.{20}fraudulent.*)\W \W(informa.{20}inver.dic.*)\W \W(mat.ria.{20}fals.*)\W \W(mat.ria.{20}fraudulent.*)\W \W(mat.ria.{20}inver.dic.*)\W \W(not.cia.{20}enganos.*)\W \W(not.cia.{20}fals.*)\W \W(not.cia.{20}fraudulent.*)\W \W(not.cia.{20}inver.dic.*)\W \W(sabidament.{20}enganos.*)\W \W(sabidament.{20}fals.*)\W \W(sabidament.{20}fraudulent.*)\W \W(sabidament.{20}inver.dic.*)\W \b fidedig.{20} inform.*\b** \b inform.{20} fidedig.*\b**	“aplicativos?” “digita[il]s?” “\W(dns)\W” “e.?mails?” “internet” “\W(ips?)\W” “link?” “on.{6}line” “urls?” “websites?” “(blog.*)\W” “blogspot” “linkedin” “face.?book” “fanpages?” “google” “instagra[mn]” “messenger” “stories” “youtube” “twitter” “telegra[mn]” “whatsapp”	“(buscador.*)\W” “(endere.{8}eletr.*)\W”*** “\W(hash.*)\W” “(m.dia.{5}soci.*)\W” “(meio.{4}eletr.*)\W”*** “(p.gina.{8}eletr.*)\W” “(p.gina.{8}socia.*)\W” “(plataforma.{8}virtu.*)\W” “(portal.{8}busca.*)\W” “\W(post.*)\W” “(provedor.{7}aplica.*)\W” “(provedor.{7}conex.*)\W” “(provedor.{7}conte.*)\W” “(rede.{6}comuni.{8}instan.*)\W” “(rede.{5}mundi.{8}compu.*)\W” “(rede.{5}soci.*)\W” “(s.tio.{5}eletr.*)\W” “\W(t[wu]it.*)\W \Wtweet”
Total de resultados: 607 expressões.	Total de resultados: 193 expressões.	

Estas expressões regulares apresentaram resultados de palavras dentre todos os casos coletados no PJe. As seguintes expressões não retornaram resultado algum: “pinterest”, “reddit”, “tumblr”, “\W(fato.{20}fraudulent.)\W”, “\W(fato.{20}enganos.*)\W”, “\W(mat.ria.{20}enganos.*)\W”.

** Adicionadas somente na segunda fase da pesquisa

***Os termos associados às expressões “meio eletrônico” e “endereço eletrônico” foram removidos do buscador de palavras na segunda fase da pesquisa, tendo em vista que quase todas as seções de julgamento foram realizadas de forma remota durante a pandemia de COVID-19, o que tornou a terminologia excessivamente comum mesmo em processos não relacionados à conteúdo online.

Após o software de *scraping* compilar com sucesso todas as decisões judiciais e informações processuais disponibilizadas no PJe dentro do recorte temporal estabelecido, o passo seguinte foi a realização de uma primeira etapa de filtragem, com

o objetivo de avaliar se o buscador de palavras automatizado estava, conforme planejado, obtendo o máximo possível de documentos relacionados ao objeto da pesquisa com volumes gerenciáveis de falsos positivos.

4.1. Rodadas de filtragem automatizada

A primeira filtragem pelo buscador de palavras foi feita por meio de busca de palavras completas, ou seja, ainda sem a utilização de *regular expressions*, e resultou em uma base de 2495 processos diferentes. Passamos, em seguida, a ler de forma preliminar uma quantidade representativa de decisões de processos em cada tribunal, de forma a verificar se os conjuntos de palavras-chave seriam, ao mesmo tempo, precisos e abrangentes o suficiente para coletar o máximo possível de resultados relacionados ao objeto de pesquisa sem incorrer em um número excessivo de falsos positivos.

Após a leitura de documentos de 900 processos, identificamos:

- (a) a presença de algumas palavras-chave contidas no Conjunto 2 que não estavam exclusivamente relacionadas aos meios de publicação digital. Essas palavras-chave geraram uma quantidade grande de resultados que não tratavam de conteúdo digital, carecendo de pertinência para nossa pesquisa. As palavras problemáticas (“aplicação,” “imagem,” “mensagem,” “foto,” “vídeo”) foram removidas do Conjunto 2 nas filtrações subsequentes.
- (b) a ausência de algumas palavras-chave que poderiam contribuir para o maior alcance do buscador de palavras, evitando falsos negativos. Essas palavras (e.g. “plataforma virtual,” “online,” “tuíte”) foram adicionadas ao buscador de palavras nas filtrações subsequentes após sua observação em decisões na amostra.
- (c) processos que, mesmo contendo palavras-chave dos dois Conjuntos, não eram relevantes para os objetivos da pesquisa e, portanto, precisariam ser analisados e filtrados manualmente (conforme procedimento descrito na seção 4.2).

Após atualizar o buscador de palavras com 29 novas entradas, foi realizada uma segunda filtragem sobre a base de dados total de processos do TSE e dos TREs no PJe, que teve como objetivo consolidar e preparar a amostra para subsequentes filtrações manuais. Dessa segunda filtragem resultou uma base de 1971 processos, dos quais 1970 já foram encontrados pela amostra anterior.

Diferentemente da primeira amostragem que coletou apenas documentos de um processo que se adequam aos dois conjuntos de palavras-chave, as filtrações

subsequentes também salvaram todos os outros documentos que compunham o mesmo processo. Em outras palavras, bastava que um dos documentos do processo cumprisse os parâmetros de busca para que o processo fosse coletado integralmente, conforme os limites levantados na [seção 3.2](#).

Após a avaliação e debate da metodologia de coleta e dos resultados da segunda amostra com especialistas num workshop de pesquisa²⁶, realizamos uma terceira filtragem automatizada implementando *regular expressions* com os termos presentes na **Tabela 1** a partir das palavras-chave utilizadas nas rodadas anteriores.

Na primeira fase da pesquisa, referente às eleições de 2018, isso resultou em uma base de 2928 processos incluindo os 1971 casos já conhecidos. Uma análise preliminar da amostra demonstrou que os resultados foram precisos o suficiente para que não fosse necessária uma nova atualização dos parâmetros de busca. Na segunda fase da pesquisa, isso resultou em uma base de 950 processos a serem filtrados manualmente.

4.2. Filtragem manual e análise de relevância

Durante a primeira fase da pesquisa, a análise da amostra resultante da filtragem automatizada revelou que precisávamos desenvolver critérios de relevância que não poderiam ser implementados por meio do buscador de palavras. A necessidade da filtragem manual e da definição destes critérios de relevância derivou do reconhecimento de que um grupo de resultados da filtragem automatizada continha as palavras-chave que definimos, mas não se relacionava necessariamente com o objeto da pesquisa. A partir das características dos grupos de processos que identificamos na filtragem automatizada, desenvolvemos os seguintes critérios para a análise manual:

(a) Processos identificados como verdadeiro positivo:

São os processos que, simultaneamente: (i) têm como cerne da discussão conteúdo apontado como inverídico, falso, fraudulento ou enganoso, conforme terminologia associada a noções de *fake news* e desinformação; (ii) tratam de

²⁶ Workshop de pesquisa fechado realizado dia 21 de Maio de 2019 com o objetivo de validação da metodologia de coleta de dados e avaliação da amostra. Contou com especialistas de diferentes setores, como acadêmicos, membros da sociedade civil, membros do empresariado e operadores do direito. Gostaríamos de agradecer imensamente pela participação de todas e todos.

conteúdo publicado e/ou compartilhado por meios digitais e (iii) se referem às eleições de 2018.

(b) Processos identificados como falso positivo:

São processos que, ainda que tenham se adequado aos parâmetros do buscador de palavras, não cumprem os critérios de verdadeiro positivo e, portanto, não são relevantes para o desenvolvimento dos objetivos de pesquisa. Estes incluem conteúdos referentes a mídias offline como rádio e televisão; conteúdo online apontado exclusivamente como ofensivo à honra, como injúrias e expressões de baixo calão; consultas a respeito de regras eleitorais; processos de validação de chapas; e despachos que não tratam do objeto de pesquisa.

Delimitados esses critérios, foi desenvolvida uma tabela a ser preenchida pelos pesquisadores para registro da filtragem manual. Essa tabela contém em cada linha um número identificador dentro da amostra, o número CNJ de um processo, seu tribunal de origem, seu município de origem, a eleição pertinente conforme dados do PJe, e a classe judicial conforme dados do PJe. A tabela também contém uma coluna “verdadeiro_c2c3” e outra coluna “verdadeiro_c1” que deveriam ser preenchidas pelos pesquisadores com os valores “Sim” ou “Não”. A primeira observa se um processo respeita os critérios (ii) e (iii), enquanto a segunda observa se o critério (i) é atendido. Optamos pela leitura integral de decisões para a filtragem, ainda que o relatório seja particularmente elucidativo da descrição do caso.

Figura 3: Excertos da tabela de filtragem da amostra

verdadeiro_c2c3	verdadeiro_c1	id	CNJ	jurisdic	municipio_uf	eleicao	classe_judicial
Sim	Sim	422	0600198-95.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	457	0600233-55.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	800	0600576-51.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	805	0600581-73.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	806	0600582-58.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	825	0600601-64.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	845	0600621-55.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	848	0600624-10.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	856	0600632-84.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	869	0600645-83.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)
Sim	Sim	881	0600657-97.2018.6.02.0000	TRE-AL	Maceió / ALAGOAS	2018 - Eleições Gerais	REPRESENTAÇÃO (11541)

freq_sabidamente	media_por_arq	freq_ofens	media_por_arquivo	assunto
0	0	4	0.3333333333	DIREITO ELEITORAL - Eleições - Pesquisa Eleitoral - Registro de Pesquisa Eleitoral
3	0.75	34	8.5	DIREITO ELEITORAL - Eleições - Propaganda Política - Propaganda Eleitoral - Propaganda Política - Imprensa Escrita - Jornal/Revista/Tabloide
3	0.25	7	0.5833333333	DIREITO ELEITORAL - Eleições - Propaganda Política - Propaganda Eleitoral - Propaganda Política - Propaganda Eleitoral - Internet
7	1.75	14	3.5	DIREITO ELEITORAL - Eleições - Candidatos - Direito de Resposta
58	6.444444444	36	4	DIREITO ELEITORAL - Eleições - Propaganda Política - Propaganda Eleitoral - Propaganda Política - Propaganda Eleitoral - Outdoors
5	0.7142857143	8	1.142857143	DIREITO ELEITORAL - Eleições - Propaganda Política - Propaganda Eleitoral - Propaganda Política - Propaganda Eleitoral - Imprensa Escrita - Jornal/Revista/Tabloide
2	0.1111111111	11	0.6111111111	DIREITO ELEITORAL - Eleições - Transgressões Eleitorais - Abuso - Abuso - Uso Indevido de Meio de Comunicação Social
7	0.4375	40	2.5	DIREITO ELEITORAL - Eleições - Propaganda Política - Propaganda Eleitoral - Propaganda Política - Propaganda Eleitoral - Imprensa Escrita - Jornal/Revista/Tabloide
20	1	8	0.4	DIREITO ELEITORAL - Eleições - Candidatos - Direito de Resposta
2	0.1333333333	11	0.7333333333	DIREITO ELEITORAL - Eleições - Propaganda Política - Propaganda Eleitoral - Propaganda Política - Propaganda Eleitoral - Imprensa Escrita - Jornal/Revista/Tabloide
6	0.5	27	2.25	DIREITO ELEITORAL - Eleições - Propaganda Política - Propaganda Eleitoral

Em caso de dúvida a respeito da classificação do processo, pesquisadores poderiam marcar a opção “Dúvida”. Os processos identificados como “Dúvida” foram posteriormente discutidos entre a equipe para classificação correta.

Além das implementações de filtragem, a tabela também foi gerada de forma a apresentar as tags de assunto registradas no PJe, bem como contadores de frequência de palavras para variantes de “sabidamente inverídico” e “ofensivo” em cada processo, auxiliando a avaliação de dados que seguiria. A escolha das expressões se dá pela proximidade entre os dois casos no texto legal, de forma que a captura do termo “sabidamente inverídico” necessariamente implica em coleta de casos relativos à ofensa à honra. Entendemos que estes dois termos não são mutuamente exclusivos, posto que casos frequentemente utilizam das duas expressões. No entanto, é importante reforçar que o objeto de pesquisa é voltado para conteúdos apontados como inverídicos, falsos, fraudulentos ou enganosos, conforme terminologia associada a noções de *fake news* e desinformação. Nesse sentido, casos de ofensa à honra que não discutem a veracidade do conteúdo questionado são compreendidos como fora de escopo.

Outros conteúdos de difícil classificação apresentam um problema semelhante, na medida em que é pouco claro se há alguma afirmação de que um conteúdo online seria inverídico. Buscamos abarcar uma linguagem ampla para o objeto de pesquisa, validando expressões que indiquem calúnia, imputação de crime, insinuação falsa e afins. Por vezes, a identificação de um conteúdo como falso não apresenta implicações para a fundamentação, porém entendemos que isto é um dado de pesquisa e reflexo da natureza do objeto estudado.

Já outras expressões não foram tomadas como suficientes para atestar uma dimensão de falsidade. Este foi o caso do termo “notícia tendenciosa”, pois imputa um caráter de intencionalidade do conteúdo sem necessariamente acusá-lo de falso ou fraudulento. Também entendemos que difamação por si só, ou acompanhada somente

de uma acusação de injúria, não é suficiente para qualificar uma acusação de conteúdo falso ou fraudulento.

Em contrapartida, elementos que aproximam um conteúdo de casos notórios de desinformação foram aceitos na filtragem como verdadeiros positivos. Por exemplo, notícias antigas veiculadas como informações novas, notícias fora de contexto, e montagens ou trucagens que manipulam conteúdos de forma fraudulenta são casos incluídos na amostragem.

Por fim, foi feito um processo de validação em todos os Tribunais de verdadeiros positivos e falsos positivos identificados como conteúdo online — isto é, avaliados como pertencentes ao Conjunto 2 e não-pertencentes ao Conjunto 1. A rotina envolveu verificar os casos de verdadeiros positivos com maior contagem de conteúdo ofensivo, seguido da verificação de casos de menor contagem de conteúdo sabidamente inverídico, e terminando com a verificação de todos os casos de conteúdo online identificados como falsos positivos. Tribunais com um número pequeno de verdadeiros positivos ($n < 50$) foram verificados na íntegra. Também verificamos palavras-chave no campo “assunto” para confirmar pertinência de objeto online para todos os casos coletados. Desse esforço resultou a reavaliação de 177 casos, cerca de 10% do total de verdadeiros positivos e falsos positivos identificados como conteúdo online (1797 casos).

O processo de filtragem da base de dados na primeira fase da pesquisa resultou em 1492 processos verificados como verdadeiro positivo para o objeto de pesquisa proposto (dentre os 2928 identificados pela busca de palavras-chave nos 95.684 processos coletados via *scraping*).

No caso da segunda fase da pesquisa, a filtragem manual resultou num total de 444 processos (dentre os 950 identificados pela busca de palavras-chave nos 31.725 processos coletados via *scraping*).

5. ESTRUTURAÇÃO DOS DADOS

Consolidadas as bases de dados, passou-se a extrair e estruturar dados relativos aos processos em bases de dados temáticas, facilmente cruzáveis, de forma a viabilizar sua descrição e apresentação organizada. Desse processo de estruturação resultaram quatro planilhas: a Planilha de Partes, a Planilha de Processos, a Planilha de URLs e a Planilha de Precedentes.

5.1. A Planilha de Partes

A Planilha de Partes contém as informações relativas às partes dos processos coletados e filtrados. Sua construção foi dividida em quatro fases.

Em primeiro lugar, foram organizados os dados que já haviam sido extraídos automaticamente na fase de coleta pelo método de *scraping* (conforme item 3.3). A ferramenta de consulta dos processos eletrônicos do TSE, conforme exposto na **Figura 2**, fornece uma série de informações relativas a cada processo: o **número do processo**; o **nome das partes**; o **órgão julgador**; o município de origem; informações sobre a **movimentação processual**; a **eleição** a que o processo se refere e a **ementa do processo**. Todos esses dados foram organizados na Planilha de Processos, (conforme item 5.2) enquanto os dados sobre as partes foram reorganizados na Planilha de Partes. Nesta, cada linha se refere a um indivíduo ou entidade que foi parte em um determinado processo de acordo com as informações disponíveis no Processo Judicial eletrônico. Se um indivíduo ou entidade foi parte em mais de um processo, seu nome foi inserido repetidamente na planilha para cada evento, de forma que há uma linha para cada participação no polo ativo, polo passivo, ou como terceiro interessado.

O segundo passo foi estruturação limpeza dos dados relativos aos nomes das partes. Essa fase teve como objetivo padronizar os nomes de partes cadastrados na base de processos, de forma que uma mesma parte, quando envolvida em mais de um processo, fosse tratada pela mesma grafia em todos os casos. Quando a parte é um candidato, optou-se por adotar a grafia de acordo com os nomes cadastrados no TSE, de forma a viabilizar o cruzamento das bases de dados. A fase de limpeza, portanto, foi necessária para dois fins: (i) para que, ao navegar na base de processos, seu usuário pudesse associar automaticamente uma mesma pessoa (física ou jurídica) a todos os

processos em que ela participou através de um sistema de busca por partes; e (ii) para que fosse possível a extração automatizada de dados relativos a essas partes, por meio do cruzamento com a base de informações sobre candidatos do Repositório de Dados Eleitorais do Tribunal Superior Eleitoral²⁷.

No caso de partes que representam redes sociais, optamos por manter os nomes originais para que não fossem excluídas da planilha informações sobre, por exemplo, qual escritório da empresa (nacional ou internacional) foi citado. Se descontada a variabilidade dos escritórios, estamos lidando com apenas 5 empresas diferentes, o que faz a limpeza manual ser rápida, quando necessária.

Após obtida a base de informações sobre candidatos do TSE, essa foi cruzada com a Planilha de Partes, de forma que as linhas que se referiam a candidatos fossem preenchidas de forma automática. Esse cruzamento foi possível porque ambas as bases de dados se referiam aos candidatos pelo mesmo nome, conforme especificado acima. As linhas da Planilha que se referem a candidatos, portanto, contém as seguintes variáveis extraídas da base do TSE:

Tabela 2: Variáveis descritivas do repositório de dados eleitorais do TSE (Aplicáveis quando a parte é Candidato)

Código	Descrição da Variável	Exemplo
eleicao	Eleição na qual o processo está vinculado.	Eleições Gerais Federais de 2018; Eleições Gerais Estaduais de 2018. Eleições municipais de 2020
uf	Estado de candidatura.	SP; MG; AP.
nm_urna_cand	Nome registrado para votação do candidato.	FERNANDO HADDAD; MARINA SILVA; VALERIANO.
tipo_colig	Tipo de candidatura (existência de coligação ou não).	COLIGAÇÃO; PARTIDO ISOLADO.

²⁷ Disponível em: <http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais>. Acesso em 01/12/2019

idade_posse	Idade do candidato no período da candidatura (em anos).	35
cor_raca	Cor ou raça auto-relatada por candidatos.	PRETA; PARDA; BRANCA
id_partid	Nome do partido cujo candidato é membro.	PARTIDO DA SOCIAL DEMOCRACIA BRASILEIRA
id_colig	Nome da coligação eleitoral do candidato.	FORÇA DO RIO
comp_colig	Composição de partidos da coligação eleitoral do candidato.	PODE / PPL
carg_disp	Cargo disputado pelo candidato.	GOVERNADOR; DEPUTADO.
eleito	Quando aplicável, se candidato foi eleito.	Eleito; Não Eleito; Suplente.
total_despesas	Total de despesas de gastos de campanha em reais.	R\$ 620979.99
despesas_advoc	Total de despesas com serviços de advocacia em reais.	R\$ 100000.00
reeleicao	Status auto-relatado de reeleição do candidato.	Não; Sim; N/A

Extraídos os dados referentes aos candidatos, os pesquisadores passaram a coletar, a partir dos processos selecionados e dos URLs que constavam nesses processos, informações sobre as partes que permitissem sua classificação em categorias (para além das já coletadas automaticamente, descritas no início desta seção). Nos casos em que as informações não puderam ser obtidas por análise do conteúdo dos processos, outras fontes foram consultadas, como links presentes nas decisões e sites citados.

A classificação das partes é relevante porque permite ao usuário da Planilha de Partes (e, conseqüentemente, da base de dados completa) buscar processos em que determinadas categorias se confrontam, o que contribui para pesquisas direcionadas. Um exemplo de busca desse tipo seria selecionar processos em que o polo ativo é composto por candidatos, partidos ou coligações, e o polo passivo composto por entidades de

imprensa ou jornalistas. Esses processos podem, por exemplo, oferecer dados interessantes sobre a pressão exercida por agentes políticos sobre a imprensa em períodos eleitorais. Assim, as informações sobre Partes não extraídas da base do TSE foram as seguintes:

Tabela 3: Variáveis Descritivas das Partes e Outras

Variável	Descrição	Exemplo
parte	Nome da parte. Quando pseudoanonimizado, adotamos uma sequência de números como o identificador único daquela parte na amostra.	Procuradoria Regional Eleitoral; #0057; Facebook Serviços Online do Brasil LTD.
polo	Identificação como polo ativo, passivo ou terceiro interessado.	polo_ativo; polo_passivo.
id	Identificação amostral de qual processo participa.	825
cnj	Identificação CNJ do processo.	0600601-64.2018.6.02.0000
jurisdicao	Tribunal no qual a decisão foi emitida.	TRE-MA.
candidat	Se a parte é candidato nas Eleições de 2018	Sim; Não.
partid	Se a parte é um partido que disputou as Eleições de 2018	Sim; Não.
colig	Se a parte é uma Coligação que disputou as Eleições de 2018	Sim; Não.
politco_mand	Se a parte ocupava cargo eletivo durante as Eleições de 2018 e não foi candidato	Sim; Não.
pod_exec	Se a parte é Entidade ou Indivíduo que compõe ou representa o poder executivo	Sim; Não.
pod_judic	Se a parte é Entidade ou Indivíduo que compõe o Poder Judiciário	Sim; Não.

minis_publ	Se a parte é Entidade ou Indivíduo que compõe o Ministério Público	Sim; Não.
imprensa	Se a parte é entidade, membro de entidade ou indivíduo que divulga conteúdo na forma de notícias.	Sim; Não.
imprensa_g	Se a parte é imprensa de grande circulação, ou seja, entidade, membro de entidade ou indivíduo que divulga conteúdo na forma de notícias que teve circulação média mensal maior que 20.000 leitores no ano de 2018 ²⁸ .	Sim; Não
rede_social	Se a parte é entidade ou membro de entidade que controla uma plataforma de rede social	Sim; Não.
usuario_pag	Se a parte é entidade ou indivíduo que utiliza ou controla página ou perfil de rede social	Sim; Não.
site	Se a parte é entidade, membro de entidade ou indivíduo que utiliza ou controla página na Internet fora de plataformas de rede social	Sim; Não.
prov_con_hosp	Se a parte é provedor de conexão, hospedagem, ou de comunicação telefônica.	Sim; Não.
empresa	Se a parte é entidade ou membro de entidade que presta serviços não relacionados à imprensa.	Sim; Não.
outros	Se a parte não se adequa nas categorias anteriores.	Sim; Não.

Outras variáveis

incumb	Se o candidato controlou o cargo em disputa até 1 ano e meio antes da eleição. Alternativamente, se o candidato exerceu mandato em cargo público até 1 ano e meio antes da eleição.	Sim; Não.
genero	Gênero auto-relatado se candidato. Gênero inferido para outras categorias.	Feminino; Masculino; Entidade.

²⁸ Conforme dados de circulação de Revistas e Jornais pagos (meio digital + meio impresso) fornecidos pelo Instituto Verificador de Comunicação (IVC Brasil). Importa destacar que os dados de circulação se referem apenas aos veículos associados ao IVC Brasil. Podemos ter certeza, contudo, que esses veículos cumprem o requisito que estabelecemos para considerá-los imprensa de grande circulação. Ver: <https://ivcbrasil.org.br/#/home>. Acesso em: 09.03.2020

Os pesquisadores entenderam que não há justificativa para publicar os nomes de alguns grupos de partes, de forma que esses nomes foram substituídos por pseudônimos na Planilha de Partes. Enquanto é interessante para fins de pesquisa que sejam identificados e divulgados os agentes que representam órgãos públicos em geral e os candidatos que participaram dos processos, a publicação dos nomes de cidadãos comuns e jornalistas envolvidos não parece levar a resultados de pesquisa ou dados de interesse público que justifiquem a não-preservação de sua privacidade e intimidade.

Dessa forma, quando a parte foi pseudoanonimizada, adotamos uma sequência de números como o identificador único daquela parte na amostra. Dessa forma, por exemplo, a primeira entrada transformada em pseudônimo é a quarta parte na sequência amostral e, portanto, seu nome é #0004.

Na base das eleições de 2018, encontramos 6453 partes. Na base de 2020, encontramos 1481 partes. Elas foram divididas conforme as tabelas abaixo. Em cada uma delas, a coluna “Total de processos” se refere ao total de processos em que uma categoria foi identificada, independentemente da quantidade de partes no processo. A coluna “Total de partes únicas” se refere ao total de partes diferentes dentro de uma mesma categoria. A participação em múltiplos processos foi contabilizada apenas uma vez.

Tabela 4. Categorização das partes

Eleições gerais de 2018

Categoria	Total de processos (polo_passivo)	Total de partes únicas (polo_passivo)
Candidato	370	227
Partido	24	15
Coligação	152	75
Pol. em Mandato	23	19
Executivo	6	6
Judiciário	25	22
M. Público	3	3
Imprensa	375	382
Imprensa Grande	14	8
Rede Social	719	5
Usuário	611	979
Site	330	333
Provedor	21	27
Empresa	22	30
Outros	40	44

Categoria	Total de processos (polo_ativo)	Total de partes únicas (polo_ativo)
Candidato	1193	481
Partido	115	36
Coligação	441	118
Pol. em Mandato	3	3
Executivo	1	1
Judiciário	0	0
M. Público	22	6
Imprensa	2	4
Imprensa Grande	0	0
Rede Social	1	1
Usuário	1	1
Site	2	4
Provedor	0	0
Empresa	1	1
Outros	4	5

Categoria	Total de processos (interessados)	Total de partes únicas (interessados)
Candidato	2	2
Partido	0	0
Coligação	1	1
Pol. em Mandato	0	0
Executivo	4	4
Judiciário	0	0
M. Público	1488	25
Imprensa	2	2
Imprensa Grande	0	0
Rede Social	25	5
Usuário	0	0
Site	1	1
Provedor	1	2
Empresa	0	0
Outros	1	1

Eleições municipais de 2020

Categoria	Total de processos (polo_passivo)	Total de partes únicas (polo_passivo)
Candidato	370	227
Partido	24	15
Coligação	152	75
Pol. em Mandato	23	19
Executivo	6	6
Judiciário	25	22

M. Público	3	3
Imprensa	375	382
Imprensa Grande	14	8
Rede Social	719	5
Usuário	611	979
Site	330	333
Provedor	21	27
Empresa	22	30
Outros	40	44

Categoria	Total de processos (polo_ativo)	Total de partes únicas (polo_ativo)
Candidato	195	159
Partido	13	12
Coligação	95	85
Pol. em Mandato	8	8
Executivo	14	4
Judiciário	181	148
M. Público	5	3
Imprensa	65	74
Imprensa Grande	4	3
Rede Social	43	8
Usuário	130	143
Site	51	56
Provedor	0	0
Empresa	2	2
Outros	15	13

Categoria	Total de processos (interessados)	Total de partes únicas (interessados)
Candidato	4	4
Partido	2	2
Coligação	1	1
Pol. em Mandato	0	0
Executivo	1	1
Judiciário	0	0
M. Público	442	19
Imprensa	1	1
Imprensa Grande	0	0
Rede Social	4	1
Usuário	3	3
Site	1	1
Provedor	0	0
Empresa	0	0
Outros	1	1

5.2. A Planilha de Processos

Como já mencionado no Item 5.1, a Planilha de Processos organiza informações que se referem a cada processo incluído na base de dados. Assim, cada linha da planilha se refere a todo o conjunto de decisões de um processo, identificado já na primeira coluna por seu número CNJ. Para além dos metadados obtidos mediante extração de dados do PJe (número do processo; nome das partes; órgão julgador; município de origem; dados de movimentação processual; eleição a que o processo se refere e ementa do processo), buscamos codificar nessa tabela dados obtidos mediante uso da ferramenta de *regex* descrita no item 4 (referentes à **linguagem** utilizada pelos juízes eleitorais para se referir ao conteúdo questionado e referentes aos **dispositivos legais** mencionados nos processos) e dados coletados manualmente (referentes a ocorrência ou não de determinados **resultados** ao fim do processo).

Quanto aos dados de **linguagem**, utilizamos a ferramenta de *regex* para verificar a presença nas decisões que formam os processos de determinadas expressões utilizadas em cenários jurídicos e não-jurídicos para tratar de conteúdo desinformativo. Foram essas as expressões buscadas (com variações): “Fake News”, “Sabidamente Inverídico”, “Inverídico em sentido amplo” (ou seja, outras expressões que incluem os adjetivos “inverídico, falso” e afins), “Desinformação”, “Calúnia”, “Difamação”, “Injúria” (essas três últimas agrupadas por se referirem a crimes contra à honra), “Ofensa à honra”, “Propaganda Irregular” e “Propaganda negativa”. A inserção desses dados é conveniente pois permite que o usuário da Planilha de Processos associe o uso de determinadas expressões a outros dados presentes ou não na planilha, como taxa de sucesso de pedidos e órgão julgador. A própria taxa de presença das expressões nos processos, contudo, é um dado importante. Isso porque pode revelar a adoção, por juízes, de linguagem que não tem definição legal clara, como “fake news” e “desinformação”, ainda que presentes ferramentas jurídicas definidas em lei que poderiam ser consideradas suficientes para a construção de sua argumentação. Os dados completos, que indicam como as expressões se distribuem entre os tribunais, podem ser encontrados no **Anexo I**.

Quanto aos dados referentes aos **dispositivos legais**, objetivamos identificar a frequência em que determinados artigos de lei são mencionados nos processos coletados. Utilizamos, novamente, as expressões regulares para encontrar a maior quantidade possível de ocorrências de citação de determinados artigos, pois com elas

pudemos captar uma ampla variação de formas de escrita de uma determinada menção de artigo. Assim, quando buscamos, por exemplo, por citações do Artigo 5º, inciso IV da Constituição Federal, pudemos encontrar, com relativa facilidade, trechos que mencionam esse artigo das seguintes formas: art. 5º, inciso IV da Constituição Federal, artigo 5, IV da Constituição Federal etc.

Tendo em vista a impossibilidade de construirmos *regex* para cada dispositivo que poderia ser citado nos processos, a escolha dos dispositivos legais que seriam efetivamente buscados se deu por um critério de relevância. A partir de uma leitura preliminar de menções genéricas a dispositivos legais nos documentos coletados, identificamos 47 que consideramos relacionados ao tema de desinformação online. A lista completa dos dispositivos escolhidos, o conteúdo que justificou sua escolha e a tabela completa que descreve sua frequência nos processos está no **Anexo II**.

Quanto aos dados de **resultados dos processos**, esses foram extraídos dos documentos manualmente. Dois pesquisadores e duas pesquisadoras estagiárias foram destacados para a leitura dos documentos coletados e para preencher, com “sim”, “não”, ou “não solicitado” três colunas da Planilha.

É importante observar que o objetivo do preenchimento aqui descrito não foi o de identificar o resultado de cada um dos pedidos de cada parte, ou o de compreender em detalhes as razões e argumentos jurídicos (e não jurídicos) que levaram a esses resultados. Por uma restrição de tempo, os pesquisadores se limitaram a registrar se, ao final do processo, algum conteúdo foi removido por ordem judicial, alguma parte exerceu direito de resposta ou se alguma multa foi aplicada, de acordo com as instruções descritas abaixo. Para a coleta desses dados consideramos irrelevante, portanto, se a ocorrência ou inoocorrência de um determinado resultado se deu com ou sem julgamento de mérito²⁹. Destaca-se também que os resultados não são exclusivos, de forma que um mesmo processo pode resultar na remoção de um conteúdo, na concessão de direito de resposta e na aplicação de multa, simultaneamente, ou em qualquer outra combinação dos três.

²⁹ Isso é mais comum no âmbito do direito eleitoral, tendo em vista que, com o término das eleições, certos processos que permanecem inconclusos podem ser excluídos sem julgamento de mérito por perda de seu objeto. Não há razão, por exemplo, para se acatar o pedido de exclusão de uma publicação potencialmente prejudicial à uma campanha eleitoral se a campanha já cessou. Nesses casos, passa a ser pouco relevante a decisão sobre o conteúdo da mensagem ser, de fato, danoso.

Tabela 5: Codificação de variáveis de resultados de processo

Código da Coluna	Instruções para preenchimento
remocao	Algum conteúdo questionado foi removido? Se houve decisão que acolheu pelo menos um pedido de remoção de conteúdo ilícito e não foi revertida ou desconfirmada até o final do processo, preencher “sim”. Se o pedido de remoção de conteúdo não foi acolhido ou julgado até a extinção do processo, preencher “não”. Se nenhum pedido de remoção de conteúdo foi realizado, preencher “não solicitado”.
dir_resp	Foi concedido direito de resposta ao conteúdo questionado? Se houve decisão que acolheu pelo menos um pedido de direito de resposta à publicação de conteúdo ilícito e não foi revertida ou desconfirmada até o final do processo, preencher “sim”. Se o pedido de direito de resposta não foi acolhido ou julgado até a extinção do processo, preencher “não”. Se nenhum pedido de direito de resposta foi realizado, preencher “não solicitado”.
multa	Foi aplicada multa pela publicação do conteúdo questionado? Se houve decisão que acolheu pelo menos um pedido de aplicação de multa por publicação de conteúdo considerado ilícito e não foi revertida ou desconfirmada até o final do processo, preencher “sim”. Se o pedido de aplicação de multa não foi acolhido ou julgado até a extinção do processo, preencher “não”. Se nenhum pedido de aplicação de multa foi realizado, preencher “não solicitado”. Obs. hipóteses de aplicação de multa em caso de descumprimento de decisão são preenchidas como “não solicitado”.

A tabela completa que contém os dados resultantes do preenchimento discriminados por órgão julgador, pode ser encontrada nos Sumário de Resultados referentes a cada uma das fases da pesquisa³⁰. Para cada tipo de resultado (remoção, direito de resposta ou multa) e para cada órgão, estão apresentados o total de processos em que houve pedido (n_resultado), o número de processos em que o resultado ocorreu (taxa_resultado_#) e a porcentagem referente a esses dados (taxa_resultado_%).

5.3. As Planilhas de URLs e de Precedentes

As planilhas de URLs (o endereço em que se encontra um conteúdo on-line) e de Precedentes contêm, respectivamente, dados sobre os **URLs citados em cada processo** e sobre os **precedentes citados em cada processo**. Em ambos os casos, utilizamos

³⁰ Ambos acessíveis pelo do endereço <https://fgv.academia.edu/fgvcepi>.

ferramentas de regex para extrair de cada processo o conteúdo relevante. Além disso, extraímos também um ou dois parágrafos adjacentes ao conteúdo relevante que permitem sua contextualização no processo sem a necessidade de releitura da peça. Partimos de termos de busca genéricos associados a URLs (e.g. “www”, “http”, “.com” etc.) e a precedentes (e.g. “acórdão n°”, “representação” etc.) e, após sucessivas revisões, aprimoramos esses termos para encontrar o maior número possível de resultados relevantes, sendo excluídos os falsos positivos.

Coletamos e estruturamos os URLs porque eles permitem a localização e análise do conteúdo que foi acusado ou apontado como falso nos processos. Com acesso à planilha de URLs, um pesquisador pode, por exemplo, identificar, ao acessar esses endereços, o formato predominante do conteúdo questionado (e.g. texto, vídeo, foto, áudio etc.). Também pode verificar se o conteúdo foi de fato removido ou, nos casos que sim, se foi restabelecido após o processo. As URLs foram divididas entre genéricas (quando levam a página inicial de um website ou a um perfil em rede social) e específicas (quando levam a um conteúdo em específico dentro do website). Também identificamos redes sociais, sites da grande mídia, sites governamentais, verificadores de fatos, sites hospedados no blogspot e outros. Por vezes a URL coletada pode aparentar incompleta ou não-funcional. Recomendamos verificar a URL completa no parágrafo contextual na coluna ao lado para eventuais correções.

REFERÊNCIAS

Allcott, Hunt; Gentzkow, Matthew (2017) Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, v. 31, n. 2, pp. 211-236.

Allcott, Hunt; Gentzkow, Matthew; Yu, Chuan (2018) Trends in the Diffusion of Misinformation on Social Media. Disponível em: <[https://www.researchgate.net/publication/327717892 Trends in the Diffusion of Misinformation on Social Media](https://www.researchgate.net/publication/327717892_Trends_in_the_Diffusion_of_Misinformation_on_Social_Media)>. Acesso em: 01 de abril de 2019.

Bastos, Marcos T.; Mercea, Dan (2019) The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*, v. 37, n. 1, pp. 38-54.

Botei, Mircea (2017) Misinformation with Fake News. *Bulletin of the Transilvania University of Brasov, Series VII*, v. 10, n. 2, pp. 133-140.

Bovet, Alexandre; Makse, Hernán A. (2019) Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, v. 10, n. 7, pp. 1-14.

Bradshaw, Samantha; Howard, Philip N. (2018) The Global Organization of Social Media Disinformation Campaigns. *Journal of International Affairs*, v. 71, n. 1.5, Special Issue: Contentious Narratives: Digital Technology and the Attack on Liberal Democratic Norms, pp. 23-32.

Brummette, John; DiStaso, Marcia; Vafeiadis, Michail; Messner, Marcus (2018) Read All About It: The Politicization of "Fake News" on Twitter. *Journalism & Mass Communication Quarterly*, v. 95, n. 2, pp. 497-517.

Carrubba, Clifford J.; Gabel, Matthew; Hankla, Charles (2008) Judicial Behavior under Political Constraints: Evidence from the European Court of Justice. *The American Political Science Review*, v. 102, n. 4, pp. 435-452.

Conselho Nacional de Justiça (2018) *Justiça em Números 2018: ano-base 2017*. Brasília: CNJ.

Deakin, Simon (2018) The Use of Quantitative Methods in Labour Law Research: An Assessment and Reformulation. *Social & Legal Studies*, v. 27, n. 4, pp. 456-474.

Farkas, Johan; Schou, Jannick (2018) Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood. *Javnost: The Public*, v. 25, n. 3, pp. 298-314.

Ferrara, Emilio (2017) Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election. *First Monday*, v. 22, n. 8. Disponível em: <<https://ssrn.com/abstract=2995809>>. Acesso em: 15 de abril de 2019.

Flynn, D. J.; Nyhan, Brendan; Reifler, Jason (2017) The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics. *Advances in Political Psychology*, v. 38, n. 1, pp. 127-150.

Gabel, Matthew J.; Carrubba, Clifford J.; Ainsley, Caitlin; Beaudette, Donald M. (2012) Of Courts and Commerce. *The Journal of Politics*, v. 74, n. 4, pp. 1125-1137.

Galloway, Kate (2017) Big Data: A case study of disruption and government power. *Alternative Law Journal*, v. 42, n. 2, pp. 89-95.

Hall, Matthew E. K.; Windett, Jason Harold (2013) New Data on State Supreme Court Cases. *State Politics & Policy Quarterly*, v. 13, n. 4, pp. 427-445.

Hirst, Martin (2017) Towards a political economy of fake news. *The Political Economy of Communication*, v. 5, n. 2, pp. 82-94.

Iaryczower, Matías; Spiller, Pablo T.; Tommasi, Mariano (2002) Judicial Independence in Unstable Environments, Argentina 1935-1998. *American Journal of Political Science*, v. 46, n. 4, pp. 699-716.

Jang, S. Mo; Geng, Tieming; Li, Jo-Yun Queenie; Xia, Ruofan; Huang, Chin-Tser; Kim, Hwalbin; Tang, Jijun (2018) A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behavior*, v. 84, pp. 103-113.

Jang, S. Mo; Kim, Joon K. (2018) Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior*, v. 80, pp. 295-302.

Jensen, Michael (2018) Russian Trolls and Fake News: Information or Identity Logics? *Journal of International Affairs*, v. 71, n. 1.5, Special Issue: Contentious Narratives: Digital Technology and the Attack on Liberal Democratic Norms, pp. 115-124.

Kapiszewski, Diana; Taylor, Matthew M. (2008) Doing Courts Justice? Studying Judicial Politics in Latin America. *Perspectives on Politics*, v. 6, n. 4, pp. 741-767.

Levinger, Matthew (2018) Master Narratives of Disinformation Campaigns. *Journal of International Affairs*, v. 71, n. 1.5, Special Issue: Contentious Narratives: Digital Technology and the Attack on Liberal Democratic Norms, pp. 125-134.

Lewandowsky, Stephan; Ecker, Ullrich K.H.; Cook, John (2017) Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition*, v. 6, pp. 353-369.

Lewandowsky, Stephan; Gignac, Gilles E.; Oberauer, Klaus (2013) The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science. *PLoS ONE*, v. 8, n. 10, e75637.

Malka, Ariel; Krosnick, Jon A.; Langer, Gary (2009) The Association of Knowledge with Concern About Global Warming: Trusted Information Sources Shape Public Thinking. *Risk Analysis*, v. 29, n. 5, pp. 633-647.

Newman, Nic, Fletcher, Richard, Kalogeropoulos, Antonis, Levy, David A. L., & Nielsen, Rasmus Kleis (2018) Reuters Institute Digital News Report 2018, pp. 19. Disponível em: <<http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475>>. Acesso em: 08 de dezembro de 2018.

Oliver, J. Eric; Wood, Thomas J. (2014) Conspiracy Theories and the Paranoid Style(s) of Mass Opinion. *American Journal of Political Science*, v. 58, n. 4, pp. 952-966.

Pérez-Liñán, Aníbal; Ames, Barry; Seligson, Mitchell A. (2006) Strategy, Careers, and Judicial Decisions: Lessons from the Bolivian Courts. *The Journal of Politics*, v. 68, n. 2, pp. 284-295.

Ponniah, Kevin (2019) WhatsApp: The 'black hole' of fake news in India's election. BBC News. Disponível em: <<https://www.bbc.com/news/world-asia-india-47797151>>. Acesso em: 15 de abril de 2019.

Rais, Diogo; Falcão, Daniel; Giacchetta, André Zonaro; Meneguetti, Pamela. (2018) Direito Eleitoral Digital. Thomson Reuters Brasil.

Rais, Diogo (2018) Fake News e Eleições. In: *Fake News: a conexão entre a desinformação e o direito*. Revista dos Tribunais, pp.103-130.

Ribeiro, Márcio Moretto; Ortellado, Pablo (2018) O que são e como lidar com as notícias falsas. *SUR: Revista Internacional de Direitos Humanos*, v. 15, n. 27, pp. 71-83.

Ruedinger, Marco A. (2018) Desinformação na era digital: amplificações e panorama das Eleições 2018. Rio de Janeiro: FGV DAPP. (Policy Paper Sala de Democracia Digital #Observa2018, 2).

Schäfer, Fabian; Evert, Stefan; Heinrich, Philipp (2017) Japan's 2014 General Election: Political Bots, Right-Wing Internet Activism, and Prime Minister Shinzō Abe's Hidden Nationalist Agenda. *Big Data*, v. 5, n. 4, pp. 294-309.

Shao, Chengcheng; Ciampaglia, Giovanni Luca; Varol, Onur; Flammini, Alessandro; Menezes, Filippi (2017) The spread of fake news by social bots. Disponível em: <https://www.researchgate.net/publication/318671211_The_spread_of_fake_news_by_social_bots>. Acesso em: 01 de abril de 2019.

Shao, Chengcheng; Hui, Pik-Mai; Wang, Lei; Jiang, Xinwen; Flammini, Alessandro; Menczer, Filippo; Ciampaglia, Giovanni Luca (2018) Anatomy of an online misinformation network. *PLoS ONE*, v. 13, n. 4, e0196087.

Smith, Joseph L.; Todd, James A. (2015) Rules, Standards, and Lower Court Decisions. *Journal of Law and Courts*, v. 3, n. 2, pp. 257-275.

Svantesson, Dan Jerker; Caenegem, William van (2017) Is it time for an offence of 'dishonest algorithmic manipulation for electoral gain'? *Alternative Law Journal*, v. 42, n. 3, pp. 184-189.

Taylor, Matthew M. (2006) Veto and Voice in the Courts: Policy Implications of Institutional Design in the Brazilian Judiciary. *Comparative Politics*, v. 38, n. 3, pp. 337-355.

Tucker, Joshua Aaron; Guess, Andrew; Barbera, Pablo; Vaccari, Cristian; Siegel, Alexandra; Sanovich, Sergey; Stukal, Denis; Nyhan, Brendan (2018) Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. Disponível em: <<https://ssrn.com/abstract=3144139>>. Acesso em: 30 de outubro de 2019.

Vargo, Chris J; Guo, Lei; Amazeen, Michelle A (2018) The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New media & society*, v. 20, n. 5, pp. 2028-2049.

ANEXOS

ANEXO I - Dispositivos em Processos

Dispositivos Buscados

Tipo normativo	Nº	Ano	Órgão	Nome	Artigo	Inciso	Conteúdo
CF		1988		CF	5º	IV	Liberdade de expressão e vedação ao anonimato.
CF		1988		CF	5º	X	Direito à indenização por violação da intimidade, vida privada, honra e imagem.
CF		1988		CF	220		Não haverá restrição à manifestação de pensamento, expressão e informação.
CF		1988		CF	221		Princípios que pautam as emissoras de rádio e televisão.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	54		Restringe o conteúdo das propagandas de rádio e televisão no horário eleitoral gratuito.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	57-D		Liberdade de expressão e vedação ao anonimato na propaganda na Internet.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	57-F		Responsabilização dos intermediários ao não acatarem ordem de remoção.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	57-C		Veda propaganda paga na Internet, com exceção do impulsionamento.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	57-B		Estabelece as formas pela qual a propaganda eleitoral pode ser realizada na Internet.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	57-A		Determina o início do período de propaganda eleitoral permitida na Internet.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	57-J		Atribui ao TSE a competência de regulamentar a lei de acordo com a tecnologia.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	57-I		Traz a possibilidade de requerimentos de suspensão por 24h de acesso ao conteúdo ilícito.
Lei	9.504	1997		Lei das	58		Assegura e regula o direito de

Tipo normativo	Nº	Ano	Órgão	Nome	Artigo	Inciso	Conteúdo
				Eleições, Lei Eleitoral			resposta nas eleições.
Lei	9.504	1997		Lei das Eleições, Lei Eleitoral	58-A		Atribui prioridade na tramitação para pedidos de direito de resposta.
Lei	12.965	2014		Marco Civil da Internet, MCI	3º		Princípios que disciplinam o uso da Internet no Brasil.
Lei	12.965	2014		Marco Civil da Internet, MCI	10		Regulamente a guarda e a disponibilização de dados pessoais, comunicações e registros de conexão.
Lei	12.965	2014		Marco Civil da Internet, MCI	19		Limita a responsabilidade dos provedores apenas a quando não acatam ordem judicial.
Lei	12.965	2014		Marco Civil da Internet, MCI	15		Regulamenta a guarda de dados por provedores de aplicação.
Lei	12.965	2014		Marco Civil da Internet, MCI	22		Regulamenta a requisição judicial de registros de conexão e de acesso.
Lei	12.965	2014		Marco Civil da Internet, MCI	23		Resguarda a intimidade na requisição judicial de registros.
Resolução	23.551	2017	TSE		33		Regras para ordens de remoção de conteúdo divulgado na Internet / menor interferência possível
Resolução	23.551	2017	TSE		34		Regras de armazenamento e disponibilização de dados por provedores
Resolução	23.551	2017	TSE		6º		Proíbe a propaganda que cria na opinião pública estados mentais, emocionais ou passionais.
Resolução	23.551	2017	TSE		22		Expressão só pode ser limitada quando ocorrer ofensa à honra de terceiros ou divulgação de fatos sabidamente inverídicos.
Resolução	23.551	2017	TSE		17	X	Não será tolerada propaganda que caluniar, difamar ou injuriar pessoas, órgãos ou entidades que exerçam autoridade pública.

Tipo normativo	Nº	Ano	Órgão	Nome	Artigo	Inciso	Conteúdo
Resolução	23.551	2017	TSE		25		Liberdade de expressão e vedação ao anonimato na propaganda na Internet.
Resolução	23.551	2017	TSE		28		Exige mecanismo de descadastramento quando o candidato envia mensagem para eleitores.
Resolução	23.551	2017	TSE		84		Crime de divulgar fatos sabidamente inverídicos.
Resolução	23.551	2017	TSE		85		Calúnia Eleitoral.
Resolução	23.551	2017	TSE		86		Difamação Eleitoral.
Resolução	23.551	2017	TSE		87		Injúria Eleitoral.
Resolução	23.551	2017	TSE		83		Crime de contratar pessoas para enviar mensagens.
Resolução	23.547	2017	TSE		15		Assegura e regula o direito de resposta nas eleições.
Resolução	23610	2019	TSE		9		Trata da presunção de fidedignidade das informações na propaganda eleitoral
Decreto	8.771	2016			13		Dita os padrões de segurança mínimos para o armazenamento de dados pelos provedores.
Lei	4.737	1965		Código Eleitoral	243		Não será tolerada propaganda que caluniar, difamar ou injuriar pessoas, órgãos ou entidades que exerçam autoridade pública.
Lei	4.737	1965		Código Eleitoral	242		Proíbe a propaganda que cria na opinião pública estados mentais, emocionais ou passionais.
Lei	4.737	1965		Código Eleitoral	323		Crime eleitoral de difamação.
Lei	4.737	1965		Código Eleitoral	324		Crime eleitoral de calúnia.
Lei	4.737	1965		Código Eleitoral	325		Crime eleitoral de injúria.
Decreto-Lei	2.848	1940		Código Penal	139		Crime de Difamação.
Decreto-Lei	2.848	1940		Código Penal	140		Crime de Injúria.
Decreto-Lei	2.848	1940		Código Penal	286		Crime de incitação ao crime.
Decreto-Lei	2.848	1940		Código Penal	307		Crime de falsa identidade.
Decreto-Lei	2.848	1940		Código	154-A		Crime de invasão de dispositivo.

Tipo normativo	Nº	Ano	Órgão	Nome	Artigo	Inciso	Conteúdo
				Penal			
Decreto-Lei	2.848	1940		Código Penal	138		Crime de Calúnia.
Lei	13.105	2015		Código de Processo Civil, CPC	485		Regula a hipótese de Perda de Objeto

Frequencia dos dispositivos

Eleições gerais de 2018

Dispositivos citados	#_casos	%_casos
RES_23551_art33	642	43.03%
LE_art58	560	37.53%
CPC_art485	489	32.77%
LE_art54D	486	32.57%
LE_art54J	433	29.02%
RES_23551_art22	402	26.94%
LE_art54A	244	16.35%
CF_art5_incIV	219	14.68%
MCI_art10	216	14.48%
RES_23551_art25	179	12.00%
MCI_art22	169	11.33%
CE_art243	152	10.19%
MCI_art19	130	8.71%
CE_art242	128	8.58%
RES_23551_art34	112	7.51%
CF_art220	94	6.30%
LE_art54B	89	5.97%
LE_art54C	87	5.83%
CE_art323	72	4.83%
RES_23551_art17_incX	68	4.56%
LE_art54F	65	4.36%
CE_art325	65	4.36%
CE_art324	56	3.75%
RES_23551_art28	38	2.55%
LE_art54I	35	2.35%
RES_23551_art6	31	2.08%
CF_art5_incX	26	1.74%
MCI_art3	23	1.54%
LE_art54	18	1.21%
MCI_art15	17	1.14%
LE_art58A	12	0.80%
CF_art221	11	0.74%

Dispositivos citados	#_casos	%_casos
RES_23551_art86	7	0.47%
RES_23551_art15	7	0.47%
CP_art139	6	0.40%
CP_art140	5	0.34%
CP_art138	4	0.27%
RES_23551_art84	4	0.27%
DEC_8771_art15	4	0.27%
RES_23551_art85	3	0.20%
RES_23551_art83	2	0.13%
CP_art286	2	0.13%
CP_art154A	2	0.13%
MCI_art23	1	0.07%
RES_23551_art87	1	0.07%
CP_art307	1	0.07%
Total de casos	1492	100.00%

Eleições municipais de 2020

Dispositivos citados	#_casos	%_casos
CPC_art485	163	36,71%
LE_art58	132	29,73%
RES_23610_art27	79	17,79%
RES_23610_art38	72	16,22%
LE_art57D	63	14,19%
CE_art243	61	13,74%
LE_art57J	37	8,33%
CF_art5_incIV	34	7,66%
CF_art5_incX	23	5,18%
LE_art57C	23	5,18%
RES_23610_art30	22	4,95%
CF_art220	19	4,28%
CE_art242	19	4,28%
LE_art57B	17	3,83%
RES_23610_art10	16	3,60%
MCI_art19	15	3,38%
RES_23551_art33	13	2,93%
RES_23610_art22_incX	12	2,70%
RES_23610_art9	11	2,48%
LE_art57A	8	1,80%
CE_art325	7	1,58%
CE_art324	6	1,35%
LE_art58A	5	1,13%
RES_23610_art33	5	1,13%

Dispositivos citados	#_casos	%_casos
LE_art57F	4	0,90%
CE_art323	4	0,90%
CP_art139	4	0,90%
CP_art286	4	0,90%
LE_art54	3	0,68%
LE_art57I	3	0,68%
MCI_art23	3	0,68%
CP_art140	3	0,68%
MCI_art10	2	0,45%
MCI_art22	2	0,45%
CP_art138	2	0,45%
RES_23610_art39	2	0,45%
MCI_art15	1	0,23%
RES_23551_art6	1	0,23%
RES_23551_art22	1	0,23%
RES_23551_art28	1	0,23%
RES_23610_art90	1	0,23%
RES_23610_art91	1	0,23%
RES_23610_art89	1	0,23%
CPC_art485	163	36,71%
LE_art58	132	29,73%
RES_23610_art27	79	17,79%
Total de casos	444	100.00%

