

PUCRS

ESCOLA DE COMUNICAÇÃO, ARTES E DESIGN - FAMECOS  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO SOCIAL  
MESTRADO EM COMUNICAÇÃO SOCIAL

LORENZO POOCH LEUCK

**A ESFERA PÚBLICA NA ERA DA INFORMAÇÃO: ESTUDANDO A CÂMARA DOS  
DEPUTADOS A PARTIR DA ATUAÇÃO PARLAMENTAR E DO USO POLÍTICO DO TWITTER**

Porto Alegre  
2022

PÓS-GRADUAÇÃO - *STRICTO SENSU*



Pontifícia Universidade Católica  
do Rio Grande do Sul

LORENZO POOCH LEUCK

**A ESFERA PÚBLICA NA ERA DA INFORMAÇÃO: ESTUDANDO A CÂMARA  
DOS DEPUTADOS A PARTIR DA ATUAÇÃO PARLAMENTAR E DO USO  
POLÍTICO DO TWITTER**

Dissertação apresentada como requisito para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Comunicação Social da Escola de Comunicação, Artes e Design da Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Prof. Dr. Roberto Tietzmann

Porto Alegre  
2022

## **AGRADECIMENTOS**

Eu agradeço a minha família por ter me dado as condições de estudar tudo o que estudei. Ao meu orientador, Roberto Tietzmann, por ter mantido contato e oportunizado mais uma vez o ingresso a este campo de pesquisa. A todo pessoal do DAVINT, especialmente às professoras Isabel e Milene, por mostrarem que por trás de todos aqueles rigorosos métodos científicos estavam tardes agradáveis de amizade e companheirismo. E finalmente a uma imensa rede de pessoas que nunca vou conhecer. Se não fossem numerosos e insistentes nossos professores e cientistas, a minha e tantas outras bolsas de pesquisa não teriam passado por este período, talvez o mais sombrio da história recente. A Ciência do Brasil é forte. Nosso país ainda vai ser um vetor do progresso social e tecnológico.

## RESUMO

Esta dissertação compara o trabalho de deputados federais à forma que estes se posicionam e interagem no Twitter. O *corpus* da pesquisa é formado por dados públicos de parlamentares que submeteram proposições e publicaram *tweets* no ano de 2020. Em um âmbito teórico, é estudada a integração da democracia brasileira com as novas tecnologias digitais. Métodos de revisão bibliográfica e análise documental são utilizados para abordar: política, tecnologia, internet e mídias sociais. Estes temas são contemplados a partir de sua conexão com o conceito de esfera pública, como proposto pelo filósofo Jürgen Habermas. Por um viés técnico, se representa a multidimensionalidade da atuação parlamentar a partir de técnicas de coleta, análise e visualizações de dados, desenvolvidas e documentadas majoritariamente na linguagem de programação Python. Deste modo foi destacado o nível de consistência entre temas recorrentes no âmbito parlamentar e na mídia social, e apontada a relação entre as redes de proposições legislativas e as formadas a partir de *tweets*.

**Palavras-chave:** Comunicação. Política. Mídias Sociais. Análise de Dados.

## ABSTRACT

This dissertation compares the work of federal deputies to the way they position themselves and interact on Twitter. The corpus of the research is formed by public data from parliamentarians who submitted proposals and published tweets in 2020. In a theoretical context, the integration of Brazilian democracy with new digital technologies is studied. Literature review and document analysis methods are used to address: politics, technology, internet and social media. These themes are contemplated based on their connection with the concept of the public sphere, as proposed by the philosopher Jürgen Habermas. From a technical perspective, this work represents the multidimensionality of parliamentary action based on data collection, analysis and visualization techniques, mostly developed and documented in the Python programming language. In this way, the level of consistency between recurring themes in the parliamentary sphere and in social media is highlighted, and the relationship between the networks of legislative proposals and those formed from tweets is pointed out.

**Keywords:** Communication. Politics. Social Media. Data Analysis.

## GLOSSÁRIO

Python	Linguagem de programação de fácil leitura que exige poucas linhas de código.
Pandas	Biblioteca de software na linguagem Python, voltada à manipulação e análise de dados.
Qcut	Função de <i>pandas</i> que cria intervalos regulares de tamanhos iguais com base no conjunto de dados estudado.
Agrupamento	Função <i>groupby</i> da biblioteca <i>pandas</i> que agrupa dados de acordo com categorias, para em seguida aplicar funções nelas.
Moda	Estatística que retorna os valores mais frequente de uma série de dados.
Perfil	No capítulo “Estudo de caso” isto se refere a conjuntos de modas obtidos a partir de funções de agrupamento.
Grafo	Estrutura de dados usada para representar relações entre objetos de um conjunto.
API	Conjunto de rotinas e padrões de programação para acesso a aplicativos de software ou serviços online.
Treemap	Gráfico que fornece uma visão hierárquica de um conjunto de dados, facilitando a descoberta de padrões.
Dashboard	Painel que fornece visualizações rápidas dos principais indicadores do objetivo de pesquisa especificado.
Opensource	Licença de direito autoral que fornece o direito de estudar, modificar e distribuir software de graça para qualquer um e para qualquer finalidade.
Expressão Regular	Método que identifica cadeias de caracteres a partir de padrões.
Notebook	Formato de arquivo que facilita a escrita e execução de códigos. Atualmente serviços como o <i>Google Collab</i> permitem rodar notebooks “na nuvem”, desonerando a necessidade de altos recursos computacionais para trabalhar com análise de dados.

## LISTA DE FIGURAS

Figura 1 – Painel de resultados legislativos da Câmara dos Deputados.....	55
Figura 2 - Painel de orçamento da Câmara dos Deputados.....	56
Figura 3 - Painel de Recursos Humanos e Concursos da Câmara dos Deputados ..	56
Figura 4 – Painel de Classificação de Deputados do Atlas Político .....	57
Figura 5 – Painel de Financiamento de Campanha do Atlas Político .....	58
Figura 6 – Painel de Espectro Ideológico do Atlas Político .....	58
Figura 7 - Conexões entre as bases de dados.....	66
Figura 8 – Quantidade de Deputados Federais por Estado .....	77
Figura 9– Quantidade de Deputados Federais com proposições e tweets em 2020	78
Figura 10 – Proporção de Deputados Federais com proposições e tweets em 2020	79
Figura 11 - Índices de consistência temática por estado.....	82
Figura 12 - Quantidade de Tweets por Estado.....	83
Figura 13 – Temas e palavras-chave mais recorrentes nas proposições – Geral.....	84
Figura 14 – Publicações mais compartilhadas no Twitter – Geral.....	85
Figura 15 – Usuários, hashtags e palavras mais frequentes no Twitter – Geral .....	86
Figura 16 - Temas mais recorrentes nas proposições - Rio Grande do Sul.....	88
Figura 17 – Publicações mais compartilhadas no Twitter – Rio Grande do Sul .....	90
Figura 18 – Usuários e hashtags mais frequentes no Twitter - Rio Grande do Sul ...	91
Figura 19 - Nuvens de palavras sobre proposições e tweets - Recorrência Parlamentar.....	92
Figura 20 - Proposições com maiores percentuais de aprovação e rejeição .....	95
Figura 21 - Menção a proposições no Twitter- Geral .....	97
Figura 22 - Menção a proposições no Twitter- Rio Grande do Sul.....	99
Figura 23 – Grafo de coautoria por deputados – Rio Grande do Sul .....	101
Figura 24 – Rede de Tweets mencionando proposições - Geral.....	103
Figura 25 - Rede de Tweets mencionando proposições – Rio Grande do Sul.....	105

## LISTA DE TABELAS

Tabela 1 – Efetivação de Deputados Federais em 2020.....	63
Tabela 2 – Características gerais da Câmara dos Deputados .....	68
Tabela 3 – Compartimentos Numéricos .....	69
Tabela 4 – Perfil dos deputados gaúchos .....	70
Tabela 5 – Perfis por ideologia .....	72
Tabela 6 – Deputados com maior número de seguidores no Twitter - Geral .....	73
Tabela 7 – Deputados com maior número de seguidores no Twitter – Rio Grande do Sul .....	73
Tabela 8 – Comparação entre perfil geral e o de consistência temática .....	75
Tabela 9 – Exemplo de análise de consistência temática .....	80
Tabela 10 – Limitações do método de Consistência Temática .....	83
Tabela 11 – Métricas usadas na Análise de Redes Sociais .....	100

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>10</b>
<b>2 A EVOLUÇÃO DO CONCEITO DE ESFERA PÚBLICA.....</b>	<b>16</b>
2.1 A SOCIEDADE CIVIL CONECTADA SEGUNDO A PERSPECTIVA HABERMESIANA .....	27
<b>3 O IMPACTO SOCIAL DA INTERNET.....</b>	<b>33</b>
3.1 O PROCESSO DE INFORMATIZAÇÃO DO BRASIL.....	38
<b>4 MÍDIAS SOCIAIS E POLÍTICA .....</b>	<b>45</b>
4.1 O USO DO TWITTER NA CÂMARA DOS DEPUTADOS .....	49
4.2 A TRANSPARÊNCIA DE DADOS PÚBLICOS NA CÂMARA DOS DEPUTADOS .....	54
<b>5 ANALISANDO A ATUAÇÃO PARLAMENTAR E O USO POLÍTICO DO TWITTER NO ANO DE 2020 .....</b>	<b>60</b>
5.1 METODOLOGIA.....	62
5.2 PERFIS PARLAMENTARES .....	67
5.3 CONSISTÊNCIA TEMÁTICA.....	74
5.3.1 ÂMBITO GERAL .....	84
5.3.2 ÂMBITO ESTADUAL .....	88
5.4 FORMAÇÃO DE REDES.....	94
<b>CONSIDERAÇÕES FINAIS .....</b>	<b>107</b>
<b>REFERÊNCIAS.....</b>	<b>113</b>
<b>APÊNDICE A – Notebook de coleta e processamento de dados .....</b>	<b>121</b>
<b>APÊNDICE B – Notebook de análise e visualização de dados.....</b>	<b>131</b>

## 1 INTRODUÇÃO

De acordo com o relatório mundial do *We are Social* (2021), 4.20 bilhões de pessoas usam mídias sociais atualmente. No Brasil, o número chega a 150 milhões, alcançando 70,30% da população (DIGITAL IN BRAZIL, 2021). Araújo (2017) põe a situação em termos práticos: hoje é inimaginável ler notícias, interagir com os outros e até mesmo tomar decisões sem estar conectado a estas redes. A predominância deste meio de comunicação na política é muito discutida e passível de diversas interpretações.

Bessa et al. (2016) examinam o panorama dos movimentos sociais na era digital. Para os autores, o surgimento da internet ampliou o potencial de participação cidadã. As novas tecnologias de informação e comunicação, principalmente as mídias sociais, seriam forças de democratização nesse cenário. Castells (2013) traz este tom otimista em *Redes de Indignação e Esperança*, onde argumenta que a autocomunicação de massa constrói a autonomia do ator social, ocasionando uma “relação de amor e ódio” entre elites políticas e econômicas com a internet.

O desafio de integrar essa “inteligência coletiva conectada” em uma nova forma de governo onde os cidadãos tenham mais poder de decisão seria central a teoria política conhecida como “democracia digital” (BESSA; DE AZEVEDO; DE SOUSA, 2016). Da Silva Resende et al (2012) trazem um estudo de caso do campo ao avaliar o programa “Gabinete Digital”, do governo estadual do Rio Grande do Sul. O serviço, lançado em 2011, abriu um canal de comunicação popular *online*, a fim de facilitar o acesso a informações, e incluir a opinião dos cidadãos na política. Para Jerônimo (2016) e Wu (2018), essa iniciativa almejava uma nova cultura de gestão pública, promovendo a colaboração entre governo e sociedade civil.

Em nenhum momento os autores citados acima apresentam o programa como uma solução definitiva à democracia digital. Pelo contrário, muitas vezes apontam que o desenvolvimento do conceito é um processo em aberto, que deve adaptar-se constantemente, acompanhando mudanças sociais e novas tecnologias de comunicação. A questão é que não existe um consenso na classe política a respeito da importância destes mecanismos digitais. 4 anos após o lançamento do “Gabinete Digital”, em um período de mudança de gestão, o programa foi cancelado (SUL 21, 2015).

É possível que a descontinuidade de projetos como este seja mais do que um “contratempo”. A própria ideia da democracia digital como algo factível, que pode ser alcançado coletivamente, tem sido questionada. Para Marazzi (1997), a entrada da comunicação na esfera produtiva possibilitou a representação irrestrita de interesses individuais, sem categoria, classe ou grupo social. E este movimento, ao invés de aumentar a participação popular na política, a colocou em xeque.

Posicionamentos como este ganham força à medida que se toma conhecimento dos novos sistemas sociotécnicos. No início do século XXI, redes globais de competitividade econômica, sem qualquer vínculo com instituições de Estado ou organizações de sociedade civil, enfraqueceram mecanismos de controle social (CASTELLS, 1999b). Para os autores mais citados desta obra, a desconfiança nas instituições e a deslegitimação da representação política despontam uma crise intrínseca à estrutura do capitalismo avançado, e um iminente fim do interesse comum (CASTELLS 2018; HABERMAS apud LUBENOW, 2015).

Este não é um diagnóstico isolado. Sérgio Abranches (2018) aponta que acusações de corrupção, pedidos de impeachment e conflitos entre os poderes são constantes no Brasil desde o período da redemocratização. E isso não seria apenas uma particularidade ou infortúnio da realidade social brasileira, visto que a “a crise de representação é global” (Abranches, 2018, p.6).

No Brasil, as eleições presidenciais de 2018 colocaram as mídias sociais em um novo patamar. Não só o presidente Jair Bolsonaro se elegeu a partir de uma campanha majoritariamente direcionada a estas plataformas, como centenas de deputados federais fizeram o mesmo (DA SILVA; KERBAUY, 2020). Saber quem são estes políticos, e se há relação entre o trabalho da Câmara com suas declarações nas mídias sociais foi a motivação principal deste trabalho, que compara a atividade parlamentar com o modo que os deputados federais se posicionam e interagem no Twitter. Deste modo, questionamos “quão consistentes são os discursos dos deputados federais no Twitter, se comparados às suas proposições parlamentares?”. Este problema de pesquisa é respondido por dois caminhos, um técnico e um teórico.

No âmbito teórico, é estudada a integração da democracia brasileira com as novas tecnologias digitais. A literatura sobre a informatização da sociedade é extensa. Há duas décadas Castells (1999a) previa que a internet iria integrar todas as formas de expressão social. Segundo Malini (2007), a comunicação mediada por computador ganhou centralidade no cotidiano ao possibilitar que trabalho, comércio,

entretenimento e política fossem realizados através de um contato imersivo com o indivíduo. O resultado disso mudou a mídia drasticamente. Silva (2012a) aponta que boa parte do consumo midiático atual é formado por dados opinativos e comportamentais.

Estas considerações dimensionam o problema teórico desta obra. A estratégia escolhida para remetê-lo consiste em duas metodologias, e é dividida em três capítulos. Métodos de revisão bibliográfica e análise documental serão utilizados para abordar: esfera pública, política, tecnologia, internet e mídias sociais. Este conjunto de temas pode parecer exagerado ou até mesmo pretensioso. De fato, seria impossível apresentá-los em sua devida profundidade. É na conexão destes conceitos que a questão principal será contemplada.

O segundo capítulo da dissertação aborda o trabalho de Jürgen Habermas sobre a esfera pública. Além de ser o principal difusor deste conceito, e cânone nas ciências sociais aplicadas, há algumas particularidades na história de Habermas que o fazem imprescindível a estudos sobre democracia. Siebeneichler (2018a) faz uma breve trajetória da vida do autor. Por ter nascido na Alemanha, em 1929, boa parte de sua infância e juventude passou pela ascensão do regime nazista. O choque que levou ao tomar ciência do extermínio em massa de judeus durante a II Guerra Mundial o afetou profundamente, fazendo com que dedicasse sua carreira profissional e acadêmica a denúncia do autoritarismo. Em 1956, ingressou no Instituto de Pesquisa Social da Universidade de Frankfurt como bolsista de Theodor W. Adorno. Vale ressaltar que ele foi um dos responsáveis por popularizar os escritos da Escola de Frankfurt. O que o faz, de certa maneira, um dos precursores da crítica da mídia (SIEBENEICHLER, 2018b).

Mas quão relevante ele é hoje? Siebeneichler (2018c) argumenta que o pensamento habermasiano se encaixa perfeitamente no tempo da ciência, da técnica e da inteligência artificial, devido a contenciosos esforços de levar a filosofia a uma era “pós-metafísica”. Deste modo, o autor conduziu a teoria da sociedade em um caminho seguro, a partir de rigorosos métodos científicos, mas sem cair em dogmas do positivismo.

Levando em conta estas questões, segue-se para o terceiro capítulo, que apresenta a complexidade do contexto social da internet, contrapondo a visão idealizada da tecnologia a sua implementação no Brasil. Para isto, serão utilizadas a trilogia de Manuel Castells (1999), “A Era da Informação” e a obra “O Brasil: Território

e sociedade no início do século XXI" de Milton Santos e María Laura Silveira (2001). A expressão "Era da Informação", presente no título deste trabalho, faz referência à terminologia e conceitos destas obras. Lilia Schwarcz e Heloisa Starling (2015) também são citadas a fim de complementar estas leituras com uma maior compreensão sobre a história do Brasil

O quarto consiste em uma revisão de literatura sobre o uso das mídias sociais na política, primeiro em um âmbito geral e depois na Câmara dos Deputados. Aqui o referencial teórico passa por Sérgio Abranches (2018), Jairo Nicolau (2017), Bruno Carazza (2018) e Cristiano de Faria (2012).

Feito o devido aprofundamento teórico, é possível se direcionar aos objetivos específicos de pesquisa no quinto capítulo. Pelo viés técnico, esta obra representa a multidimensionalidade da atuação parlamentar a partir de técnicas de análise desenvolvidas e documentadas na linguagem de programação Python.

Faz parte do *corpus* do trabalho: dados públicos da Câmara dos Deputados, Tribunal Superior de Contas (TSE) e Twitter, relativos à deputados federais que submeteram proposições e publicaram *tweets* no ano de 2020. A partir deste material, se destaca o nível de consistência entre temas recorrentes no âmbito parlamentar e na mídia social, e é apontada a relação entre as redes de proposições legislativas e as formadas a partir de *tweets*. Resultados obtidos no âmbito geral da Câmara dos Deputados são comparados aos da bancada do Rio Grande do Sul, a fim de distinguir tendências predominantes e nuances da base de dados.

Boa parte dos métodos desta seção são da estatística básica, providos a partir da biblioteca *pandas*. Outro método importante é a Análise de Redes Sociais (ARS). Recuero et al. (2015) o define como um conjunto teórico de cunho estruturalista que descreve a relação entre nós (atores e vértices) e suas conexões (arestas). Para Smith (2015), as ferramentas que a integram estão para o mundo das conexões como a gravação de vídeos e a fotografia estão para o mundo físico - por meio delas foi liberado o registro e a compreensão de fenômenos até então inacessíveis.

O trabalho de conclusão de curso do autor desta dissertação apresentou uma análise quantitativa do uso de mídias sociais nas eleições presidenciais brasileiras de 2018. Uma série de visualizações de dados, baseadas no aporte teórico da análise de redes sociais, foram utilizadas para representar o engajamento de eleitores e as estratégias políticas nas plataformas Facebook, Instagram e Twitter

Neste ponto fica evidente que a parte técnica deste trabalho não acontece “no vácuo”. Hoje não seria possível replicar este tipo de pesquisa. Em 2016, a campanha eleitoral de Donald Trump contratou a empresa *Cambridge Analytica* (CA), que utilizou dados de 84 milhões de usuários do *Facebook* de forma ilegal. Christopher Wiley, ex-analista de dados da CA, depôs ao Congresso Americano que esta operação desenvolveu algoritmos para estimular paranoias e preconceitos raciais em diversos grupos de eleitores (WYLIE, 2018). Segundo Magrani e Oliveira (2018), a campanha de Jair Bolsonaro em 2018 utilizou estratégia semelhante no Brasil, ao gastar 12 milhões de reais com disparos de notícias falsas no WhatsApp.

A noção de que existem dissensos irremediáveis nas mídias sociais é preocupante, uma vez que um consenso mínimo sobre fatos é necessário para discutir temas de interesse público (MAGRANI; OLIVEIRA, 2018). Contudo, o que mais inviabiliza a perspectiva democrática é o fato destas divisões, e o nível de desinformação nas redes sociais, serem ampliados de modo artificial por agentes econômicos (WU, 2018). Isto vai além das eleições. Magrani e Oliveira (2018) apontam que a elaboração de políticas públicas pode partir de vontades populares forjadas, ocasionando gastos estatais alheios a necessidades reais dos cidadãos.

A reação do Facebook a estes problemas foi restringir o acesso a dados de usuários e promover uma campanha de “humanização da rede”, demonstrando uma súbita preocupação com a saúde mental coletiva (TARNOFF; WEIGEL, 2018). Isso reduziu substancialmente as possibilidades de pesquisa com dados de mídias sociais, impossibilitando a coleta de informações em duas das maiores plataformas da atualidade (DIGITAL 2021, 2021).

Para Magrani e Oliveira (2018), a resposta a estas fraudes digitais tem que se dar de forma institucional, por meio da lei. Aqui trabalha-se com uma ideia parecida: a divulgação de dados públicos e o envolvimento dos cidadãos na política devem ser direitos assegurados transversalmente pela sociedade civil, imprensa, academia e governo. Operações de engenharia social, como a de *Cambridge Analytica*, não ocorrem pelo excesso de transparência, mas pela falta dela.

Somente uma das três mídias sociais estudadas em 2018 mantém certo nível de escrutínio público. Muito pode ser dito sobre o uso político do Twitter e sua adaptabilidade a debates públicos. Esta mídia social possui 353 milhões de usuários ativos mensalmente, sendo a 16ª mais utilizada no mundo e a 6ª no Brasil (DIGITAL

IN BRAZIL, 2021). Segundo Silva (2012a), os perfis são minimalistas, com foco na constante publicação de mensagens curtas e instantâneas. Ademais, o fluxo de *tweets* é multimídia, rico em textos, imagens, áudios e vídeos (HU et al., 2016). Estas características tornam a rede um recurso sem precedentes para cientistas sociais, políticos, profissionais de marketing e jornalistas. (LIN et al., 2013). Ainda assim, a escolha desta plataforma como objeto de estudo se deve ao fato desta continuar com uma política aberta para análise de dados, já que as únicas mudanças em sua API foram restrições a novos pesquisadores, requerendo autorização de funcionários da empresa para evitar fraudes.

As técnicas de análise de dados do Twitter foram desenvolvidas no grupo de pesquisa multidisciplinar *Data Visualization and Interaction Lab* (DAVINT) da PUCRS. O repositório *Tweet Utils*<sup>1</sup> é a ferramenta principal, que faz a coleta dos dados e gera relatórios sobre as métricas de engajamento. Este projeto remonta anos de trabalho do grupo, sistematizados pelo pesquisador Pedro Sanvido. Neste contexto, a maior contribuição do autor desta dissertação foi outro módulo de análise chamado *Twitter Network Graphs*, que gera arquivos e imagens das redes de compartilhamentos, menções, hashtags e palavras mais utilizadas na plataforma.

Todos estes códigos estão abertos ao público. Isso faz parte da cultura *Open-source*. Seria impossível responder os questionamentos e objetivos propostos aqui sem o manejo de referenciais teóricos e outras formas tradicionais de pesquisa. O mesmo pode ser dito a combinação de múltiplas ferramentas digitais, como o Python. Estas tecnologias não são produtos finais criados exclusivamente por indivíduos. É mais adequado pensá-las como construções coletivas, em constante evolução e desdobramento por redes de desenvolvedores.

Inúmeras contribuições deste cenário, na forma de frameworks, módulos, e repositórios, possibilitaram o protótipo de novas técnicas de análise e visualização de dados. Toda etapa deste processo foi documentada a fim de garantir a reprodutibilidade dos resultados obtidos e a apropriação de suas técnicas em outros contextos<sup>2</sup>. O *notebook* de coleta e processamento de dados é apresentado aqui no apêndice A. Já o de análise e visualização de dados, no apêndice B. Esta medida visa ampliar a análise científica de fenômenos comunicacionais e promover transparência a informações de utilidade pública.

---

<sup>1</sup> Disponível em: <https://github.com/DAVINTLAB/TweetUtils>

<sup>2</sup> Disponível em: <https://github.com/lorenzo-leuck/esfera-publica-na-era-da-informacao>

## 2 A EVOLUÇÃO DO CONCEITO DE ESFERA PÚBLICA

Além de ser um lugar comum na Comunicação, o conceito de esfera pública é uma referência em todas as sociedades democráticas (OLIVEIRA; CANÇADO; PEREIRA, 2010). Normalmente, este se refere à espaços de discussões, formais ou informais, que fazem intermédio entre a sociedade civil e o estado. Ainda que esta aproximação esteja correta, empregá-la à produção de saber científico sem considerar suas origens e significados particulares é um equívoco.

Por este motivo, a tese de doutorado de Jurgen Habermas (1989), “Mudança estrutural da esfera pública: Investigações sobre uma categoria da sociedade burguesa”, será utilizada a fim de apresentar a evolução social, política e midiática do conceito. Para entender melhor este último aspecto, será abordado o trabalho de Melvin DeFleur e Sandra Ball-Rockeach (1993) sobre os meios de comunicação de massa. “A teoria da ação comunicativa” de Habermas (1984; 2015) também será apresentada aqui, junto ao trabalho interpretativo de Rüdiger (2009), para situar a grande mudança de perspectiva do autor após a publicação de sua tese. Além disso, interpela-se uma breve revisão bibliográfica ao longo destas leituras, a fim de auxiliar a compreensão do tema.

Habermas (1989) faz uma análise sociológica e histórica da esfera pública e sua decorrente transformação no estado de bem estar social. Segundo o autor, este conceito remonta sistemas de participação política da Grécia Antiga. Contudo, durante séculos a esfera pública não teve qualquer aplicação técnica ou jurídica na sociedade. Há dois momentos cruciais depois deste hiato: a esfera pública da baixa idade média, imersa nas tensões culturais do surgimento do capitalismo, e a esfera pública liberal, institucionalizada mediante a ascensão da burguesia (OLIVEIRA; CANÇADO; PEREIRA, 2010; SIEBENEICHLER, 2018a).

Esta transição é descrita em detalhes por Habermas (1989). Segundo o autor, no início do século XVI, a expansão do comércio exigia uma série de comportamentos de orientação racional, baseada em informações frequentes e exatas sobre o mundo. A partir dessa necessidade, surgem as primeiras rotas de correio e, quase simultaneamente, a imprensa e as bolsas de valores. As instituições acima foram como mecanismos de defesa aos interesses da burguesia.

O sistema mercantilista da época não favorecia o estabelecimento de empresas privadas dirigidas de maneira capitalista. Na verdade, O consumo das

grandes cidades era constantemente afetado por seus regulamentos. “Quando havia escassez de trigo, o consumo de pão nas noites de sexta-feira era proibido por decreto oficial” (HABERMAS, 1989, p.24). Assim, Impostos, taxas e intervenções do governo no domínio privado constituíam o alvo de uma esfera pública em desenvolvimento.

Segundo DeFleur e Ball-Rockeach (1993), neste ponto as elites educadas começaram a usar a tecnologia da impressão para distribuir os primeiros periódicos. Habermas (1989) conta que estes continham notícias atualizadas sobre guerras, colheitas e comércio exterior, e eram enviados para seletos grupos de comerciantes.

A politização da esfera pública se intensificou através destes circuitos literários e fóruns de discussão institucionalizados (DA SILVA, 2001; OLIVEIRA; CANÇADO; PEREIRA, 2010). Conforme aumentava a insatisfação com as administrações estatais, escritores passaram a informar “verdades úteis” a seus públicos, contrapondo nesse processo o monopólio da igreja e do estado sob a interpretação da filosofia, literatura e arte (DA SILVA, 2001; HABERMAS, 1989).

Os governos locais, conscientes dos riscos que corriam, emitiram proclamações sobre os perigos de discussões em locais públicos e nos periódicos. Como foi o caso da Inglaterra, em que o autor se foca.

Os homens assumiram para si mesmos uma liberdade, não apenas nos cafés, mas em outros lugares e reuniões, públicas e privadas, para censurar e difamar os procedimentos do Estado, falando mal das coisas que eles não entendem, e se esforçando para criar e nutrir um ciúme e uma insatisfação universal nas mentes de todos os bons súditos de Sua Majestade.

(HABERMAS, 1989, p.59)

Habermas (1989) destaca que o objetivo dos burgueses não estava somente em assumir o papel de autoridade, mas sim reivindicá-la ao debate público, regulamentando o estado de direito. Dessa maneira, é concebida a ideia de tornar racional o poder político, o que resultaria em uma nova "ordem natural" para ser reproduzida socialmente. Sem um estado moderno, estruturado e autônomo, não seria possível chegar a este nível de organização. Oliveira et al. (2010) afirmam que somente após sua formação que as categorias de público e privado passaram a ter uma efetiva aplicação processual jurídica (OLIVEIRA; CANÇADO; PEREIRA, 2010). Em sua tese, Habermas mostra como a esfera pública mudou o sistema político e administrativo da época. O autor faz isso através de uma pesquisa empírica (de

caráter sociológico e histórico, como dito anteriormente), mas também a incluindo em uma teoria ampla da democracia (LUBENOW, 2015).

Nesta abordagem, destaca-se o aspecto multidimensional do conceito, que conecta cultura, economia e política, criando assim diversas sobreposições, como esferas sociopolíticas, socioeconômicas, socioculturais e assim por diante (FUCHS, 2016). A esfera pública seria como uma “caixa de ressonância dotada de um sistema de sensores sensíveis ao âmbito de toda a sociedade” (LUBENOW, 2015, p.6). Seguindo esta analogia, a sua função seria captar problemas sociais que ressoam nas esferas privadas, filtrando e sintetizando os temas mais importantes, a fim de transmiti-los institucionalmente no sistema político (JERÔNIMO, 2016; LUBENOW, 2015).

A imprensa do século XVIII teve um papel fundamental neste processo. A partir do declínio do regime feudal, e surgimento de novos conceitos como liberdade de expressão e direito ao voto, a comunicação de massa se viabiliza (DEFLEUR; BALL-ROCKEACH, 1993). O que trouxe, segundo Habermas, uma série de mudanças significativas à sociedade ocidental.

Intelectuais desta época acreditavam que quando todas as pessoas virassem estudiosos “cujos escritos falam ao mundo”, haveria o domínio da “preocupação comum” (HABERMAS, 1989). E é esse “mundo” que designa o escopo da esfera pública. Este senso de cosmopolitismo e progresso foi articulado por Immanuel Kant (apud HABERMAS, 1989), para quem o “mundo” consistia na comunicação de seres racionais. Apesar deste entendimento ter sido basilar aos princípios do iluminismo, ele se mostrou facilmente refutável. Hegel (apud HABERMAS, 1989) contesta a extensão desse mundo ao apontar uma profunda cisão na sociedade civil, proveniente da desigualdade de habilidades intelectuais e até mesmo morais.

Fuchs (2016) alerta sobre leituras desatentas de “Mudança Estrutural da Esfera Pública” que criticam a idealização do conceito. Em nenhum momento Habermas afirma que a Esfera Pública existe em todo lugar, ou que faz uma representação perfeita dos interesses populares. Pelo contrário, constantemente aponta sua imanência as falhas da sociedade, instigando que é preciso superar as incongruências entre a teoria e a prática.

A tese inclusive relata a exclusão de mulheres neste âmbito, apesar de participarem mais ativamente dos círculos literários que os chefes de família burgueses (HABERMAS, 1989). Vale complementar este adendo, lembrando que ao constituir-se apenas de homens brancos, letrados e ricos, também foram silenciadas

as reivindicações da população negra e indígena, explorada e desumanizada por este grupo social durante séculos (FUCHS, 2016).

Marx (apud HABERMAS, 1989) rejeita qualquer possibilidade de emancipação desse meio, que segundo ele promove uma falsa consciência - a favor dos interesses burgueses e contra os da classe trabalhadora. Entre as contradições da esfera pública do século XIX estariam as oportunidades de acesso desiguais e a legitimação de interesses privados, apresentados como universais (DA SILVA, 2001). Para Marx (apud DA SILVA, 2001), uma esfera pública radicalmente democrática teria deliberação sobre todos os aspectos da coletividade, sendo a condição necessária para isso a socialização dos meios de produção. Somente assim o conceito poderia alcançar seu objetivo real, que seria “a racionalização da dominação política, e a consequente dissolução do poder político em poder público” (ibid, 2001, p.120).

Críticas como essas expuseram a visão idealizada da esfera pública nas sociedades europeias. Segundo Habermas (1989), a primeira vez que esta se integrou completamente ao campo político foi na Grã Bretanha, na virada do século XVIII. Então, já não era mais considerada como uma base de unidade e verdade, mas como um conjunto de opiniões da sociedade civil, constitutiva a base do governo. Nesse âmbito, só eram considerados aqueles que subissem "ao nível de participação em questões de estado". Entre os requisitos necessários estava a “divisão do homem entre a pessoa pública e a pessoa privada”. Isto é, interesses particulares não seriam pesados na legislação, pois esta seria resultado de acordo racional em busca do bem comum.

Isto aproximou a prática às concepções originais do conceito, supostamente livre da censura do estado e de controles particulares (FUCHS, 2016). Habermas (1989) conta que só no ano de 1803 o parlamento britânico liberou acesso a jornalistas, que passaram quase um século entrando ilegalmente na câmara dos comuns. A partir daí, a publicidade das informações se tornaria um princípio organizacional dos órgãos do governo, assegurando a conexão dos eleitores com seus delegados.

Simultaneamente, discussões sobre política viravam um hábito entre todas as classes da sociedade. As “pessoas comuns” já não dependiam da igreja, do estado, ou de intelectuais renomados para formar suas opiniões, pois surgiam pensamentos e discursos que se dirigiam diretamente a elas. Neste contexto, a legitimidade do

sistema político era assegurada por uma rede de pontos de vista e opiniões, transmitidos sob argumentações racionais (WU, 2018).

No século XX, se tornou uma trivialidade dizer que a opinião pública dominava o mundo da política. Os meios de comunicação de massa atingiram um alcance e eficácia outrora inimagináveis, porém não estabeleceram garantias quanto ao desenvolvimento da democracia (HABERMAS, 1989). Habermas (1989) conta que a difusão de ideias passou a acontecer mais por pressões coletivas do que pelo uso da razão. Nos Estados Unidos, epicentro desse fenômeno, o negócio de fornecer opiniões prontas prosperava.

DeFleur e Ball-Rockeach (1993) afirmam que a invenção do rádio, da televisão e do filme geraram pouca participação política, se comparados ao da imprensa nascente. Na verdade, a proliferação destes meios pode muito bem ter causado o efeito oposto. A grande mídia detinha uma força de persuasão incomparável. Em seus primórdios foi utilizada na segunda guerra mundial, sendo encarada como uma ferramenta capaz de “moldar a opinião pública e inclinar as massas para quase qualquer ponto de vista desejado pelo comunicador” (DEFLEUR; BALL-ROCKEACH, 1993, p.98).

Para Habermas (1989), é nesse momento que os fundamentos sociais da esfera pública burguesa começam a entrar em colapso - enquanto seu alcance se expandia consideravelmente, sua função se tornava progressivamente insignificante. O direito à participação civil e à publicidade das informações não conseguiu conter o avanço dos interesses particulares na política.

Problemas estruturais no âmbito discursivo, e a falta de condições institucionais adequadas, a levaram de “um espaço de discussão e de exercício da crítica a uma esfera dominada pelos meios de comunicação de massa infiltrada pelo poder” (LUBENOW, 2015, p.2). Além disso, a ideia de uma vida estritamente privada se esvaziava. No momento em que as pessoas se retiravam de papéis socialmente controlados no mundo do trabalho, para os puramente "pessoais", entregavam boa parte do seu tempo ao uso não comprometido do lazer (HABERMAS, 1989).

Sendo que as principais formas de lazer da época, como televisão e rádio, se resumiam a um consumo cultural programado, sob a influência de autoridades semi públicas. Esse tempo livre se tornou parte do ciclo de produção e consumo da sociedade, incapaz de constituir um mundo emancipado das necessidades básicas da sobrevivência (HABERMAS, 1989). Desse modo a mídia moldava, monitorava e

modificava o significado de tudo, de produtos comerciais à orientações políticas (DEFLEUR; BALL-ROCKEACH, 1993). Aqui, a própria imprensa jornalística delega sua função crítica para se tornar um complexo de poder social (HABERMAS, 1989). Este seria o ápice de um processo em voga desde a metade do século 19: a comercialização da imprensa. A partir dele, a opinião pública, ora emergente, se decompõe em opiniões informais de cidadãos privados (FUCHS, 2016).

Habermas (1989) afirma que neste contexto a esfera pública da sociedade civil voltou a ter “características feudais”. A capacidade de alcançar as massas simulava o prestígio e autoridade de outrora. E nesse quesito, a grande mídia e o estado estavam par a par. Na democracia de massa, o processo no qual poder social se torna poder político é alvo tanto de crítica como de construção de consensos. Um exemplo disso é o período de eleições. Em tese, esse arranjo democrático requer cidadãos bem-informados e envolvidos debatendo racionalmente para decidir seus representantes. O que recebe na prática é uma ampla camada da população rígida em convicções, muitas vezes envoltas de “preconceitos profundamente enraizados referentes a experiências distantes no passado e transmitidas por gerações” (HABERMAS, 1989, p.214).

E é esse o eleitorado visado por agências de relações públicas alheias a demandas democráticas. A partir delas, candidatos lançam ofertas psicologicamente apelativas a inclinações subscientes da população, a fim de provocar reações previsíveis sem qualquer intenção de formar acordos plebiscitários. Assim, argumentos e indagações são substituídos por símbolos de identificação (HABERMAS, 1989). Essas táticas só eram possíveis devido a poderosos grupos midiáticos, que acompanhavam o comportamento das audiências a fim de segmentar campanhas de publicidade de acordo com interesses, necessidades e valores (DEFLEUR; BALL-ROCKEACH, 1993).

Neste cenário, a liberdade de expressão se mostrou insuficiente para garantia de direitos sociais, pois a formação da opinião pública estava sujeita a interesses privados que pautavam e desconsideravam temas de discussão a seu favor. Aqui a ideia de Kant (apud HABERMAS, 1989) de que o mundo seria constituído pelo diálogo de seres racionais fica ainda mais distante, pois o público já não é composto por pessoas formal e materialmente em pé de igualdade (HABERMAS, 1989).

Perlatto (2012) ressalta que a mudança estrutural da esfera pública é esta ter passado de um espaço de convencimento para um de articulação de interesses e

pressões políticas. O que é compreensível, levando consideração o aumento do público entre os séculos XVIII e XIX. Contudo a comunicação de massa do século XX levou isto a um novo patamar, transformando os cidadãos em consumidores e destruindo a divisão entre as esferas privada e pública. Isto se dá devido a “instituições privadas assumirem cada vez mais os espaços do poder público, e o Estado penetrar com maior intensidade no domínio privado” (PERLATTO, 2012).

O modo que Habermas descreve este fenômeno parte de seu meio de formação intelectual, a Escola de Frankfurt. Tradições e métodos deste grupo se encontram na tese do autor a partir de elementos como: estudo crítico da ideologia, denúncia do impacto da mídia na democracia moderna, e a noção de que o capitalismo impossibilita o desenvolvimento de uma esfera pública ideal (FUCHS, 2016; MAGRANI; OLIVEIRA, 2018).

Em “Teoria da Ação Comunicativa” há uma profunda transformação no pensamento de Habermas. Ao romper com seus posicionamentos anteriores, ele se distancia também da perspectiva frankfurtiana (RÜDIGER, 2009). Rüdiger (2009) destaca que a esfera pública foi basilar a modernidade, enquanto projeto coletivo que buscava o progresso através da ciência. Habermas a caracteriza como a “época da racionalização” que lidou com questões sociais e políticas como se fossem técnicas (VILLA, 1992). Isto resultou em uma série de patologias, como a hipertrofia da racionalidade econômica e da burocratização do sistema político (MAGRANI; OLIVEIRA, 2018).

Ao apontar a decadência da esfera pública, Habermas também assumia o fracasso da modernidade e, ainda que indiretamente, a capacidade emancipatória da razão. A mudança significativa em seu trabalho veio em parte de uma questão em aberto na tese de doutorado: a necessidade de uma comunicação intersubjetiva. Siebeneichler (2018b) explica que nesta nova perspectiva a consciência deixa de ser vista como algo privado, devido ao entendimento de que esta depende de uma rede cultural de informações, valores e pensamentos compartilhados intersubjetivamente.

Nas palavras do autor: “Jamais consegui aceitar a ideia de que a autoconsciência constitui por si mesma um fenômeno originário [...] somente nos tornamos conscientes de nós mesmos nos olhares que um ‘outro’ dirige a nós” (HABERMAS apud SIEBENEICHLER, 2018b, p.50).

A partir daí a teoria crítica da sociedade é reconstruída através da razão comunicativa - força que gera consenso na linguagem (RÜDIGER, 2009). Toda a

experiência discursiva, dos processos políticos e democráticos até a formação da vontade dos cidadãos, passa por essa concepção (JERÔNIMO, 2016). A lógica interna desta ação parte de interações livres de contextos normativos da esfera privada familiar e da esfera pública marcada pelos meios de comunicação de massa (HABERMAS, 1984).

Isso permite que Habermas reformule o diagnóstico pessimista sobre a dialética da razão, reabilitando o projeto de realização humana concebido pela modernidade (RÜDIGER, 2009). Deste modo, a modernidade não teria esgotado suas potencialidades positivas, apesar das trágicas experiências históricas do século XX. Existiria nela um potencial prático e cognitivo não explorado totalmente devido às perversidades do capitalismo (RÜDIGER, 2009).

A nova teoria de Habermas aponta que a comunicação comanda o processo dialético de evolução social. Esta, enquanto processo de entendimento recíproco entre as pessoas, continuamente transmite e renova o conhecimento comum gerado no passado (RÜDIGER, 2009). Todo este processo ocorre dentro de sua respectiva tradição cultural, independente se esta é empregada, reproduzida, criticada ou renovada. Ou seja, sujeitos sociais precisam cumprir certas regras que “não só regulam sua pertença a certos grupos como reforçam, por extensão, a coesão de toda a sociedade” (RÜDIGER, 2009, p.102). Neste contexto, as condições de validade de um argumento dependem de conhecimentos intersubjetivos. Habermas (1984) coloca nos seguintes termos:

O mundo ganha objetividade apenas por contar como um e o mesmo para uma comunidade de sujeitos falantes e atuantes. O conceito abstrato de mundo é uma condição necessária para que os sujeitos que atuam comunicativamente possam compreender entre si o que se passa no mundo ou que nele será afetado. Por meio dessa prática comunicativa, eles se asseguram ao mesmo tempo de suas vivências comuns, de um mundo da vida compartilhado intersubjetivamente.

(HABERMAS, 1984, p.62)

Segundo o autor, a suposição de um mundo comum a todos não é sustentável logicamente, porém não pode ser refutada pois parte da experiência direta dos observadores. A única “solução” estaria em desacreditar seus relatos, rebaixando-os a situações de “falsa consciência”, “paranoia” ou “preconceito” (HABERMAS, 1984, p.63). Esta abordagem coloca em questão não a intersubjetividade do mundo, mas os métodos para lhe experimentar e relatar.

O que levanta um problema. É a partir destes mecanismos de socialização que as pessoas tomam parte dos processos de compreensão mútua e formam suas identidades. Assim, pode-se dizer que a comunicação cotidiana sempre surge pré-interpretada por determinados valores e conhecimentos de fundo. O que a difere do “mundo da vida”:

Lugar transcendental no qual o falante e o ouvinte saem ao encontro; no qual podem colocar a pretensão de que suas emissões concordam com o mundo (objetivo, subjetivo e social); e em que podem criticar e exhibir os fundamentos destas pretensões de validade, resolver suas discordâncias e chegar a um acordo

(RÜDIGER, 2009, p.102)

De acordo com Habermas (1984), o consenso sempre está no horizonte do mundo da vida. Já que este teria como princípio básico comunidades de comunicação onde os participantes compartilham uma série de pressupostos sobre um único mundo, formado intersubjetivamente por indivíduos e coletividades.

Feita esta distinção, o conceito de sociedade passa a ser formado por dois domínios estruturalmente diferentes: o mundo da vida e o sistema, que inclui a função burocrática do estado, assim como a lógica do dinheiro e do poder (MAGRANI; OLIVEIRA, 2018). Lubenow (2015) rejeita a comparação desta dicotomia com a de Estado/sociedade civil. O mundo da vida seria mais do que a reprodução institucional de tradições, solidariedades e identidades, encarregada pela sociedade civil. Uma vez que envolve também uma dimensão linguístico-simbólica, manifestada por processos comunicativos de transmissão cultural e socialização.

Habermas (1984) pontua que o mundo da vida armazena o trabalho interpretativo de gerações anteriores, resultando em contrapesos conservadores a mudanças sociais. Ou seja, o processo real de alcançar o entendimento não é livre de desacordos. A renovação da tradição depende de indivíduos prontos para criticar e inovar. Os pesos desta balança mudariam a partir da descentralização de visões de mundo. Isso quer dizer que quanto mais acessível o estoque cultural de conhecimento, menor é a necessidade de cobrir a compreensão com noções previamente estabelecidas (HABERMAS, 1984).

Quando os participantes passam a criticar o próprio mundo da vida, mudam da conversação para a discussão (*Diskurs*). Neste âmbito a comunicação cotidiana é suspensa (RÜDIGER, 2009). A ação comunicativa ocorre a medida que indivíduos formam as suas identidades e entendem o sentido que imprimem às suas ações, ou

seja, no mundo da vida (PERLATTO, 2012). As pessoas abrem mão de motivos ou interesses na ação comunicativa e buscam apenas um novo consenso sobre o mundo e suas vidas (RÜDIGER, 2009).

O tema da esfera pública, anteriormente analisado por uma perspectiva histórica, passa a ser visto de modo socio teórico (LUBENOW, 2010). O motivo para isso é a crescente dificuldade desta exercer seu potencial normativo nas instituições do capitalismo avançado (LUBENOW, 2015). Nesta nova definição, há uma “zona de intercâmbio” entre mundo da vida e sistema (MAGRANI; OLIVEIRA, 2018). Há duas possibilidades neste processo de mediação:

1) A esfera pública cede a complexa rede administrativa e econômica que retoma as patologias da modernidade (JERÔNIMO, 2016). Neste ponto, a noção de “feudalização da esfera pública” é atualizada para “colonização do mundo da vida”, ainda que a essência disso continue a mesma: elementos morais e práticos saem de âmbitos privados e políticos em prol de questões técnicas, abstratas mas necessárias a manutenção dos sistemas que o integram (FUCHS, 2016). Em outras palavras, dinheiro e poder automatizam a reprodução da linguagem e se tornam imperativos funcionais da sociedade (PERLATTO, 2012).

2) A esfera pública garante a autonomia do mundo da vida frente a estes impulsos sistêmicos, protegendo a função simbólica de integração social (LUBENOW, 2015). A condição necessária para isso seria uma comunidade real de comunicação, onde todas as partes tenham condições de compreender o sentido dos argumentos e de ajuizar em definitivo a verdade (RÜDIGER, 2009). O que consiste também na razão comunicativa articular questões éticas, estéticas e de justiça para promover a tolerância mútua e a convivência em comum (PERLATTO, 2012)

Habermas (1984) frisa que este tipo de ação, voltada unicamente para alcançar o entendimento, “de forma alguma está em todos os lugares e sempre é encontrado como o caso normal na prática cotidiana” (ibid, 1984, p.199). Contudo, ao exercê-la as pessoas podem transpor coerções provenientes de processos de trabalho e de relações de poder, para uma forma de vida livre e igualitária (RÜDIGER, 2009).

Em suma, se evidenciam dois mecanismos de integração social. Os mecanismos de poder e de troca, e os de entendimento e formação de consenso (SIEBENEICHLER, 2018b). Costa (1995) destaca a ambivalência desta análise, uma vez que a esfera pública não seria mais constituída apenas por manipulação de opiniões, manobras políticas e discursos que legitimam o poder de grupos

econômicos. Nesta reavaliação, encontra-se um espaço onde fluxos de comunicação condensam tópicos de opinião pública relevantes para o conjunto da sociedade. (COSTA, 1995; RÜDIGER, 2009). Contudo, Habermas (2015) ainda vê a mídia como algo exógeno, que coordena processos de compreensão através de valores instrumentais generalizados como dinheiro e poder. Estes meios substituem o papel da linguagem do cotidiano, “libertando” a ação social da incumbência de criar consenso de valores (HABERMAS, 1984, p.431).

Neste ponto, o filósofo traça uma distinção entre formas generalizadas e meios sistêmicos de comunicação. Os meios de comunicação de massa continuam sendo vistos de forma negativa por reproduzirem o mundo da vida de modo sistematizado, degradando a formação de consenso em estereótipos para o controle social (RÜDIGER, 2009). É deste modo que vários papéis sociais se cristalizam em torno de relações de intercâmbio, tais como funcionário e cliente e cidadão e Estado (HABERMAS, 1984).

As mídias sistêmicas, prescritas na racionalização do mundo da vida, desoneram as pessoas da necessidade de se comprometerem com a própria comunicação (RÜDIGER, 2009). Castells (2013) considera que apesar de cada indivíduo poder interpretar em seus termos o que é transmitido pela imprensa, o ambiente comunicacional onde está inserido afeta diretamente a construção de significado e relações de poder. Resta assim às instâncias intermediadoras (parlamentos, conselhos, colegiados, etc.) atuarem sobre a formação da vontade política nos fluxos comunicativos do cotidiano (COSTA, 1995).

Na visão de Habermas (2015), o fato de um potencial de dominação e distorção sem precedentes ser paralelo a um potencial emancipatório, capaz de promover uma comunidade ideal de comunicação, representa o verdadeiro paradoxo do processo histórico universal do esclarecimento.

Neste ponto é possível aproximar outros autores a sua teoria. Rüdiger (2009) faz uma conexão com Norbert Wiener, fundador da cibernética que difundiu a noção de que a comunicação constitui a base de todos os fenômenos, sejam naturais ou artificiais. Esta ideia daria base a uma “sociabilidade pós-moderna”, que se caracteriza pelo valor conferido à conexão do indivíduo a um sistema de mídia global, sobrepondo relações de convívio direto.

A perspectiva da internet como elo entre a esfera política e civil, sem intermédio dos meios de massa (WU, 2018), também se origina desta formulação. Assim como

discussões sobre a propagação da opinião pública nos meios digitais (BASTOS, 2012), e sua subsequente apropriação por políticos, a fim de se envolver com seus apoiadores compartilhando ideias e recebendo feedback sobre diferentes questões (NGUYEN, 2018). É possível encontrar sustentação teórica à “redenção da esfera pública” na própria obra de Habermas (1984), pois ele assume que a criação de esferas públicas livres de coerções do poder político e econômico poderiam por si só instaurar convicções democráticas em toda as pessoas.

Rüdiger (2009) reconhece a dimensão comunicativa do processo social, porém alerta que é preciso relativizar criticamente sua importância teórica e implicação prática a fim de evitar a “ilusão de uma democracia baseada apenas em participação virtual” (p.143). Ainda assim, como Castells (2008) destaca em sua leitura, a esfera pública varia em contexto, história e tecnologia. Por isso é preciso realizar novas investigações capazes de atualizar adequadamente o conceito. Seguindo estas recomendações, será empregado no decorrer deste capítulo uma aproximação da teoria de Habermas ao atual cenário midiático.

## **2.1 A SOCIEDADE CIVIL CONECTADA SEGUNDO A PERSPECTIVA HABERMESIANA**

Essa dissertação trabalha com a ideia de que a difusão das mídias sociais levou o conceito de esfera pública a uma nova categoria histórica. É seguro fazer este tipo de afirmação, já que até os processos de comunicação digital mais básicos suprem necessidades de arenas discursivas horizontais, como liberdade de expressão e ampla oferta de informações (WU, 2018). Magrani e Oliveira (2018) consideram que espaços digitais conectados, como fóruns de discussão, mídias sociais, e plataformas de troca de mensagens, são como a esfera pública de Habermas, só que em menor escala. Castells (2013) dá mais importância a este fenômeno, postulando que a transformação tecnológica e organizacional da “autocomunicação de massa” constitui a nova estrutura social do século XXI.

Nesta seção será abordado como transição da mídia de massa para esse novo sistema de comunicação afetou espaços de discussões e o emprego da opinião no mundo contemporâneo. Em sua tese de doutorado, Habermas (1989) descreve a decadência da esfera pública diante do predomínio da mídia de massa. Posteriormente ele afirma que esta universaliza reivindicações de validade, se

associando desonestamente a expansão da democracia e dos processos de formação de vontade (HABERMAS, 2015).

Castells (1999a) vai nessa direção ao afirmar que em uma sociedade organizada em torno da grande mídia, opiniões divergentes tendem a ficar restritas a redes interpessoais. Contudo, ele rejeita a ideia de que as pessoas eram apenas receptores passivos à manipulação ideológica, argumentando que a homogeneização das audiências ocorria devido ao controle deste sistema de mídia por parte de governos e oligopólios empresariais (CASTELLS, 1999b).

Habermas não fala diretamente, mas acena à informatização da sociedade, afirmando que “interações orientadas pela mídia podem ser interconectadas espacial e temporalmente em teias cada vez mais complexas”, mesmo que na forma de conhecimento compartilhado coletivamente, sem qualquer mecanismo de controle ou coordenação (HABERMAS, 2015).

Cabe destacar que a diversificação de conteúdos e formas de expressão nos meios digitais não trouxe o fim da grande mídia (CASTELLS, 1999b). Tanto que, na última década do século XX, diversos agentes econômicos promoviam uma série de investimentos massivos e alianças estratégicas no campo das comunicações para adquirir fatias deste mercado em transformação (CASTELLS, 1999c).

O que levanta a questão: Quem são os donos das plataformas mais utilizadas na internet? É necessário desafiar o idealismo da esfera pública e dos meios de comunicação digitais, visto que o conceito original de Habermas parte de uma crítica a sociedade (FUCHS, 2016). Além do mais, a teoria do autor engloba outros temas como poder, ideologia, capitalismo e democracia (FUCHS, 2016).

Neste novo cenário midiático, formou-se uma interseção entre meios de comunicação verticais e horizontais. As tecnologias digitais permitiram que jornais, rádios e emissoras se organizassem globalmente e interagissem com suas audiências a partir de redes de informação *online*. Ao passo que a possibilidade de criar, reformular e compartilhar todo tipo de conteúdo levava a sociedade civil a gerar seus próprios meios de comunicação (CASTELLS, 1999a).

DeFleur e Ball-Rockeach (DEFLEUR; BALL-ROCKEACH, 1993) afirmavam, ainda na década de 1960, que a iminente transição à era dos computadores iria alterar todos os processos de comunicação existentes. Para os autores, as tecnologias digitais não viriam para substituir os veículos de massa, mas sim para remodelá-los e prolongá-los (DEFLEUR; BALL-ROCKEACH, 1993). De fato, houve um período de

adaptação necessário a coexistência entre esses sistemas. Porém isso não significa que a mídia tradicional absorveu ou suplantou as novas formas de produção e distribuição de conteúdo. Logo, a internet iria se infiltrar em todas as áreas da vida social. E os efeitos desse processo de convergência tornavam cada vez mais difícil o controle dos tráfegos de informação (CASTELLS, 1999a).

Pelo menos da forma como as empresas noticiosas estavam acostumadas. Negroponte (1995) previa outra maneira de ver os jornais. O autor questionava porque as pessoas se contentavam em ler o que outras consideravam novidade ou dignas de destaque. Segundo ele, a “vida digital” podia mudar completamente o modelo econômico de seleção de notícias. Interesses pessoais teriam um papel bem maior, talvez até centralizador nas produções jornalísticas. Nessas futuras interfaces de notícias, todas as reportagens e transmissões do planeta estariam disponíveis em um resumo personalizado por respectivas predileções, hábitos de entretenimento e comportamento social (NEGROPONTE, 1995). Para isso era necessário uma explosão exponencial na quantidade de dados gerados, capturados, analisados e visualizados no âmbito cultural (MANOVICH, 2009).

Neste cenário hipotético, a mídia se torna “maior e menor ao mesmo tempo”, uma vez que seu público alvo passa de um grande grupo para grupos pequenos, e por fim a “unidade demográfica de um” (NEGROPONTE, 1995, p.164). Castells (1999b) observa o mesmo fenômeno ao relatar que os computadores viabilizaram a coleta, processamento e utilização de um enorme conjunto de informações individualizadas. Negroponte (1995) considerava isso uma espécie de triunfo do indivíduo. Já que o eu digital não seria um “subconjunto estatístico”, pois não produz informações significativas nem pode ser reduzido pela demografia clássica, que embasava as ações das grandes emissoras. Ao longo do tempo, as máquinas entenderiam pessoas com mais sutileza que elas mesmas, o que aprimoraria significativamente a forma como estas consomem a mídia (NEGROPONTE, 1995).

Esta ideia se materializou em um curto espaço de tempo. Manovich (2013) resume em uma frase o novo paradigma da comunicação: a mídia se tornou programável. O processo de segmentação de conteúdo fragmentou ainda mais a base de usuários da internet. Nesse cenário, a intensidade das interações era maior entre pessoas da mesma rede, e menor entre redes distintas. Assim, a capacidade de estabelecer interconexões através das tecnologias da informação configurou processos e funções predominantes da sociedade (CASTELLS, 1999a).

A figura da rede passa a ser indispensável em análises sociais e até epistemológicas do mundo. Bruno Latour (1994) critica o conceito de universalidade, apontando que este partia de uma fina linha de práticas e instrumentos a favor dos que o compunham. Nesta visão, certos pensamentos, situações e regras se estendiam a todos os lugares, quando na verdade, segundo o autor:

Os laboratórios, as coleções, os centros de cálculo e de lucro, os institutos de pesquisa e os escritórios de desenvolvimento traçam diariamente os limites da liberdade dos grupos sociais e transformam as relações humanas em coisas duráveis que ninguém criou.

(LATOURE, 1994, p.42)

Portanto, pensar a partir de categorias “circunstanciais” e “universais” seria um equívoco, pois estas seriam redes de fatos e conhecimentos mais, ou menos, conectadas. Para Castells (CASTELLS, 1999c) a própria estrutura do poder perpassa atores e instituições organizados em rede. Isso remonta o diagnóstico de Habermas (1984) sobre a situação do racionalismo ocidental. De acordo com o autor, este não é mais aceito sem discussão pois depende da estabilização de condições internas e de relações entre superpotências, que gera grandes custos sociopsicológicos e culturais.

Desse modo, a inclinação de transformar “redes ampliadas” em totalidades sistemáticas e globais trouxe consequências severas ao laço social, substituindo “ideias pelas práticas, os raciocínios apodícticos pela *doxa* controlada, e o consenso universal por grupos de colegas” (LATOURE, 1994, p.27). Castells (2013) afirma que as redes de comunicação, programadas em cada domínio da atividade humana, se tornaram decisivas a construção do poder e a defesa de interesses e valores de atores sociais habilitados. Segundo Durkheim (apud HABERMAS, 2015), a identidade coletiva e coesão de um grupo social está estritamente associado a imagem idealizada que este tem da sociedade.

Essa afirmação se encaixa perfeitamente a epítome da internet como base da esfera pública, as mídias sociais. A popularização dessas plataformas permitiu que usuários fizessem conexões, formassem comunidades e publicassem o que pensavam. Toda esta liberdade veio com um preço. Castells (1999a) aponta que a circulação e difusão de mensagens no sistema de comunicação digital traria batalhas culturais em um novo ambiente histórico.

Diversos tensionamentos nesta esfera pública contemporânea retomam o debate de Jürgen Habermas com Jean-François Lyotard, um dos seus grandes críticos. A obra deste autor contesta a possibilidade de uma esfera pública unificada

e baseada no consenso em uma era que “testemunhou o fracasso da legitimação de narrativas” (VILLA, 1992, p.713). Segundo Lyotard (1986), Habermas defendia que as pessoas, enquanto inteligências conhecedoras e vontades livres, chegariam a emancipação por meio do diálogo. E que, a legitimidade de qualquer enunciado deveria residir na contribuição a este objetivo.

Lyotard (1986) problematiza essa posição, alegando duas inconsistências. A primeira seria que não há divergências sobre as regras dos “jogos de linguagem”, que em sua definição abrangem todo e qualquer campo de expressão humana, como ciência, política, arte etc. A segunda seria que a finalidade do diálogo é o consenso. Quando este é um estado das discussões, e não o seu fim.

Apesar de ter apresentado essa réplica antes da difusão da internet, Lyotard (1986) ponderava como a crescente onda de informatização, descrita como “instrumento sonhado de controle”, afetaria as discussões da sociedade. Concluiu então que, por um lado isso seria capaz de transformar as pessoas em meros componentes de um sistema regido unicamente por melhorias de desempenho. Por outro, poderia dispor informações para que tomassem decisões corretas. O fator determinante aí estaria no público ter ou não livre acesso “às memórias e aos bancos de dados” (LYOTARD, 1986).

A partir deste capítulo, é possível concluir que a esfera pública está em constante transformação. Tanto nos meios que se dispõe, indo da ágora às plataformas digitais, ao seu condicionamento social, ou seja, se de fato dá voz a todos ou se restringe a um seleto grupo de participantes. Habermas não descreve precisamente as fronteiras do conceito, já que este por definição é um espaço de comunicação que não pode ser limitado ou restringido - contudo, o autor demarca a importância do processo de deliberação pública, que leva temas e contribuições de esferas autônomas a foros formais e institucionalizados do sistema político e administrativo (LUBENOW, 2015).

Neste ponto há uma conexão com o objeto de estudo desta obra: a Câmara dos Deputados. Apesar de não ser abordado diretamente nos livros percorridos nesta seção, o poder legislativo é imprescindível a teoria Habermasiana. Pinto (2005) explica que quando a comunidade jurídico-política alcança consensos de modo intersubjetivo, e cidadãos veem nas leis a manifestação racional e livre de suas próprias vontades, o processo de integração social está completo.

A centralidade deste tema é compreensível, visto sua grande contribuição e reconhecimento na área do direito:

Para Jürgen Habermas, nas sociedades modernas, o Direito tem a tarefa preponderante de ser uma das suas formas principais de integração, assumindo algo do papel transcendente que a religião e as tradições desempenharam no passado. Para isso, no entanto, é necessário haver condições processuais adequadas que propiciem o fluxo de formas comunicativas democráticas por meio das quais os cidadãos, em condições de igualdade, possam participar da formação de uma decisão político-jurídica oponível a todos

(VIEIRA, 2019, p.40)

Castells (2008) acede a definição de esfera pública como espaço de comunicação de ideias entre a sociedade e o âmbito das decisões institucionais, acrescentando que o processo de globalização rompeu com as fronteiras desse debate, que agora se dá através de redes digitais descentralizadas. Este contexto possibilita produções acadêmicas voltadas a análise quantitativa de canais de comunicação política e da utilização das novas plataformas midiáticas (ENJOLRAS, 2014).

Antes de partir para exploração dos dados nesta perspectiva, é preciso investigar se os fluxos de informações digitais são de fato neutros e livres, ou estão comprometidos pela falta de transparência no recorte algorítmico das grandes plataformas (MAGRANI; OLIVEIRA, 2018). Para abordar esta questão, será empregada uma revisão de literatura sobre o impacto social da invenção e popularização da internet.

### 3 O IMPACTO SOCIAL DA INTERNET

Em 1945 o engenheiro Vannevar Bush elaborou uma metáfora sobre as dificuldades de inventar a “máquina do futuro”, capaz de armazenar e consultar informações com extrema facilidade: Se fosse dado a um faraó a planta baixa de um automóvel, e ele destinasse todos os recursos do seu reino para construí-lo, este quebraria na primeira viagem a Gizé (BUSH, 1945).

Desenvolver um computador que pudesse ser usado por multidões parecia ser apenas um problema técnico, porém havia nele uma dimensão social que passava despercebida aos colegas de Bush. Seguindo sua linha de raciocínio, se todas as instituições dos anos 40 direcionassem seus recursos a esse feito, talvez obtivessem sucesso. Porém, isso não contemplaria a todos. Primeiro porque a produção em massa desse computador ainda estaria economicamente muito distante. Segundo porque exigiria um novo nível de educação e diversos outros fatores que, em síntese, apontariam à uma sociedade completamente diferente.

Na época, computadores só eram usados para fins militares e acadêmicos, por pessoas versadas em programação e manipulação de dados. Ou seja, por uma ínfima parcela da população. Bush (1945) clamava para que cientistas enfrentassem a enorme tarefa de torná-los mais acessíveis, pois à medida que a civilização se complexificava, a mecanização dos seus registros se mostraria necessária. Segundo o autor, durante anos invenções científicas ampliaram os poderes físicos do ser humano. A partir daí, deveriam ampliar os poderes de sua mente, o elevando para “rever melhor seu passado sombrio e analisar mais completa e objetivamente seus problemas atuais” (BUSH, 1945).

Em questão de décadas, esse salto paradigmático aconteceu, superando todas as expectativas de Bush. Segundo Manovich (2013), além de armazenar e produzir qualquer tipo de mídia, o computador pessoal se mostrou capaz de estabelecer redes de comunicação, integrando sistemas sociais e econômicos a uma sociedade global da informação. Para Castells (1999a), a ubiquidade dos sistemas tecnológicos alterou todos os aspectos da vida cotidiana. Ainda que diferentes contextos históricos, culturais e institucionais permaneçam relevantes, essa nova sociedade vinha com sua própria lógica, informacional, sim, mas também capitalista (ibid, 1999a).

De acordo com Milton Santos e María Laura Silveira (2001), aqui a história passa a ser regida por fluxos imateriais, seja capitais, informações, mensagens ou

ordens. É deste modo que a “inteligência do capital” reúne distintos circuitos espaciais de produção, estabelecendo redes hierárquicas que vão da escala regional até a escala mundial (ibid, 2001).

Manovich (2013) relata que até o início dos anos 90 as maiores empresas do mundo produziam bens de consumo, materiais, ou processavam matéria física. Ademais, a maioria das redes de televisão era governamental, e estações de rádio e jornais estavam sujeitos a restrições por parte das autoridades, mesmo em países democráticos (CASTELLS, 1999b).

A partir dessa década, o cenário econômico passou por grandes mudanças. Fusões e aquisições de empresas no setor do audiovisual, das telecomunicações, da produção gráfica e dos serviços postais tornaram a Comunicação o negócio mais rentável da época (GOMES, 2009). As bases de regulamentação da mídia foram completamente destruídas pelo surgimento de um poder de transmissão e difusão de informações sem precedentes, decorrente não só da comunicação via computador mas também de sinais de comunicação via satélite, impossíveis de serem contidos em fronteiras nacionais (CASTELLS, 1999b).

Segundo Gomes (2009), a consolidação dos novos impérios midiáticos implicava na transnacionalização de indústrias culturais, principalmente a países em desenvolvimento, com grande potencial consumidor. Ao mesmo tempo, o progresso na eletrônica e na informática possibilitavam a interligação, em tempo real, de bolsas, bancos, e praças financeiras, fazendo com que fluxos de dinheiro circulassem internacionalmente em velocidades cada vez maiores (SANTOS; SILVEIRA, 2001).

Castells (1999b) frisa que Estados só conseguiram manter a influência neste meio através de aquisições de ações de capital e relações próximas com suas organizações. Para Santos e Silveira (2001), a nova divisão territorial do trabalho parte de uma trama densa e complexa de sistemas de engenharia, que delimitam a capacidade de agir de pessoas, empresas e instituições. Em outras palavras, “um novo conjunto de técnicas torna-se hegemônico e constitui a base material da vida da sociedade”, tornando o meio “técnico-científico-informacional” a expressão geográfica da globalização (SANTOS; SILVEIRA, 2001, p.24).

Na outra margem desse processo estava a pretensão de resolver todos os problemas do mundo a partir da informática. Gillespie (2018) afirma que a criação da internet promoveu a ideia de uma cultura mais igualitária, criativa e inclusiva. Lanier (2018) traz ressalvas a esse movimento, apesar de ser um de seus precursores. Ele

conta que o “espírito libertário” da primeira geração de desenvolvedores deixou necessidades fundamentais do mundo digital, como mecanismos de identidade pessoal e transações financeiras, a encargo de empreendedores, ao invés de delegar por regulamentações do governo. O que resultou na “estupidez de preparar o terreno para monopólios globais” (LANIER, 2018, p.26).

Santos e Silveira (2001) não teorizaram sobre este estágio do capitalismo informacional, contudo, há em sua obra uma consideração fortuita a este contexto. Segundo os autores, a adequação de técnicas e políticas a serviço do lucro de grandes empresas representa sempre uma drenagem de recursos sociais para a esfera do setor privado, ou ainda uma “racionalidade privada obtida com recursos públicos” (SANTOS; SILVEIRA, 2001, p.341)

Esse cenário contribuiu para que empresas de tecnologia da informação se tornassem as maiores marcas globais, como aponta Manovich (2013). Na década de 2010, boa parte de todos os discursos públicos e produções culturais se davam inteiramente em plataformas digitais. Além disso, estas acabaram integrando funções sociais, financeiras, e até mesmo institucionais. O que as aproximou mais do conceito original de computador, uma “máquina universal” para todo tipo de troca de informações (GILLESPIE, 2018).

Segundo Manovich (2013) a partir daí se tornou impossível estudar a produção, disseminação e consumo de mídia como processos separados. Já que esses serviços borraram a fronteira entre “informações pessoais” e “informações públicas, à medida que mais pessoas os usavam para se comunicar e compartilhar conteúdo (MANOVICH, 2013).

5 multinacionais dominam este mercado no ocidente: Google, Apple, Facebook, Amazon e Microsoft, conjunto apelidado pelo acrônimo GAFAM<sup>3</sup> por Nieborg e Poell (2018). Gillespie (2018) coloca que o gerenciamento desses espaços, pelo menos no que toca as mídias sociais, é uma faca de dois gumes - Facilitam processos de comunicação, trazendo um sentimento de conexão e pertencimento entre as pessoas. Porém, ao tratar sua base de usuários, já na casa dos bilhões, como uma comunidade, desconsideram que esta possui línguas, crenças e hábitos e diferentes.

---

<sup>3</sup> Acrônimo desatualizado no momento da redação deste texto, visto que as companhias Google e Facebook mudaram seus nomes para Alphabet e Meta, respectivamente.

Até mesmo o jornalismo fica sujeito aos algoritmos da GAFAM, que afetam a visibilidade tanto de conteúdos individuais, quanto de agências de notícias inteiras (NIEBORG; POELL, 2018). Ao se alinhar por marcadores de popularidade como curtidas e acessos, em vez dos seus indicadores de qualidade tradicionais, organizações noticiosas acabam ignorando valores e sensibilidades históricas, culturais e políticas do seu público (NIEBORG; POELL, 2018).

Tudo isso parte do modelo de negócios das plataformas digitais, que armazena e analisa dados pessoais para segmentar conteúdo e anúncios publicitários (GILLESPIE, 2018). Isso tornou realidade a antiga projeção de Negroponte (1995) sobre vida digital: cada usuário recebe uma resposta personalizada do sistema que está inserido, a partir de seus hábitos de entretenimento e comportamento social. Segundo De Faria (2012), utilizar informações dos usuários, como preferências pessoais, idade, gênero e estado civil para fins comerciais foi uma das grandes revoluções do mercado publicitário na história recente.

Aqui, a análise de Habermas (1989) sobre o uso programado do tempo de lazer nas sociedades capitalistas encontra sua hipérbole. Rüdiger (2006) aponta que a modernidade trouxe o predomínio do cálculo a todas as ordens de existência. Obras, experiências e situações são multiplicadas e consumidas em massa para em seguida perderem seu valor. Assim, o mundo se reduz a uma “corrente de imagens que jamais se furta a nosso olhar” (ANDERS apud RÜDIGER, 2006).

Essas constatações valem tanto para o entretenimento como para política. De acordo Lanier (2018), a indústria da propaganda digital não se importa com quem são os anunciantes, nem com o que eles fazem com os dados. E isso deixa margem para que agentes econômicos se utilizam das mídias sociais como ferramentas para modificação massiva de comportamentos. Habermas (1989) constata que na época da mídia de massa agências de comunicação construía discursos políticos a partir da observação detalhada do eleitorado, a fim de manipulá-lo. Segundo o pensador, essa prática era incompatível a democracia. Teóricos contemporâneos vêm chegando a mesma conclusão sobre os temas levantados acima. Se torna cada vez mais evidente que o papel democrático da esfera pública conectada não dever ser encarado como a “panaceia para a salvação da legitimidade do sistema político contemporâneo” (MAGRANI; OLIVEIRA, 2018, p.45).

No âmbito dos debates e campanhas políticas, isso causa uma profunda inquietação, pois é aí que se põe a prova a neutralidade das mídias sociais. De Faria

(2012) afirma que a facilidade e a liberdade de apresentar opiniões políticas naturalmente gera fragmentação e polarização nos debates, tornando as pessoas mais radicais e reduzindo o sentimento de comunidade (DE FARIA, 2012). Nos últimos anos, esta predisposição tem sido instrumentalizada para ganhos eleitorais.

De acordo com Lanier (2018), o conteúdo com maior alcance nas eleições presidenciais americanas de 2016 consistia em informações não verificadas que inflamavam o ânimo dos eleitores. Esse episódio trouxe à tona discussões sobre a “compra de atenção de usuários” e predominância dos “filtros bolha” - espaços digitais nos quais são mostrados apenas conteúdos de interesse do usuário – por muitos considerados prejudiciais a formação de consenso na esfera pública conectada (MAGRANI; OLIVEIRA, 2018). Gillespie (2018) aponta que todas as mídias sociais proíbem comportamentos abusivos e a promoção de violência. Contudo, estas regulações muitas vezes não são aplicadas.

Neste ponto os autores apresentam visões opostas sobre o mesmo problema. Lanier (2018) não acredita em uma solução técnica para isso. Todas as eventuais medidas tomadas contra campanhas de desinformação, discursos de ódio ou contas falsas seriam alegóricas. As mídias sociais não teriam interesse em lidar com este tráfego anômalo de informações, visto que este aumenta a atividade do serviço e geram realidades sociais genuínas, deixando-as mais relevantes. Para ele, isso não seria algo necessariamente negativo, pois pedir para estas companhias tratarem comportamentos abusivos apenas aumentaria o controle delas sobre a vida das pessoas. O posicionamento do autor está exposto no título do seu último trabalho: é preciso abandonar as mídias sociais.

Magrani e Oliveira (2018) analisam este debate sob um viés habermasiano, chegando à conclusão de que esferas virtuais verdadeiramente democráticas e a “descolonização” do mundo da vida arrefecem diante da intervenção de agentes não-humanos (como filtros algorítmicos) que influenciam discursos e visões políticas.

Gillespie (2018) é mais otimista. Para ele, o sonho da “web aberta” não falhou, nem se tratava de promessas vazias. Todas as controvérsias mencionadas anteriormente devem ser vistas como um amadurecimento lento e instável das mídias sociais no que seria uma poderosa infraestrutura de conhecimento, participação e expressão pública.

Estas não só podem como devem moderar a experiência dos usuários, só que de maneira transparente, assumindo o compromisso de revelar o critério de

visibilidade das publicações e possibilitar avaliá-lo. Nesse cenário, conferir mais agência aos usuários significa mostrar o mesmo mundo de discursos disponíveis a todos, sem incentivá-los a navegar por discursos sobrepostos (GILLESPIE, 2018).

De certa forma, Castells (1999a) antecipa esse debate. Para o autor, o período de grande transformação tecnológica no início do século XXI fez com que fluxos de mensagens e imagens entre redes se tornassem o encadeamento básico da sociedade. Nessa “era da informação”, a espécie humana finalmente alcançou um mundo predominantemente social. Isso não quer dizer que “a história terminou em uma feliz reconciliação da humanidade consigo mesma”, mas que seria preciso encarar sua realidade histórica ou viver com as consequências (CASTELLS, 1999a). Santos e Silveira (2001) também preconizam o tema afirmando que o crescente consumo de informações da “psicosfera modernizadora” iria entretecer racionalidades e emoções convergentes e conflitantes.

A internet e as mídias sociais não trazem por si só uma sociedade conectada, plural e inclusiva. E não há ajustes técnicos que garantam isso. Utilizando a metáfora de Bush, a planta baixa está à disposição, e a máquina já foi montada. Agora resta entender até onde ela nos levou e porque parou de funcionar como devia. Para cumprir este objetivo serão apresentadas as consequências reais do processo de informatização do Brasil.

### **3.1 O PROCESSO DE INFORMATIZAÇÃO DO BRASIL**

A informatização da sociedade se mostrou mais complexa do que os primeiros diagnósticos de revolução tecnológica. De Faria (2012) afirma que todas as inovações no campo da Comunicação são acompanhadas por uma euforia, que no melhor dos casos só se concretiza de forma parcial. Na visão do autor, este ciclo de altas expectativas e decepções se deve à falta de compreensão sobre o papel das instituições políticas.

Pode-se supor que a solução para este problema estaria em comparar a teoria com a prática; O que autores com maior destaque falavam sobre a internet e o que de fato aconteceu quanto esta foi difundida. Não seria possível, pelo menos nesta dissertação, alcançar isto abrangendo todos os países do mundo. Contudo, pode-se partir à realidade local, apontando como fatores socioeconômicos do Brasil alteraram seu processo de informatização.

Segundo Castells (1999c), a sociedade em rede surge no final do século XX a partir de três grandes mudanças: a revolução da tecnologia da informação; a reestruturação do capitalismo e o apogeu dos movimentos sociais. A simultaneidade destes fenômenos, considerada pelo autor como uma coincidência histórica, resultou em uma capacidade produtiva e potencial criativo sem precedentes.

Por outro lado, redes globais de competitividade econômica, sem qualquer vínculo com instituições de Estado ou organizações de sociedade civil, enfraqueceram mecanismos de controle social e representação política (CASTELLS, 1999b). Santos e Silveira (2001) listam uma série de motivos que países subdesenvolvidos teriam para rejeitá-las, entre eles o peso da ideologia do crescimento, o imperativo de afirmar o Estado sobre a nação, e a indispensabilidade de um comando eficaz sobre o próprio território. À despeito de tudo isso, a adaptação ao modelo capitalista internacional, e a respectiva ideologia de racionalidade e modernização a qualquer custo, prevaleceram de modo indiscutível ao redor do globo (SANTOS; SILVEIRA, 2001).

No Brasil, este processo não foi pontual nem espontâneo. Milton Santos e María Laura Silveira o analisam através da história territorial do país, separada por três longos períodos. Primeiramente foi buscada a unificação do território, do mercado e da linguagem. Em seu extenso estudo sobre a história brasileira, Schwarcz e Starling (2015) discorrem sobre um período correlato a este, entre a colonização e a independência. Segundo as autoras, sempre houve tensão entre um poder central, proveniente do império português, e poderes regionais, formados por elites econômicas e aristocráticas. Incluídas, neste ínterim, as relações perversas que se constituíam na escravidão. Este cenário ebulia no final do século XVIII, causando diversas insurreições a coroa.

A articulação de um movimento de revolta em Salvador, conhecido como a “Conjuração Baiana”, traz semelhanças ao desenvolvimento de uma esfera pública moderna:

No início do ano de 1798, na cidade de Salvador, amanheceu queimada a forca instalada no largo em que se erguia o Pelourinho — símbolo máximo do poder da Coroa portuguesa [...] Quem fez isso permaneceu anônimo, mas deixou um lembrete e expressou uma opinião tangível ao pé da madeira queimada e das portas que protegiam a cidade: afixou “ludibriosos pasquins” [...] De diversas maneiras, seu significado político era notável, a começar por sua aparição pública, quando fizeram transitar livremente pelas calçadas ideias, notícias e opiniões que até então só podiam ser transmitidas por meio de canais clandestinos. Além disso, o conjunto de panfletos que sobreviveram até o século XXI aponta outra novidade: por seu intermédio colonos articularam as ideias de República e de democracia, organizando, pela primeira vez, o repertório político de baixo para cima na estrutura social e expandindo para a população pobre da cidade o atributo da igualdade política.

(SCHWARCZ; STARLING, 2015, p.215)

De acordo com Perlatto (2012), diversos estudos afirmam que uma esfera pública brasileira só foi surgir após a redemocratização do país, na década de 1980. Esta ideia seria omissa a recentes avanços historiográficos, que afastaram a imagem de “sociedade civil apática, desinteressada e bestializada” ao encontrar evidências de formas de associação e organização que, apesar de não seguirem padrões europeus, “possibilitaram a coordenação de discursos públicos capazes de problematizar elementos da ordem vigente” (PERLATTO, 2012, p.85).

Mesmo após o processo de independência, o ambiente político interno era marcado pela divisão, a ponto de não haver acordo acerca das estruturas básicas que o Estado iria se organizar. Apesar disso, a primeira Constituição brasileira foi concluída em 1823, contemplando de modo equilibrado diferentes interesses, mas deixando de fora “projetos democráticos e populares na condução do país” (SCHWARCZ; STARLING, 2015, p.342).

O processo de formação do Estado brasileiro difere ao de países europeus, como descrito por Habermas (1989). O que mais se distingue é a falta de uma classe burguesa, capaz de regular relações sociais por mecanismos do mercado. O próprio Estado tomou esta função, povoando seu setor burocrático com profissionais liberais e advogados, “os famosos bacharéis” (SCHWARCZ; STARLING, 2015, p.419). Contudo, a característica mais tenaz se manteve a mesma. A instância das tomadas de decisão estava no jugo das elites brancas. Schwarcz e Starling (2015) a descrevem como uma ilha de letrados em um mar de analfabetos, destacando que até 1872 apenas 16% da população era alfabetizada.

Este quadro se perpetuou devido ao êxito em bloquear canais adequados de formação e ascensão de novas lideranças, mantendo assim o alinhamento político a oligarquias, a manutenção do status quo e a falta de representação a grande parte da população (ABRANCHES, 2018). É desta maneira, suprimindo diversidades e postergando crises, que se completa o processo de unificação de um sistema nacional. No segundo período da história territorial do país, o objetivo recai na construção de uma indústria própria. Este último ponto é discutido extensivamente por Santos e Silveira (2001) a partir de dados sobre a urbanização tardia, voltada fortemente a região sudeste.

As metrópoles mais buscadas no início do século XX eram Rio de Janeiro, amplamente beneficiada por ter sido a capital do país por quase duzentos anos, e São Paulo, cujo desenvolvimento industrial garantiria centralidade no âmbito econômico. No período atual, simultâneo ao processo de globalização, dois eventos ganham destaque: disparidades regionais tornam-se estruturais devido a concentração de investimentos e infraestrutura na Região Concentrada (termo usado para se referir as Regiões Sudeste e Sul), e o “motor” da sociedade deixa de ser a indústria e passa a ser a informação (SANTOS; SILVEIRA, 2001).

Neste contexto, o fator crítico para a geração de riqueza está na capacidade de utilizar e se adaptar a novas tecnologias (CASTELLS, 1999c). Chegando a este estágio, boa parcela da economia urbana pode se desenvolver sem investimentos adicionais em infraestrutura, a ponto de o custo das metrópoles informacionais ser menor ao que se tinha anteriormente (SANTOS; SILVEIRA, 2001). Ainda assim o “informacionalismo” depende da matriz histórica de onde está inserido, já que a dinâmica globalizante não apaga restos do passado, apenas modifica seu significado e acrescenta características do novo tempo (CASTELLS, 1999c; SANTOS; SILVEIRA, 2001). A trilogia de Castells sobre a era da informação foi publicada em 1998. Mais de dez anos depois, Manovich discorre sobre o panorama do seu campo de pesquisa, as humanidades digitais:

Centenas de milhares de estudantes, artistas, designers agora têm acesso às mesmas ideias, informações e ferramentas. Não é mais possível falar em centros e províncias. Na verdade, com base em minhas próprias experiências, acredito que os estudantes, profissionais da cultura e governos em países recém-globalizados estão frequentemente mais prontos para abraçar as ideias mais recentes do que seus equivalentes nos "velhos centros" da cultura mundial.

(MANOVICH, 2009, p.2)

Ainda que esta impressão não possa dar fundamento a quaisquer conclusões sobre geração e distribuição de renda no mundo contemporâneo, dá indícios de que estes processos não são predeterminados. Santos e Silveira (2001) entendem informação como o recurso que gerencia meios geográficos sob a égide da ciência e da técnica. Os autores rejeitam termos absolutos: existiriam territórios bem-informados, pouco informados e um número infinito de situações intermediárias. Assim, pode-se imaginar que isso varie de país a país, de região a região etc. Resta a questão de como isso acontece, qual é o fator determinante para que uma nação

tenha protagonismo, ou pelo menos agência, em uma economia global completamente digitalizada.

Segundo Castells, décadas de pesquisa sobre sociologia do desenvolvimento revelaram sem sombra de dúvida que “processos de crescimento econômico e transformação estrutural são embutidos nas instituições, direcionados pela cultura, apoiados pelo consenso social, modelados pelos conflitos sociais, disputados pela política e conduzidos por políticas e estratégias” (CASTELLS, 1999c, p.254).

Cabe ressaltar que a dificuldade não está só em pensar um modelo de desenvolvimento à par de seu tempo, mas de implementá-lo diante de influentes parcelas da população, protegidas dos efeitos socialmente desintegradores de um crescimento desenraizado, mas retoricamente afirmado por suas visões de mundo (HABERMAS, 1984). Segundo Jurgen Habermas:

Os neoconservadores querem manter a qualquer preço o padrão capitalista de modernização econômica e social. Eles dão a mais alta prioridade ao crescimento econômico que o compromisso do estado de bem-estar social promove e também restringe cada vez mais.

(HABERMAS, 1984, p.46)

A nova geografia, configurada pelos fluxos de informações e finanças, trouxe necessidades produtivas e divisões sociais do trabalho que em um primeiro momento só foram contempladas na “Região Concentrada” do país, onde havia um denso sistema de relações urbanas e elevados padrões de consumo de empresas e famílias (SANTOS; SILVEIRA, 2001). Esta realidade fez com que o “novo” custasse a ser completamente difundido pelo território nacional – Santos e Silveira (2001) trazem dados sobre a disseminação de diversas tecnologias no século XX, como rodovias, eletricidade e meios de comunicação. O mais relevante à esta dissertação concerne o uso da internet. Segundo os autores, no início de 1996 havia 170.429 usuários, dois anos mais tarde o número cresce para 1.310.001 (SANTOS; SILVEIRA, 2001).

A partir deste ponto, os autores não apresentam mais dados. Felizmente, a série histórica da pesquisa TIC Domicílios (CGI.BR, 2019) do Comitê Gestor de Internet do Brasil permite observar o processo de difusão do acesso domiciliar à rede. Em 2008, primeiro ano em que a pesquisa foi realizada em áreas urbanas e rurais, já havia mais de 50 milhões de usuários. Em 2019, cerca de 134 milhões, o equivalente a 74% da população com dez anos ou mais. Uma a cada quatro pessoas não usava a rede no país. Neste contingente, de aproximadamente 47 milhões de pessoas, 95%

pertenciam às classes C e DE e 85% possuíam até o ensino fundamental. Este quadro demográfico é o que mais impulsiona os índices de acesso atualmente, mesmo permeado em severas disparidades:

"A presença de computadores está associada a fatores sociodemográficos: em 2019, eles estavam presentes em 95% dos domicílios da classe A, mas em apenas 44% dos domicílios da classe C e 14% dos domicílios das classes DE."

(CGI.br, 2019, p.366)

De acordo com a pesquisa, o motivo mais citado para falta de internet nos domicílios foi o valor do serviço (59%), em seguida falta de interesse e o fato de as pessoas não saberem usar a rede. Este último dado traz uma problemática maior, que pode ser observada no modo como as pessoas usam a internet.

O CGI.BR (2019) revela que atividades de comunicação foram as mais usadas na rede, sendo o envio de mensagens realizado por 92% dos usuários, depois o uso de mídias sociais (76%) e chamadas por voz ou vídeo (73%). 47,6 milhões de indivíduos, cerca de um quarto da população, criaram e postaram algum conteúdo próprio na Internet, como textos, imagens, vídeos e músicas. Este índice foi mais frequente em camadas da população com maior grau de escolaridade: 46% entre quem tinha Ensino Superior, 18% entre quem tinha apenas Ensino Fundamental e apenas 3% entre analfabetos (CGI.BR, 2019).

Autores citados neste capítulo supuseram que a proeza com as tecnologias da informação seria o fator-chave para o desenvolvimento social e econômico no século XXI. Isto foi comprovado com a pandemia do novo coronavírus. O controle desta crise exigia uma população bem-informada, devido à ausência de um medicamento e a necessidade de atenção pessoal para evitar o contágio (DE SOUSA JÚNIOR et al., 2020). Contudo, a praticidade de compartilhar informações nas mídias sociais trouxe alguns problemas. O primeiro estaria na falta de “alfabetização digital”, que leva muitas pessoas a divulgar mensagens sem verificar suas fontes, o segundo, na multiplicação de rumores e a manipulação de informações com intenção duvidosa (GARCIA; DUARTE; DE SOUSA JÚNIOR et al., 2020).

Diante desta situação atípica, o CGI.BR realizou uma edição extraordinária da pesquisa TIC Domicílios (CGI.BR, 2020). No início da adesão ao isolamento social, o tráfego de Internet registrou um pico de cerca de 13,5 *terabits* por segundo - volume inédito no país (CGI.BR, 2020). Também se sobressaiu a exclusão digital de uma parcela significativa da população. Como foi visto anteriormente, a barreira do acesso

vem sendo superada nos últimos anos. Contudo, há um segundo nível de exclusão, relativo à qualidade do acesso. O uso da Internet exclusivamente por celular, predominante entre a população de baixa renda, está diretamente relacionado com um menor aproveitamento de “atividades culturais, pesquisas escolares, cursos à distância, trabalho remoto e utilização de governo eletrônico” (CGI.br, 2019, p.366).

Santos e Silveira (2001) apresentam uma espécie de paradoxo que impediria a solução definitiva deste problema: O Estado brasileiro precisa reunir forças “mais concentradas e maciças” para operacionalizar a economia globalizada e estar a par com os países desenvolvidos. Porém, o preço disso é renunciar funções de regulação social, desconsiderando assim tanto a nação quanto o território (SANTOS; SILVEIRA, 2001).

Neste ínterim, as capitais mantêm centralidade através de uma nova conjuntura. Legalmente, a regulação normativa do país reside em Brasília, através da criação de leis impulsionadoras ou limitadoras de ação. Contudo, este comando político está subordinado ao mesmo fator que sustenta a posição hierárquica de São Paulo sobre a vida econômica nacional: a concentração de informações e serviços na tomada de decisões (SANTOS; SILVEIRA, 2001).

Bases de dados e ferramentas de análise dispostas nesta dissertação permitem investigar como agentes de poder político, no caso deputados federais, direcionaram recursos da máquina pública em um momento de crise sem precedente no cenário internacional, que acelerou a tendência da informatização e contrariou a de redução do papel do Estado na sociedade. Ademais, como estes parlamentares usaram a popularidade no Twitter e influência na Câmara para promover suas agendas. Para compreender melhor estes fenômenos, será levantado um histórico de como as mídias sociais vem sendo utilizadas na política.

## 4 MÍDIAS SOCIAIS E POLÍTICA

A Internet desempenhou um papel importante em todas as eleições americanas desde sua apresentação ao público na década de 1990 - isso evoluiu à medida que as campanhas tentavam encontrar uma vantagem que lhes permitisse disseminar ainda mais suas mensagens (MASCARO; AGOSTO; GOGGINS, 2016). As mídias sociais catalisaram esse processo ao permitir que pessoas não só as usassem para interagir umas com as outras, mas também para ler e compartilhar notícias, discutir eventos importantes e se envolver em todo o tipo de discussões (YAQUB et al., 2017).

Assim, grandes plataformas como Facebook, Instagram e Twitter se tornaram indispensáveis à sociedade moderna, mas o Twitter, de longe, virou o espaço mais proeminente para conversas políticas (POSEGGA; JUNGHERR, 2019). Neste serviço de *microblog*, usuários publicam e compartilham diariamente um grande volume de mensagens curtas chamadas “*tweets*” (MONTI et al., 2013). Segundo Silva (2012a), o criador deste serviço não o reconhece como rede social, mas como uma “rede de informações”. Independente de quão válida é essa afirmação, cabe considerar o seguinte: hoje o que se toma por “rede social” parte de sistemas como os do Facebook, baseados em gostos e relacionamentos pessoais, que enviam fontes de informação (CHAGAS; TOTH, 2016).

O Twitter não está livre desse dilema. A questão aqui é que entre o usuário e todo esse contingente está apenas a função de “seguir”. É claro que existem mediações algorítmicas, sujeitas a alterações por parte dos desenvolvedores, porém o apelo da plataforma está justo nesse sentimento de estar apenas a “um passo de distância” de qualquer pessoa do mundo.

A massiva discussão pública gerada por este contingente provou ser uma fonte de dados relevante para explorar tendências de sentimento e opiniões políticas (MONTI et al., 2013). Além disso, sua Interface de programação de aplicações (API), permite que terceiros “dialoguem” com seus servidores. A soma destes fatores faz com que esta provavelmente seja a mídia social mais analisada cientificamente hoje (POSEGGA; JUNGHERR, 2019).

A partir de um recurso nativo chamado *trending topics*, é possível acompanhar quais são os assuntos mais discutidos em sua região. Entrevistas e análises de conteúdo realizadas por Posegga e Jungherr (2019) apontam que a “agenda” de tópicos do Twitter tem pouca ou nenhuma relação com a agenda pública e uma

relação um pouco mais forte com a cobertura da mídia de massa. Os temas mais discutidos geralmente são polêmicas com alto envolvimento de ativistas online, e fortes críticas ao extremismo de direita e ao populismo. Portanto, é possível concluir que a conversa política no Twitter é muito mais do que um simples reflexo das agendas públicas ou da mídia (POSEGGA; JUNGHERR, 2019).

Hu et al. (2016) argumentam que esse índice de popularidade pode funcionar, mas não contempla toda a atividade da rede pois as postagens mais populares geralmente são produzidas por pequenos grupos de formadores de opinião. Além disso, esse indicador é passível de fraudes. Lokot e Diakopoulos (2016) indicam que até 8,5% das contas podem ser automatizadas - muitas dessas servem como vetores poderosos de spam que manipulam percepções no discurso político (LOKOT; DIAKOPOULOS, 2016).

A estratégia de Barack Obama nas eleições presidenciais de 2008 foi um marco no uso do Twitter. A combinação de habilidades tecnológicas com proezas políticas articulou um movimento popular que lhe deu a vitória, mesmo ele sendo um senador relativamente desconhecido (MASCARO; AGOSTO; GOGGINS, 2016). Desde então, campanhas políticas em todo o mundo passaram a utilizar a plataforma (POSEGGA; JUNGHERR, 2019).

Em teoria, eleitores obtinham atualizações personalizadas dos candidatos e interagiam diretamente com eles. Na prática, foi um pouco diferente (MASCARO; AGOSTO; GOGGINS, 2016). Mascaro et al. (2016) examinaram 53 milhões de *tweets* coletados durante a eleição presidencial americana de 2012. Embora a tecnologia de interação entre cidadãos e figuras públicas existisse na época, ela não foi totalmente utilizada. Os candidatos usavam o Twitter apenas para divulgar informações de interesse próprio, não dando continuidade na maioria das interações. No entanto, os autores afirmam que o engajamento político na Internet iria se fortalecer com o envelhecimento da geração *millennial*.

Na eleição presidencial dos EUA em 2016, o anúncio das candidaturas foi feito diretamente no Twitter a fim de gerar mais engajamento online (LE et al., 2017). Yaqub et al. (2017) analisaram o comportamento de mais de um milhão de usuários, investigando se a plataforma foi usada para compartilhar opiniões originais e gerar debates, ou se poucas opiniões foram repetidas continuamente com pouco diálogo. Naquela ocasião, era de senso comum que os políticos não podiam mais ignorar o papel das mídias sociais como um canal de comunicação bidirecional para monitorar

e medir continuamente as reações do público (YAQUB et al., 2017). No entanto, os resultados deste artigo não são muito diferentes do estudo de Mascaro et al. (2016). Em termos de comportamento do usuário, pouco conteúdo original foi criado pelos usuários durante as discussões, e a maioria dos usuários não se envolveu em conversas diretas entre si.

A campanha de Donald Trump em 2016 usou estrategicamente o Twitter para desencadear cobertura midiática. Segundo Wells (2020), o então candidato se valia de posicionamentos populistas, ataques pessoais a oponentes e discursos xenofóbicos na plataforma para se tornar pauta na imprensa. A retomada desta estratégia nas eleições presidenciais de 2020 levou ao banimento permanente de sua conta. Em nota, o Twitter justificou esta medida afirmando que o ex-presidente estava incitando violência e difundindo sem provas que as eleições foram fraudadas (BBC, 2021).

Os Estados Unidos podem ser o principal exemplo sobre o uso político de mídias sociais, mas estas vem desempenhando um papel cada vez mais significativo em eleições ao redor do mundo (GORKOVENKO; TAYLOR, 2016; LE et al., 2017). Conseqüentemente, o número de pesquisas sobre esse fenômeno também aumenta. A questão de quão representativa é a base de usuários do Twitter se agrava em países em desenvolvimento, onde grande parte da população não tem acesso à Internet (YAQUB et al., 2017). Dados da pesquisa *We are Social* (2021) revelam que esta é a 6ª mídias social mais usada no Brasil.

Outros problemas, como notícias falsas, filtros bolha, mineração de dados pessoais e anúncios políticos direcionados indicam que a mídia social está tendo um impacto dramático em diversos processos democráticos (GORKOVENKO; TAYLOR, 2019). Panda et al. (2020) analisaram seu uso na Índia, concentrando-se em discursos extremistas. Os autores afirmam que as desigualdades socioeconômicas do país ficam evidentes nesta mídia social, visto que políticos não apenas são protegidos por exposições que cidadãos comuns podem enfrentar, mas também são recompensados por serem controversos ou incendiários (PANDA et al., 2020). Embora o Estado não seja responsável por incivildades no Twitter, o legado colonial da Índia delimita o posicionamento de seus cidadãos. O que favorece atores institucionais e reforça hierarquias existentes. Gonawala et al. (2020) estudaram o impacto de *tweets* polarizantes nas eleições gerais indianas de 2019. Este artigo afirma que as mídias

sociais libertaram os políticos de convenções tradicionais de discurso, e acabaram favorecendo conteúdos antagônicos.

Ao lado da Índia, o Brasil é uma das maiores democracias do mundo, então entender as necessidades e os assuntos discutidos pelos eleitores é de extrema importância, não só para políticos e profissionais da área, mas também para o cidadão comum (BERTALAN; SERON RUIZ, 2019). Bertalan e Seron Ruiz (2019) usaram técnicas de Processamento de Linguagem Natural (PNL) para encontrar os principais tópicos de discussão em sites políticos brasileiros. Os autores indicam que as redes de mídia convencional não são mais as únicas instituições com credibilidade popular. Em todos os lados do espectro político, as pessoas começaram não apenas a consumir notícias e textos políticos, mas também a produzir seu próprio material. A principal consequência desse novo modelo de mídia é a descentralização da produção de conteúdo. Surge assim a oferta de interpretações de fatos que, muitas vezes, contradizem as interpretações dadas pelos meios mais tradicionais (BERTALAN; SERON RUIZ, 2019).

Ray e George (2019) alegam que a própria imprensa depende cada vez mais de mídias sociais para relatar suas histórias. Na visão dos autores isso é um problema devido ao grande volume de conteúdos tendenciosos, subjetivos por natureza, e criados com a intenção de atrair tráfego online. Para Gorkovenko e Taylor (2016), à medida que as mídias sociais se tornaram a plataforma de discussão e campanha política, todos puderam dar sua opinião, mas ficou mais difícil ser ouvido. Dois processos se destacam nesse cenário: A influência online acaba se concentrando em pequenos nichos de jornalistas, políticos e especialistas, e tornam-se mais comuns grupos de pessoas afins, que reforçam as opiniões uns dos outros (GORKOVENKO; TAYLOR, 2019).

Isso vai de acordo com as observações de Garcia et al. (2012) sobre o sucesso de campanhas online. Este não parte apenas de um simples conceito de popularidade, baseado em métricas estacionárias como a quantidade de visualizações, mas também de propriedades dinâmicas das comunidades dirigidas. Em última instância, isso poderia ser definido pela troca de informações entre usuários (GARCIA et al., 2012).

#### 4.1 O USO DO TWITTER NA CÂMARA DOS DEPUTADOS

A presença dos deputados federais no Twitter é muito expressiva, seja na proporção ou no alcance dos parlamentares conectados a plataforma. Esta afirmação se sustenta a partir da análise de dados desta obra. Contudo, o significado cultural e político disso está além de qualquer abordagem técnica, necessitando reflexões e referenciais teóricos adequados. Levando isto em consideração, foram buscadas publicações sobre o papel das novas tecnologias digitais no Poder Legislativo. Pouquíssimas foram encontradas. Na visão de De Faria (2012), a baixa atividade deste campo de estudo decorre da tendência de sobrevalorizar o potencial das tecnologias digitais, e subestimar o papel das instituições. Para não cometer este erro, e entender como se configura a função e a influência dos deputados federais, foi feita uma revisão bibliográfica que reúne fatos históricos e sociológicos sobre o tema.

O princípio da lei como fonte reguladora do poder e das próprias relações sociais é uma noção iluminista, consagrada pela criação de estados democráticos de direito das nações ocidentais modernas, que incluem garantias mínimas à vida, à igualdade e à liberdade (VIEIRA, 2019). Por definição, o processo legislativo expressa nas leis a vontade geral, canalizando diversas perspectivas, visões e valores em arranjos institucionais indispensáveis ao funcionamento da sociedade (DA SILVA, 2010; VIEIRA, 2019). Este ainda exerce uma função de controle, necessária para conter a criação de poderes soberanos (DA SILVA, 2010).

A primeira Câmara de Deputados do Brasil, gerada a partir da Constituição de 1824, era formada por uma elite política a serviço do rei, e decisiva a organização do Império, uma vez que fortalecia a unidade nacional mediando conflitos entre as regiões Norte, Nordeste e Sudeste (SCHWARCZ; STARLING, 2015). Segundo Abranches (2018), ainda que o sistema de representação da época fosse omisso às demandas populares, ao menos as reconhecia a fim de garantir sua legitimidade.

Infelizmente, é preciso destacar que a história do país está repleta de arroubos autoritários, como a ditadura Vargas (1937-1945) e o regime militar (1964-1985), responsáveis pela intermitência de governos democráticos, onde o acesso ao poder se dá através de eleições periódicas, e governantes são responsabilizáveis por meios jurídicos e civis (LAMOUNIER, 2015).

No que toca contemporaneidade, a maior referência de participação popular na política foi a Constituinte de 1987. Schwarcz and Starling (2018) contam que durante

um ano e oito meses o Congresso se transformou no centro da vida pública nacional, reunindo diversas associações e comitês para assentar as bases da democracia através de instituições sólidas e direitos sociais.

Houve também um grande número de publicações acadêmicas sobre o fim do regime militar. Esta produção dividia-se em dois campos: trabalhos ancorados no institucionalismo e aqueles voltados a sociedade civil – ambos fortemente influenciados pela obra de Habermas (PERLATTO, 2012). Não é de se surpreender que a constituinte seja considerada uma síntese de processos teorizados pelo autor, como a formação da esfera pública e do Estado democrático de direito (OLIVEIRA; CANÇADO; PEREIRA, 2010). A Constituição de 1988 trouxe avanços significativos ao campo das políticas públicas, sendo moderna nos direitos, sensível às minorias políticas, avançada nas questões ambientais, e determinada em limitar o poder do Estado sobre o cidadão (ABRANCHES, 2018; SCHWARCZ; STARLING, 2015).

O art. 2º do texto constitucional declara que são poderes da União, independentes e harmônicos entre si, o Legislativo, o Executivo e o Judiciário. Estas três funções fundamentais são exercidas (nos termos e limites da Constituição) por seus respectivos órgãos, a fim de não concentrar poder político e estatal (DA SILVA, 2010)

O Poder Legislativo compreende as regras e os processos que envolvem o ato de produzir leis, sendo exercido pelo Congresso Nacional, composto pela Câmara dos Deputados e pelo Senado Federal (ANDRADE; COUTINHO, 2019). É característica das constituições brasileiras esta estrutura bicameral, onde a Câmara geralmente inicia os projetos de lei e o Senado os revisa (DA SILVA, 2010).

Ao todo são 513 deputados federais e 81 senadores. As eleições da Câmara ocorrem de quatro em quatro anos, a partir de um sistema proporcional, onde os estados devem ter representação equivalente à sua população, não podendo ter menos de oito ou mais de setenta deputados (NICOLAU, 2017). O Senado tem mandatos de oito anos, e realiza suas eleições a partir de um sistema majoritário: três representantes para cada Estado e Distrito Federal (ANDRADE; COUTINHO, 2019).

Da Silva (2010) encara a iniciativa legislativa como o poder de escolha de interesses sociais a serem tutelados pela ordem jurídica, o que denota uma íntima relação entre ideologia e formação das leis. Este encargo assume o antigo desafio de manter uma ordem política e institucional apta a responder a pressões de uma sociedade “marcada por heterogeneidades e discrepâncias” (VIEIRA, 2019, p.96).

Segundo Abranches (2018), a resposta a esta conjuntura, desenvolvida a longo de séculos, resultou no Brasil como única democracia moderna a combinar representação proporcional, multipartidarismo e presidencialismo de coalizão. Este último conceito, essencial a obra do autor, diz respeito a tensão do poder central com os regionais (debatido na seção 3.1 - “O Processo de Informatização do Brasil”).

A história e as regras da política brasileira seriam, em boa parte, advindas de uma relação de “freios e contrapesos” entre os poderes Executivo e Legislativo. O primeiro, exercido pelo presidente, cujo mandato representa a integração dos votos nacionais, advoga pela centralização. O segundo, na forma de deputados e senadores, representantes de interesses estaduais, a favor da descentralização (ABRANCHES, 2018).

Carazza (2018) alerta sobre o custo de funcionamento do presidencialismo de coalizão brasileiro. Para manter sua governabilidade, o presidente da República faz diversas concessões, como nomeação de apadrinhados políticos em cargos importantes e execução de emendas parlamentares. Costa (1995) evidencia outra tendência preocupante: o executivo evita temas potencialmente conflitantes, mesmo que relevantes a sociedade, a fim de preservar seu capital político.

Deste modo, as coalizões governamentais não são montadas a partir de ideais ou projetos de estado, mas por negociatas de votações importantes do Congresso nacional (CARAZZA, 2018). Abranches (2018) acrescenta que a promessa de verbas a parlamentares se torna a moeda de troca pelo voto em matérias de interesse do presidente. Esta situação estaria tão naturalizada que a maioria dos senadores e deputados federais se dedica prioritariamente ao intermédio com a União para obtenção de recursos. Dada esta conjuntura, “o manejo do gasto público sempre favorece os setores politicamente mais fortes, com as clientelas mais poderosas e extensas” (ABRANCHES, 2018, p.360).

Esta problemática traz semelhanças à crítica de Habermas (2015) às democracias representativas. De acordo com o autor, neste tipo de governo a participação política se limita ao recrutamento de quadros de liderança, o que neutraliza a agência dos cidadãos na formação de políticas públicas. A partir desta perspectiva, o intercâmbio entre os domínios comunicativamente estruturados da esfera privada e pública se dá através de “hipotecas de legitimação”, alheias a formação da vontade democrática-competitiva (HABERMAS, 2015, p.353).

Da Silva (2010) aponta que a insatisfação com o processo legislativo tem aumentado nos últimos anos. Os parlamentos seriam considerados corruptos e ineficientes por exercerem seus mandatos à revelia dos eleitores, fazendo parte de jogos de interesses ilegítimos (DE FARIA, 2012). Neste interim destaca-se outra disformidade da política brasileira, o clientelismo - troca de favores pessoais, muitas vezes na forma de benefícios econômicos, por vantagens políticas ou privilégios em serviços providos pelo Estado (PEREIRA, 2008).

Ainda que eticamente condenáveis, estas práticas nem sempre ultrapassam os limites da lei. Carazza (2018) destaca que comissões da Câmara estratégicas para a economia (Finanças e Tributação, Constituição e Justiça, Relações Exteriores), geralmente são dominadas por parlamentares que receberam doações de setores interessados em assuntos que compõem o seu rol de competências.

O surgimento e continuidade deste sistema em parte está relacionado com à impossibilidade de exercer democracia direta nas sociedades de massa. Contudo, novas possibilidades de comunicação direta e instantânea têm mudado o formato do debate político e das esferas públicas, em tese permitindo que a sociedade interaja mais diretamente com os representantes do Estado (DE FARIA, 2012; SILVA, 2012b).

Diante destas possibilidades, mudanças institucionais que preservam o funcionamento eficiente e democrático da Câmara dos Deputados transparecem como necessárias ou mesmo inevitáveis (VIEIRA, 2019). Isto implica em ajustar as organizações governamentais para o desempenho da atividade legislativa, de modo que esta atenda melhor às necessidades e reivindicações da sociedade (DA SILVA, 2010).

Oliveira et al. (2010) reconhecem a formação de uma esfera pública nos moldes habermasianos em experiências como orçamento participativo, conselhos gestores e fóruns temáticos. Contudo, alertam sobre a fragilidades destes mecanismos, alegando ser preciso um esforço maior, seja conceitual ou empírico, para a efetiva legitimação destes espaços. Não são poucas as vozes a favor das tecnologias digitais nos processos de participação cidadã, visto a “facilidade do contato entre representantes e representados, a agilidade na circulação de informações e fortalecimento dos movimentos sociais” (WU, 2018, p.20)

De Faria (2012) assera que a internet redimensionou a atividade parlamentar. Funções consideradas simples hoje, como organizar uma enquete entre eleitores, consistiam antigamente no deslocamento entre cidades ou no registro de cartas e

telefonemas. Ademais, a prestação de contas dos mandatos só era possível através do intermédio de líderes locais ou da mídia de massa, nem sempre sujeitos aos interesses dos eleitores (DE FARIA, 2012). No entanto, como o próprio autor coloca, esta revolução não implicou necessariamente na elevação da democracia.

Santos e Silveira (2001) dão um panorama sobre a crescente complexidade do consumo político nas metrópoles informacionais, ressaltando que demandas puramente eleitorais e clientelistas disputam espaço com as mais gerais e sistêmicas. É inegável que as decisões legislativas afetam de muitas maneiras a vida cotidiana de toda a sociedade, alcançando profissionais, empresas, associações, famílias – e não apenas aqueles que atuam diretamente na política (VIEIRA, 2019).

Representar todos os interesses sociais de forma equânime pode muito bem ser algo impossível. De Faria (2012) discorre sobre o fracasso de muitas iniciativas de participação digital. Neste universo de demandas variadas e descoordenadas, só são respondidas as de curto prazo, ou que de alguma maneira beneficiam seu respectivo parlamentar. Assim, demandas estratégicas e com impacto mais estruturante sobre a vida coletiva geralmente são deixadas de lado (DE FARIA, 2012).

Ainda que seja necessário uma série de rupturas para “atualizar o Legislativo”, um elemento deve permanecer o mesmo: princípios, teoricamente estabelecidos, como o da transparência, o do acesso a informações, e da participação de todos os interessados no debate público (VIEIRA, 2019).

Habermas indicava, muito antes da revolução digital, a necessidade de conectar decisões políticas aos fluxos de comunicação do cotidiano. E isso só viria acontecer se houvesse entendimento genuíno entre a esfera civil e governamental (JERÔNIMO, 2016). Para Pereira (2008), este processo depende de uma série de fatores sócio-históricos relativos a um contínuo processo de nivelamento social.

É possível fazer dois paralelos aqui. O primeiro concerne as condições necessárias à “comunidade ideal de comunicação” da teoria Habermasiana. Wu (2018) as leva em consideração ao afirmar que a qualidade da democracia está sujeita capacidade de debate livres orientados por argumentos racionais. Ou seja, maior investimento em educação, equanimidade etc. O segundo, sobre a qualidade do acesso à internet e alfabetização digital. Apesar dos diversos obstáculos a serem superados, o poder público vem se integrando a internet. Dados da pesquisa TIC Governo Eletrônico 2019 indicam que 85% de órgãos públicos federais e estaduais e 82% das prefeituras tem perfis nas mídias sociais (CGI.BR, 2019).

Silva (2012b) estuda perfis de deputados federais da 54ª Legislatura no Twitter. As respectivas eleições, realizadas em 2010, trouxeram uma mudança de compreensão por parte dos políticos brasileiros. Esta mídia social não serviria para estender a influência obtida por meios de comunicação tradicionais. Era preciso saber usá-la para obter reconhecimento e visibilidade (SILVA, 2012b). A autora faz uma série de observações a partir de sua base de dados: muitos deputados abandonam o Twitter depois do período de campanha; deputados bem relacionados em seu estado, e com elevadas votações, não necessariamente tem a mesma audiência no Twitter; deputados mais jovens tendem a ser mais participativos; e há coincidência entre liderança no Parlamento e liderança no Twitter. Estas observações foram pertinentes a esta obra, pois deram um ponto de referência que possibilitou a comparação de resultados e avaliação de quanto o uso desta mídia social mudou ao longo de uma década. Antes de chegar a estes resultados, será apresentado um breve estudo documental sobre iniciativas de transparência do Congresso Nacional e de terceiros.

## **4.2 A TRANSPARÊNCIA DE DADOS PÚBLICOS NA CÂMARA DOS DEPUTADOS**

Já houve iniciativas por parte da Câmara dos Deputados para inteirar a população nos processos que regulam a vida coletiva. De Faria (2012) cita o “e-Democracia”, que não gerou o resultado esperado devido à baixa escala de participação popular. O autor levanta uma série de hipóteses para explicar a descontinuidade deste e de outros projetos semelhantes: custo financeiro exacerbado; inaptidão para operacionalizar o desenvolvimento de novas tecnologias no ritmo do mercado; e resistência de “grupos políticos mais conservadores” à transparência e interação digital (DE FARIA, 2012, p.101). Pereira (2008) afirma que a introdução de mecanismos de *accountability* é particularmente difícil no Brasil devido a histórica falta de distinção entre público e privado, e cultura política que garante certa impunidade aos representantes populares.

Seja qual for o motivo, a Câmara optou, no início da década de 2010, por delegar estas iniciativas à sociedade civil. Esta escolha pode parecer displicente, mas como coloca De Faria (2012), acabou por ser a mais sensata. Institucionalizar canais virtuais de participação diante da ubiquidade de plataformas como o Facebook e o Twitter seria apenas um desperdício de recursos públicos, na medida que é possível

estimular o interesse da população pelo debate político a partir delas, como muitos deputados já vinham fazendo (DE FARIA, 2012). Atualmente o *site* da Câmara dos Deputados dispõe de diversas informações sobre atividades parlamentares. Esta não é uma iniciativa isolada, visto que o governo federal, assim como os estaduais e municipais desenvolveram plataformas digitais de colaboração para dar visibilidade, informar e promover a participação política on-line (JERÔNIMO, 2016).

**Figura 1 – Painel de resultados legislativos da Câmara dos Deputados**



**Fonte: Câmara dos Deputados<sup>4</sup>**

No primeiro painel do site é possível conferir os resultados legislativos de qualquer deputado. Nele se encontram visualizações de dados sobre “Áreas temáticas mais votadas”, “Plenário” e “Comissões”. Nos três casos há a opção de “Ver todas”, que leva uma lista dos projetos de leis respectivos.

<sup>4</sup> Disponível em: <https://www.camara.leg.br/transparencia/resultados-legislativos> . Acesso em: 15 mai. 2021.

**Figura 2 - Painel de orçamento da Câmara dos Deputados**



**Fonte: Câmara dos Deputados<sup>5</sup>**

No painel referente ao gasto de dinheiro público, há mais opções para buscas personalizadas, permitindo que o usuário filtre por deputados, estados e partidos.

**Figura 3 - Painel de Recursos Humanos e Concursos da Câmara dos Deputados**



**Fonte: Câmara dos Deputados<sup>6</sup>**

Na última seção há um gráfico no estilo *treemap* com dados sobre a distribuição de “Recursos Humanos e Concursos”. Apesar deste tipo de gráfico ser utilizado para

<sup>5</sup> Disponível em: <https://www2.camara.leg.br/transparencia/receitas-e-despesas> . Acesso em: 15 mai. 2021.

<sup>6</sup> Disponível em: <https://www.camara.leg.br/transparencia/recursos-humanos>. Acesso em: 15 mai. 2021.

buscas segmentadas, o usuário não pode esmiuçar mais as distribuições apresentadas.

A Câmara dos Deputados também adere a uma política de Dados Abertos. Isso significa, em teoria, que qualquer informação referente ao trabalho dos parlamentares está disponível ao público. É possível fazer uma coleta manual dos dados, baixando os respectivos arquivos por ano, ou utilizar a Aplicação de Interface de Programação (API), desenvolvida e documentada pela equipe técnica da instituição.

O *site* da Câmara possui uma seção com projetos de terceiros que utilizam essa interface. Até o momento 12 projetos são listados. Destes, 10 tem o objetivo de informar o público por meio de *newsletters* e outras formas de conteúdo personalizado. Dois são semelhantes ao propósito desta dissertação. O aplicativo Kan, ainda em fase de desenvolvimento, permite acompanhar gastos públicos. Já o projeto Atlas Político disponibiliza diversas informações de parlamentares, que podem ser buscados a partir de parâmetros como nome, cargo, partido e localidade.

**Figura 4 – Painel de Classificação de Deputados do Atlas Político**



**Fonte: Atlas Político<sup>7</sup>**

O *site* também apresenta um sistema de ranqueamento que considera 5 fatores. São estes: Representatividade, Campanha Responsável, Ativismo Legislativo, Debate Parlamentar e Fidelidade Partidária.

<sup>7</sup> Disponível em: atlasintel.org. Acesso em: 15 mai. 2021.



A deliberação de projetos como este é um ponto importante, visto que os mecanismos de prestação de contas só se tornam realidade quando a cidadania tem acesso às informações e a documentação relacionada com as ações de seus representantes, “seja para aferir como eles se comportam em suas funções legislativas, na aprovação de leis e de planos, seja na forma como direcionam o uso do patrimônio público no intuito de atender as necessidades de seus eleitores e da população como um todo (PEREIRA, 2008, p.42).

Destaca-se aqui a importância de conectar dois campos: academia e sociedade civil. Apesar do “jogo legislativo” ser público, pouquíssimas pessoas tem o conhecimento necessário para julgar ações e resultados de seus representantes, pois os procedimentos parlamentares geralmente são herméticos e intrincados para o observador comum (VIEIRA, 2019).

Segundo Perlatto (2012), a ampliação de uma esfera pública renovada depende tanto da institucionalização de procedimentos democráticos capazes de dar vazão a argumentos críticos racionais, quanto destes mecanismos, que a tornam mais porosa aos valores, demandas, reivindicações e manifestações da sociedade.

Vieria (2019) afirma que a transparência do processo decisório do Legislativo, proveniente a divulgação e análise de dados do Congresso Nacional, contribui não apenas a teorias sobre parlamento, Direito e democracia, mas também a participação individual coletiva no debate sobre leis e políticas públicas. Levando isto em consideração, o estudo das interfaces oficiais da Câmara dos Deputados e do Atlas Políticos auxiliou a conceituar um modelo de análise próprio, que será apresentado no próximo capítulo.

## **5 ANALISANDO A ATUAÇÃO PARLAMENTAR E O USO POLÍTICO DO TWITTER NO ANO DE 2020**

Feito o devido aprofundamento teórico, deve-se responder quão consistentes são os discursos dos deputados federais no Twitter, se comparados às suas proposições parlamentares, principal problema de pesquisa aqui proposto. Manovich (2009) afirma que é impossível entender desenvolvimentos culturais globais usando ferramentas e métodos teóricos do século XX. A solução estaria na análise massiva de dados. Prática esta que começa nos anos 1990, a partir dos primeiros motores de busca na Web, se consolida na década de 2010, junto a crescente base de usuários das mídias sociais, e hoje é realizada em tempo real por toda grande empresa (MANOVICH, 2018).

Para Santos e Silveira (2001), a demanda por mercadorias informacionais constituiu o chamado “setor quaternário”, voltado ao processamento de dados, publicidade e propaganda. Os autores, oriundos da geografia, também seguem esta tendência, avaliando que a noção de território não pode mais ser interpretada somente por métricas tradicionais, como a densidade demográfica, pois passou demonstrar densidades “técnicas, informacionais, normativas e comunicacionais” (Milton Santos, 2001, p.296).

Considera-se que a multidimensionalidade da atuação parlamentar, discutida nos capítulos anteriores, se encaixa perfeitamente nesta definição. Portanto, técnicas de análise de dados, provenientes da linguagem de programação Python, foram empregadas no decorrer desta obra. O fator em comum na grande variedade de códigos utilizadas está na sua anuência à estatística.

Segundo Agresti e Finlay (2012), o uso da estatística nas ciências sociais se resume a um conjunto de métodos para obter e analisar características de interesse sobre determinado assunto que, na definição dos autores, pode ser descrito como “dados”. A primeira etapa à aplicação destes métodos está no “delineamento”, que concerne o ato de planejar a coleta e a organização dos dados. A partir deste ponto, há dois tipos de análise estatística voltada à fenômenos sociais: inferencial e descritiva. Previsões feitas usando dados se encaixam na inferência estatística. Representar resultados encontrados, invés de listar todas as informações disponíveis, está dentro da estatística descritiva. O surgimento deste campo de estudo se justifica pelo fato de que dados brutos não são fáceis de entender. É preciso resumi-los a sua

forma mais simples através de gráficos, tabelas ou métricas, sem que estes percam seu significado (AGRESTI; FINLAY, 2012). Esta dissertação atua apenas no âmbito descritivo. O que não implica em tentativas de medir a eficácia dos mandatos. Considera-se que a atuação parlamentar não pode ser resumida a quantidade de proposições aprovadas, ou outras métricas facilmente quantificáveis (ANDRADE; COUTINHO, 2019). Ainda que este trabalho levante índices e sistemas de classificação, estes não são utilizados de forma taxativa, tendo como principal objetivo remeter o problema de pesquisa proposto.

Grande parte da análise de dados se deu através de dois arquivos<sup>8</sup> em formato *notebook*, que facilita a escrita e execução de códigos. O primeiro concerne a coleta e o processamento dos dados, e é apresentado aqui como Apêndice A. O segundo traz métodos voltados a análise e visualização de dados, e está no Apêndice B.

A próxima seção, referente a metodologia, apresenta um resumo das técnicas utilizadas. Nela, cada unidade de código é referenciada como “célula”, seguido de sua numeração, e nome do apêndice quando necessário. Cabe destacar que o conteúdo dos *notebooks* não é expresso de modo sequencial, tendo em vista a grande quantidade de informações técnicas e o fato de que a ordem das células nem sempre interfere na obtenção dos resultados. Deste modo, são apresentados os pontos principais à formação do *corpus*, assim como as dificuldades e soluções encontradas fora do âmbito dos *notebooks*, de forma a ilustrar esta etapa do desenvolvimento da pesquisa.

Em “Perfis Parlamentares”, há uma aproximação geral à base de dados a partir de perfis dos deputados estudados. Estes são apresentados através de tabelas, e foram elaborados pela sistematização de informações demográficas, como gênero, etnia, idade e ideologia. Com maior ciência sobre a composição da Câmara, se parte para o subcapítulo “Consistência Temática”. Nele é destacado o nível de consistência entre temas recorrentes no âmbito parlamentar e na mídia social. Para isso foram gerados gráficos de barra e de radar, nuvens de palavras, e outros tipos de visualizações de dados voltadas a destacar o aspecto semântico da amostra.

Próximo a conclusão, o objetivo se torna apontar a relação entre redes de proposições legislativas e as formadas a partir de *tweets*. Na seção “Formação de Redes”, tanto a metodologia quanto a forma de representar visualmente as

---

<sup>8</sup> Disponível em: <https://github.com/lorenzo-leuck/esfera-publica-na-era-da-informacao>

informações muda ligeiramente. Continua-se no campo da análise descritiva, porém se acrescenta métricas e conceitos da Análise de Redes Sociais (ARS) e da teoria dos grafos. Neste ponto, o foco vai do conteúdo discursivo ao registro das interações nos meios digitais e legislativos.

## 5.1 METODOLOGIA

A coleta dos dados consistiu primeiramente no *download* de todos os materiais relevantes aos objetivos de pesquisa no site da Câmara dos Deputados<sup>9</sup>. Os arquivos, em formato csv, foram baixados de modo manual através dos links dispostos na Célula 0.5 do Apêndice A. Desta maneira, a amostra inicial continha seis fontes de dados: “deputados.csv”, “proposicoes-2020.csv”, “proposicoesAutores-2020.csv”, “proposicoesTemas-2020.csv”, “votacoesProposicoes-2020” e “votacoes-2020.csv”.

O arquivo “deputados.csv” serviu como ponto de partida por conter informações básicas como nome, gênero e idade de todos deputados federais do Brasil, desde os tempos da primeira república. Entretanto, filtrando a base à parlamentares da 56ª Legislatura (2019-2023), obteve-se 560 registros. O correto seria 513 (PORTAL DA CÂMARA DOS DEPUTADOS, 2021a). A solução encontrada consistiu em baixar as bases de dados do Tribunal Superior Eleitoral (TSE). Acrescentam-se assim duas amostras: “candidatos2018.csv” e “bensCandidatos2018.csv”, cujos *links* também se encontram na Célula 0.5 do Apêndice A.

Para usá-las, primeiramente foi preciso filtrar “candidatos2018.csv” à deputados federais eleitos, procedimento feito na Célula 2.1. A base não só tinha o número correto de registros, como continha informações significativas ao desenvolvimento da pesquisa, como partido, escolaridade, estado civil, etnia e ocupação anterior.

Dois campos exigiram etapas adicionais à fruição da análise. A quantidade de votos, distribuídas por zonas eleitorais na base do TSE, foi alvo de funções de agrupamento e soma através dos registros únicos de cada candidato, como disposto no código da Célula 2.2. O valor dos bens declarados por cada deputado teve tratamento semelhante, como pode ser visto na Célula 2.10. Pelo fato desta informação partir de um arquivo separado, foi necessário executar uma função de pareamento na célula 2.11.

---

<sup>9</sup> Disponível em: <https://dadosabertos.camara.leg.br/>

As bases de dados do TSE e da Câmara foram combinadas através de uma função que une listas de dados a partir de semelhanças de texto, presente na célula 2.6. Ainda havia um problema. As eleições municipais de 2020 trouxeram mudanças consideráveis ao quadro da Câmara. Muitos deputados entraram de licença para se candidatar ou ajudar colegas de partido. Ou seja, o resultado das eleições ocasionou na abdicação de mandatos, já que alguns se elegeram e outros passaram a exercer cargos de confiança. Portanto, a base do TSE estaria desatualizada.

Para obter a situação dos mandatos, a API da Câmara foi utilizada a partir de uma função customizada no programa Google Sheets<sup>10</sup>. Assim, registros de deputados que abdicaram de seus mandatos foram substituídos pelo dos seus suplentes (Célula 2.3 e 2.4). Esta última informação não estava disponível nos arquivos nem na API, tendo que ser buscada manualmente no site da Câmara (PORTAL DA CÂMARA DOS DEPUTADOS, 2021b).

**Tabela 1 – Efetivação de Deputados Federais em 2020**

Deputados afastados	Deputados efetivados
Ney Leprevost	Evandro Rogerio Roman
Sandro Alex	Reinhold Stephanes Junior
Paulo Foletto	Ted Conti
Onyx Lorenzoni	Marcelo Costa
Manuel Marcos	Leo de Brito
Nelson Pellegrino	Joseildo Ribeiro Ramos
Roberto Pessoa	Danilo Forte
Josias Gomes	Paulo Magalhães
Fábio Faria	Carla Dickson
Covatti Filho	Ronaldo Santini
Tereza Cristina	Bia Cavassa
Jean Wyllys	David Miranda
Margarida Salomão	Aelton José de Freitas
Eduardo Braide	Josivaldo dos Santos Melo
Pedro Paulo	Marcos Bezerra Ribeiro Soares
João H. Campos	Milton Coelho
Sergio Vidigal	Neucimar Ferreira Fraga

<sup>10</sup> Disponível em: <https://github.com/lorenzo-leuck/chamada-apis-goglesheets>

Alexandre Serfiotis	Pedro Augusto Palareti
João Henrique Caldas	Pedro Torres Brandão Vilela
Edmilson Rodrigues	Viviane da Costa Reis

**Fonte: Câmara dos Deputados, TSE e Twitter**

Na Tabela 1 se encontra a relação entre deputados afastados, seja por licença ou cassação de mandato, e os efetivados, temporariamente ou de modo definitivo, no ano de 2020. A API também foi utilizada para conseguir o partido atual dos deputados, visto que houve mudanças no quadro desde 2018. O *site* Congresso em Foco (SARDINHA; COSTA, 2019) foi a referência para situar o espectro ideológico em “esquerda”, “centro” e “direita”. Estas informações foram pareadas na base de dados a partir da Célula 2.12.

Ainda que o arquivo original “deputados.csv” contivesse um campo destinado a *links* de mídia social, este não distinguia entre os perfis de cada plataforma, e em muitos casos estava desatualizado. Portanto, assim que constituída a lista definitiva de deputados, o *script* da Célula 2.7 foi rodado para encontrar seus respectivos perfis no Twitter. Os resultados foram comparados com um levantamento feito pelo site Auditoria Cidadã (2020). Todos os registros de deputados sem perfis no Twitter foram conferidos posteriormente por buscas no Google.

Depois de encontrar todos os perfis oficiais, uma série *scripts* acessou a API do Twitter para registrar o número de seguidores e de publicações dos deputados (Célula 2.8), e armazenar os *tweets* com suas respectivas métricas de compartilhamentos e curtidas (Célula 5.1). Este procedimento resultou em um arquivo *json* por conta, contendo até 3 mil publicações. Na Célula 5.2 estes são concatenados em um único arquivo, intitulado “tweets2020.csv”.

Todos os códigos de coleta de dados nesta mídia social partiram de funções da biblioteca *tweepy*, e adaptações do repositório *TweetUtils*<sup>11</sup>. Estes trechos do trabalho só funcionam se a Célula 0.4 for executada com chaves de API do Twitter, que podem ser requisitadas no portal de desenvolvimento da plataforma<sup>12</sup>.

O processamento dos arquivos “proposicoes-2020.csv”, “proposicoesAutores-2020.csv”, e “proposicoesTemas-2020.csv”, essencial à seção “Consistência Temática”, foi a que mais apresentou dificuldades técnicas. Antes de partir ao manejo

<sup>11</sup> Disponível em: <https://github.com/DAVINTLAB/TweetUtils>

<sup>12</sup> Disponível em: <https://developer.twitter.com/en/portal/petition/academic/is-it-right-for-you>

dos dados, foi preciso entender, a partir de fundamentação teórica, do que estes se tratam.

De acordo com Andrade e Coutinho (2019), proposição parlamentar é toda matéria sujeita à deliberação na Câmara dos Deputados. Nem toda proposição resulta em norma jurídica, porém toda proposição é o instrumento no qual os parlamentares formalizam suas ideias e propostas. Atualmente, existem oito tipos de proposições: “a) proposta de emenda à Constituição; b) projeto; c) emenda; d) indicação; e) requerimento; f) recurso; g) parecer; e h) proposta de fiscalização e controle” (ANDRADE; COUTINHO, 2019, p.43).

Devido a impossibilidade de ler e ordenar as milhares de proposições presentes na amostra, esta dissertação se valeu do sistema de classificação oficial da Câmara dos Deputados, expresso nos campos “TEMA” e “KEYWORDS”. O primeiro passo para utilização dos dados ocorreu na Célula 3.1, que combinou os três arquivos a partir do identificador “idProposicao”. Isso não foi suficiente, pois cada arquivo possuía um ordenamento próprio.

Em “proposicoes2020.csv” (29.626 linhas), ficam todas as proposições do ano, porém há múltiplos registros para cada uma, devido diferentes redações dos textos no decorrer dos trâmites legislativos. Este é o arquivo que detém o campo “KEYWORDS”. Somente 5.594 linhas (18,88%) o tem preenchido. Neste conjunto, Projeto de Lei (PL) e Projeto de Decreto Legislativo (PDL) são as categorias mais recorrentes.

O arquivo “proposicoesTemas2020”, que contém o campo “TEMAS”, conta com 18.910 linhas, porém o número de proposições únicas é 8.616. Projeto de Lei (PL) é o tipo de proposição que mais se repete, seguido por Requerimento de Informação (RIC). Cabe destacar aqui que o conjunto de proposições com “KEYWORDS” está contido no de temas. Ou seja, toda linha com “KEYWORDS” no arquivo “proposicoes2020.csv” tem uma correspondente no “proposicoesTemas2020.csv”, sendo que o inverso não se aplica.

Já “proposicoesAutores-2020.csv” se mostrou o arquivo mais volumoso, com 55.360 linhas. Esta diferença de tamanho se explica pela forma como dispõe suas informações. Se, por exemplo, um projeto de lei tiver cinco autores, este será representado em cinco linhas.

Para solucionar as diferentes ordenações de dados foram executadas, na Célula 3.2, funções de agrupamento e de filtro que retornaram o arquivo “db-

proposicoes.csv”. Este, formado apenas por proposições únicas que tivessem o campo dos temas preenchido (e consequentemente o de *Keywords*), e com os registros de coautoria separados por vírgula.

Comparado a outras amostras, o processamento de “votacoesProposicoes-2020.csv” e “votações.csv” se mostrou relativamente simples. O primeiro arquivo continha o resultado das votações feitas em 2020, o segundo a identificação do que foi votado. Deste modo, na Célula 4.2 se executa uma função para combiná-los.

Diversos arquivos compuseram o fluxo de trabalho da análise quantitativa. É importante destacar que cada arquivo, individualmente, não seria capaz de entregar os resultados esperados. A articulação da metodologia utilizada partiu primeiramente da conexão entre estas diferentes partes para formar bases de dados de fácil consulta e adaptáveis à filtros, agrupamentos e diversas outras funções necessárias ao decorrer desta obra.

**Figura 7 - Conexões entre as bases de dados**



**Fonte: Câmara dos Deputados, TSE e Twitter**

A Figura 7 ilustra o processo de concatenação do material, não representando em *stricto sensu* a ordem das operações. O primeiro retângulo se refere a tabela de deputados montada a partir do arquivo e da API da Câmara. Através das colunas “NM\_URNA” e “Nome Deputado” se estabeleceu uma conexão com “deputados-tse”, que proveu o número correto de registros a base de dados. A ligação com “ideologia-partido” se deu por “PARTIDO\_ATUAL” e “Partido”. Já “tweets-2020” foi conectado através dos campos “PERFIL\_TWITTER” e “screen\_name”. Por meio da coluna

“ID\_DEPUTADO”, forma-se uma conexão com “proposições-2019”. Esta, por sua vez, está ligada com “proposicoesAutores-2020”, “proposicoesTemas-2020” e “votacoesProposicoes-2020” por meio de “ID\_PROPOSICAO”. A última tabela estabelece uma conexão com “votacoes-2020.csv” através de “ID\_VOTACAO”.

Isto resume os procedimentos realizados na etapa de coleta e processamento de dados. No restante do capítulo serão descritas etapas adicionais, desenvolvidas e executadas através do *notebook* de análise e visualização de dados, presente no apêndice B.

## 5.2 PERFIS PARLAMENTARES

Este subcapítulo levanta o perfil de deputados federais, tanto em um âmbito geral quanto no subconjunto do Rio Grande do Sul, a partir da sistematização das informações demográficas presentes na amostra estudada. De Faria (2012) afirma que para entender o sistema representativo é preciso estudar o comportamento dos parlamentares, sujeito a pressões e influências de diversas naturezas. Este trabalho dá um passo atrás, representando quem são os parlamentares antes de procurar por padrões no seu comportamento. É importante tomar ciência dos diferentes grupos que compõe a Câmara devido a uma especificidade em voga “desde os tempos do império”: a maioria, já estabelecida no condomínio do poder, tende a manter posição e usá-la na defesa dos seus interesses (ABRANCHES, 2018).

As Características predominantes da atual legislatura trazem o seguinte perfil: homem (84,80%), branco (75,63%), entre 40 e 55 anos (40,54%), casado (69,98%), com ensino superior completo (55,36%) e ocupação anterior de deputado (44,25%). Apenas 78 das parlamentares (15,2%) são mulheres. O número de deputados autodeclarados pardos, negros, amarelos e indígenas é 125, o equivalente a 24,37% da Câmara.

Neste conjunto somam-se 99 deputados pardos (19,30%), 23 negros (4,48%), 2 amarelos (0,39%) e 1 indígena (0,19%). Joênia Wapichana (Rede-RR) é a primeira mulher indígena no cargo de deputada federal (PORTAL DA CÂMARA DOS DEPUTADOS, 2018a). A parlamentar foi eleita com 8.491 votos, menor resultado da eleição de 2018. Estes índices representam uma diversidade maior do que a da legislação anterior (PORTAL DA CÂMARA DOS DEPUTADOS, 2018a, 2019).

Outro fator de destaque é o percentual de novos deputados em atividade. A Câmara dos Deputados divulga que o índice de renovação parlamentar da 56ª Legislação foi de 47,30%, o maior desde a redemocratização do país (PORTAL DA CÂMARA DOS DEPUTADOS, 2018b). A mudança do quadro parlamentar, em decorrência das eleições municipais de 2020, elevou o índice para 48,73%.

Dos 33 partidos registrados atualmente, 24 possuem deputados eleitos. O PT e o PSL estão empatados no topo com 53 deputados (10,33% cada). É importante frisar aqui que a Constituição concentrou o controle do processo legislativo na presidência da Câmara e do Senado e nas lideranças partidárias (ABRANCHES, 2018). Contudo, o sistema de partidos não expressa necessariamente uma divisão profunda no campo ideológico ou na estrutura social, apresentando diferenças programáticas que poderiam ser traduzidas em um número menor de legendas (NICOLAU, 2017). A distribuição do espectro ideológico desta legislação, gerada a partir da classificação do site Congresso em Foco, aponta que 273 deputados (53,22%) são de direita, 142 (27,68%) de esquerda e 98 (19,10%) de centro.

Um último dado importante para este levantamento é a porcentagem de deputados que possuem conta no Twitter: 96,30%. Isso confirma a possibilidade de interações e debates envolvendo quase todo o parlamento na plataforma. Entretanto, a predominância do número de contas registradas não implica no uso ou na importância do Twitter no contexto da Câmara dos Deputados, visto que diversos parlamentares tiveram poucos *tweets* ou sequer publicaram durante o período estudado.

**Tabela 2 – Características gerais da Câmara dos Deputados**

Gênero	Homem (435    84.8%)
Cor/Raça	Branca (388    75.63%)
Faixa etária	Entre 40 e 55 (208    40.55%)
Estado Civil	Casado (359    69.98%)
Escolaridade	Superior (284    55.36%)
Ocupação anterior	Deputado (227    44.25%)
Partido	PSL (53    10.33%)

Ideologia	Direita (273    53.22%)
Tem Twitter	Sim (494    96.3%)
Estado	SP (70    13.65%)
Região	Sudeste (179    34.89%)
Reeleição	Sim (263    51.27%)

**Fonte: Câmara dos Deputados, TSE e Twitter**

A Tabela 2 reúne as informações apresentadas anteriormente, evidenciando o número de deputados e porcentagens específicas a cada característica. O método utilizado aqui consiste em funções de moda para encontrar os valores mais recorrentes nas colunas (Célula 1.3 – Apêndice B). Para aplicá-lo ao subconjunto de deputados gaúchos, foi executada uma função de agrupamento, que redistribui a base de dados a partir de suas respectivas células (Célula 1.2). Outra etapa foi necessária, visto que quatro colunas são numéricas (Valor dos bens declarados, quantidade de votos, número de tweets e número de seguidores). Lidar com estes campos exigiu o uso da função *qcut*, que gera grupos igualmente distribuídos a partir das faixas de dados escolhidas (Célula 1.05). Na Tabela 3 se encontram as definições dos novos compartimentos.

**Tabela 3 – Compartimentos Numéricos**

Compartimento	Valor
Declaração de bens	
Até R\$38 milhões (G1)	R\$1.892.000 A R\$38.844.000
Até R\$1 milhão e 892 mil (G2)	R\$938.000 A R\$1.892.000
Até R\$938 mil (G3)	R\$407.000 A R\$938.000
Até R\$407 mil (G4)	R\$900 A R\$407.000
Votos	
Até 1 milhão e 843 mil (G1)	115.000 a 1.843.000
Até 115 mil (G2)	87.000 a 115.000
Até 87 mil (G3)	60.000 a 87.000

Até 60 mil (G4)	8.000 a 60.000
Seguidores no Twitter	
Até 1 milhão 988 mil (G1)	15.350 a 1.988.000
Até 15 mil e 350 (G2)	4.200 a 15.350
Até 4 mil e 200 (G3)	1.150 a 4.200
Até 1.150 (G4)	0 a 1.150
Número de tweets	
Até 114 mil e 966 (G1)	8.475 A 114.966
Até 8 mil e 475 (G2)	2.849 A 8.475
Até 2 mil e 849 (G3)	586 a 2.849
Até 586 (G4)	0 a 586

**Fonte: Câmara dos Deputados, TSE e Twitter**

Estes procedimentos possibilitam a exploração dos dados a partir de qualquer segmento escolhido. Seguindo a linha de pesquisa proposta por esta dissertação, a amostra foi filtrada pelo campo “ESTADO\_ELEICAO”, tendo como parâmetro “RS” - isto é, por deputados federais eleitos pelo estado do Rio Grande do Sul (Célula 1.4).

**Tabela 4 – Perfil dos deputados gaúchos**

Gênero	Homem (28    90.32%)
Cor/Raça	Branca (31    100.0%)
Faixa etária	Entre 55 e 70 (13    41.94%)
Estado Civil	Casado (18    58.06%)
Escolaridade	Superior (14    45.16%)
Ocupação anterior	Deputado (17    54.84%)
Partido	PT (5    16.13%)
Ideologia	Direita (14    45.16%)
Tem Twitter	Sim (29    93.55%)

Reeleição	Sim (18    58.06%)
Declaração de bens	Até R\$407 mil (G4) (11    35.48%)
Votos	Até 115 mil (G2) (16    51.61%)
Seguidores no Twitter	Até 4 mil e 200 (G3) (13    41.94%)
Número de Tweets	Até 114 mil e 966 (G1) (13    41.94%)

**Fonte: Câmara dos Deputados, TSE e Twitter**

Percebe-se, através da Tabela 4, que as características gerais deste subconjunto não diferem substancialmente as do âmbito geral. Em certos casos, pronunciam-se ainda mais. A porcentagem de homens eleitos fica em 90,32% (28 de 31 deputados). No quesito “Cor/Raça: Branca”, chega a 100%, marcando zero diversidade racial na 56ª legislatura.

Entre os fatores distintos encontram-se a faixa etária, onde prevalece deputados mais velhos (Entre 55 e 70 anos) e PT como maior partido, tendo 5 de 31 deputados (16,13%). Direita continua sendo a ideologia dominante, ainda que tenha um percentual menor do que do quadro geral (45,16%). Nota-se outras diferenças sutis, como menor proporcionalidade de deputados com ensino superior (45,16%) e maior número relativo de recorrência no cargo (58.06%).

Os compartimentos numéricos contam que há mais deputados no menor quadrante de renda (35.48% com até R\$407 mil em bens declarados). A prevalência do segundo maior grupo em número de votos (51.61% com até 115 mil) é um reflexo do alto índice de recorrência parlamentar, evidenciando que deputados previamente eleitos tendem a manter ou aumentar suas bases.

93,55% dos deputados deste subconjunto têm conta no Twitter, percentual 3 pontos menor que o do âmbito geral. O uso da plataforma é um fator determinante, visto que 41.94% despontam a maior métrica possível, com até 114 mil e 966 *tweets* publicados. Contudo, a influência destes parlamentares não se destaca, pois 13 de 31 (41.94%) tinham até 4 mil e 200 seguidores no momento da coleta, segundo pior compartimento desta categoria.

Concluído o levantamento do perfil da bancada gaúcha, cabe esclarecer alguns conceitos utilizados daqui para frente. Como foi dito na seção sobre a metodologia, a lista sobre a orientação ideológica dos deputados veio de um levantamento do site Congresso em Foco. Por meio dos métodos aqui apresentados, é possível basear noções sobre “Esquerda”, “Centro” e “Direita” a partir de características gerais decorrentes do levantamento de perfis.

**Tabela 5 – Perfis por ideologia**

	Centro	Direita	Esquerda
Gênero	Homem	Homem	Homem
Cor/Raça	Branca	Branca	Branca
Estado Civil	Casado	Casado	Casado
Escolaridade	Superior	Superior	Superior
Ocupação anterior	Deputado	Deputado	Deputado
Faixa etária	Entre 40 e 55	Entre 40 e 55	Entre 40 e 55
Partido	MDB	PSL	PT
Estado	MG	SP	SP
Região	Sudeste	Sudeste	Nordeste
Reeleição	Sim	Não	Sim
Declaração de Bens	Até R\$1 milhão e 892 mil (G2)	Até R\$38 milhões (G1)	Até R\$938 mil (G3)
Votos	Até 60 mil (G4)	Até 115 mil (G2)	Até 1 milhão e 843 mil (G1)
Seguidores no Twitter	Até 4 mil e 200 (G3)	Até 1.150 (G4)	Até 1 milhão 988 mil (G1) Até 15 mil e 350 (G2)
Número de Tweets	Até 2 mil e 849 (G3)	Até 586 (G4)	Até 114 mil e 966 (G1)

**Fonte: Câmara dos Deputados, TSE e Twitter**

Os perfis por ideologia não quebram o conjunto de características referente à Gênero, Cor/Raça, Estado Civil etc. No entanto, alguns fatores ainda podem ser destacados. Os partidos mais frequentes são MDB, PSL e PT. A esquerda é o único campo onde a maioria não é milionária e há os maiores valores possíveis em votos e métricas digitais. A direita, o único onde predominam deputados de primeiro mandato

e maior compartimento de bens (Até R\$38 milhões). O agrupamento de votos que se destaca é o G2 (Até 115 mil) e as métricas digitais tem os piores agrupamentos possíveis. Já o centro desponta o último compartimentos de votos (Até 60 mil), mas supera a direita nos quadrantes do Twitter.

A partir da Tabela 5 é possível afirmar que proporcionalmente à Esquerda tem mais “influenciadores digitais” do que a direita. Não seria correto dizer o mesmo com números absolutos, visto que na distribuição de seguidores G1 consta: Direita (58 / 47,54%), Esquerda (54 / 44,26%) e Centro (10 / 8,19%). Muito menos concluir que a Esquerda domina o Twitter. Mesmo que este campo ideológico tivesse o maior valor absoluto, faltaria considerar que o nível de influência é diferente para cada deputado.

**Tabela 6 – Deputados com maior número de seguidores no Twitter - Geral**

	Nome	Seguidores	Ideologia
1	Eduardo Bolsonaro	1.988.692	Direita
2	Marcelo Freixo	1.526.223	Esquerda
3	Carla Zambelli	1.138.247	Direita
4	Rodrigo Maia	953.604	Direita
5	Gleisi Hoffmann	779.286	Esquerda
6	Bia Kicis	763.075	Direita
7	Pr. Marco Feliciano	666.725	Direita
8	Aécio Neves	636.175	Centro
9	Sargento Fahur	618.711	Direita
10	Kim Kataguirí	608.447	Direita

**Fonte: Câmara dos Deputados, TSE e Twitter**

A Tabela 6 mostra os dez parlamentares com o maior número de seguidores. Em primeiro lugar está Eduardo Bolsonaro (PSL), e em segundo Marcelo Freixo (PSOL). A distribuição de ideologia aqui é: Direita (7), Esquerda (2) e Centro (1).

**Tabela 7 – Deputados com maior número de seguidores no Twitter – Rio Grande do Sul**

	Nome	Seguidores	Ideologia
1	Maria do Rosário	399.436	Esquerda
2	Marcel van Hattem	367.196	Direita
3	Paulo Pimenta	362.908	Esquerda

4	Osmar Terra	334.959	Centro
5	Bibo Nunes	175.762	Direita
6	Fernanda Melchionna	121.852	Esquerda
7	Henrique Fontana	63.158	Esquerda
8	Bohn Gass	42.769	Esquerda
9	Sanderson	37.229	Direita
10	Pompeo de Mattos	15.609	Esquerda

**Fonte: Câmara dos Deputados, TSE e Twitter**

Na Tabela 7 são apresentados os dez deputados gaúchos com o maior número de seguidores no Twitter. Maria do Rosário (PT) está em primeiro, e Marcel van Hattem (NOVO) em segundo. Na distribuição ideológica a Esquerda fica em destaque (6), seguida pela Direita (3), e Centro (1). Mais uma vez é preciso reavaliar conclusões vagamente baseadas em dados. Estes resultados não indicam que a direita domina o quadro geral da Câmara, e a esquerda o âmbito Estadual. Posteriormente será mostrado, a partir de dados de publicações e formações de rede, que o desempenho e o alcance dos espectros ideológicos variam a cada contexto.

Feitas estas considerações, resta aproximar a análise quantitativa aos objetivos desta obra, i.e., comparar pragmaticamente o conteúdo das proposições parlamentares ao dos *tweets*, e observar como a dinâmica intrínseca da formação de leis reflete nas interações do site de rede social.

### 5.3 CONSISTÊNCIA TEMÁTICA

Esta subseção investiga o quanto temas de proposições parlamentares de 2020 incidiram nos posicionamentos e mobilizações digitais dos seus respectivos autores. O termo “Consistência Temática” remete ao grau de similaridade entre estas duas áreas de atuação e manifestação. Esta investigação se baseou primariamente no sistema oficial de classificação das proposições, dividido entre temas e palavras-chave, e na coleta de publicações do Twitter.

Os procedimentos de análise começam na Célula 2.1, que carrega o arquivo “db-listaDeputados.csv”, obtido no *notebook* anterior. Na seguinte é executado um filtro que mantém apenas proposições com tipo de autor “Deputado”, deixando de fora órgãos públicos e entidades não discutidas nesta dissertação, e na categoria “Projeto

de Lei”, que contém o maior número de registros com temas e palavras-chave preenchidos, e que mais contempla os objetivos de pesquisa aqui propostos.

Na Célula 2.3 é feita uma função de agrupamento, tendo como parâmetro o campo “idDeputadoAutor”. Isto resultou em uma tabela com 453 linhas, onde cada uma contém o código de determinado deputado, e todas as informações sobre os projetos que publicou, separada por vírgula. Na mesma Célula, esta é unida com a base de dados “db-deputados.csv” a partir do campo “idDeputadoAutor”, a fim de integrar informações referentes aos perfis parlamentares.

Como foi visto anteriormente, 494 dos 513 deputados federais (96,30%) tem perfil no Twitter. Em dezembro de 2020, quando a coleta de dados foi feita, era possível resgatar apenas 3.000 publicações por perfil devido a restrições na API da plataforma. Isso resultou em um total de 1.021.532 *tweets*. No recorte temporal deste trabalho (01/01/2020 à 29/12/2020) o número fica em 349.362. Apenas publicações autorais foram consideradas nesta subseção. Isto é, todos os *retweets* foram removidos (Célula 2.4), para trabalhar somente com posicionamentos e opiniões dos deputados no ano de 2020. Isto reduziu o número de *tweets* a 285.726.

A função de agrupamento foi usada mais uma vez para associar texto e métricas das publicações para cada linha/perfil (Célula 2.5). Feito isto, a nova tabela é unida com a anterior através do campo “PERFIL\_TWITTER”. Desta vez, a função foi executada no modo *inner*. Ou seja, ficou apenas a interseção das duas tabelas, resultando assim em 370 linhas (Célula 2.6). Isso significa que apenas 72,12% da Câmara dos Deputados proporcionou condições necessárias para o tipo de análise proposto (ao menos uma proposição e *tweet* publicado no ano de 2020). Diante desta redução, exportada ao arquivo “usuários-proposicoes.csv”, foi produzido mais um perfil a fim de investigar se houve mudanças nas características predominantes.

**Tabela 8 – Comparação entre perfil geral e o de consistência temática**

	Toda base de dados (513 deputados)	Consistência temática (370 deputados)
Gênero	Homem (84,79%)	Homem (83,51%)
Raça	Branca (75,63%)	Branca (77,29%)
Idade	Entre 40 e 55 (40,54%)	Entre 40 e 55 (40,00%)
Estado Civil	Casado (69,98%)	Casado (69,72%)

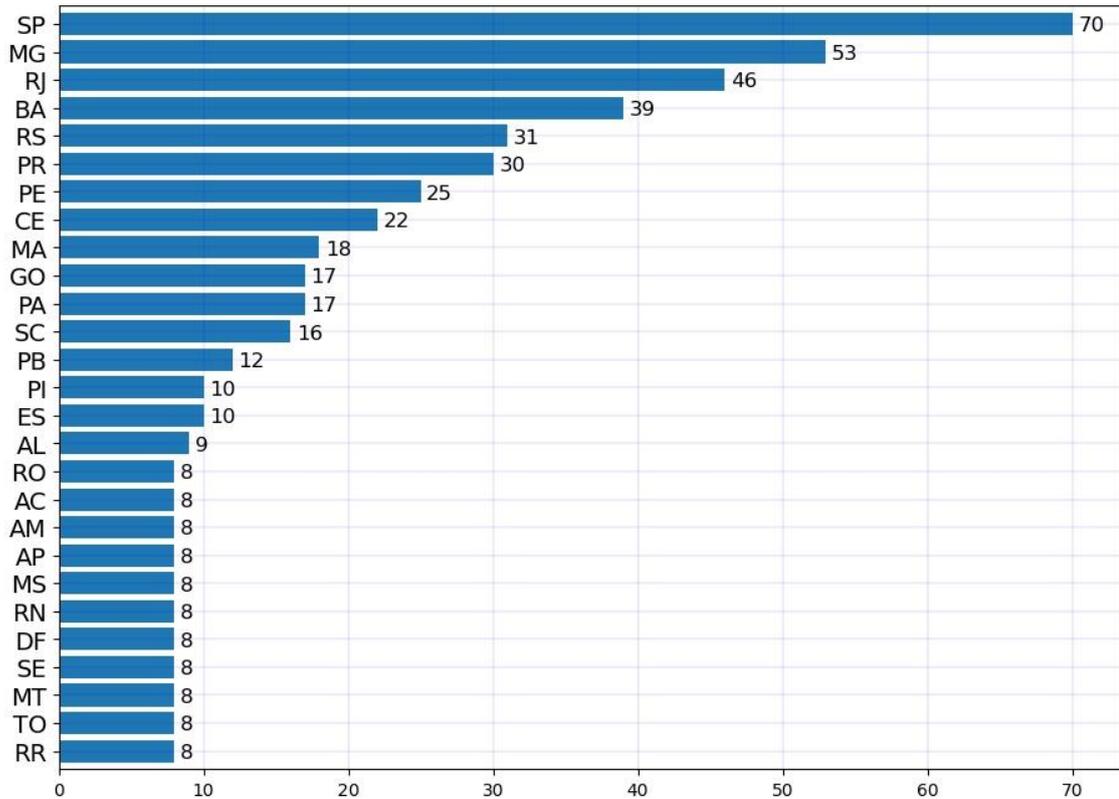
Ocupação anterior	Deputado (44,24%)	Deputado (42,70%)
Escolaridade	Superior (55,36%)	Superior (53,51%)
Partido	PSL e PT (10,33% cada)	PT (13,78%)
Ideologia	Direita (53,21%)	Direita (48,64%)
Recorrência	Sim (51,26%)	Não (53,24%)
Estado	SP (13,64%)	SP (14,59%)
Região	Sudeste (34,89%)	Sudeste (36,21%)
Agrupamento Bens		Até R\$407 mil (G4) (25,67%)
Agrupamento Votos		Até 87 mil (G3) (26,21%)
Agrupamento Seguidores		Até 15 mil e 350 (G2) (30,54%)
Agrupamento Número de tweets		Até 114 mil e 966 (G1) (30,81%)

**Fonte: Câmara dos Deputados, TSE e Twitter**

O novo perfil é ligeiramente diferente do anterior. No conjunto original, PT e PSL estão empatados em primeiro lugar, com 10,30%. Agora PT é o que prevalece, com 13,78%. No quesito da recorrência parlamentar também houve mudanças. 263 dos 513 deputados (51,26%) já cumpriram outros mandatos. A partir deste filtro, a relação se inverte. 197 dos 370 deputados (53,24%) estão em seu primeiro mandato.

Os agrupamentos numéricos da primeira coluna estão vazios, pois todos os valores são proporcionais (25,00%). Na segunda coluna, Agrupamento de Bens (G4) e de Votos (G3) se destacam por margens inferiores à 2,00%. As métricas digitais, representadas nos agrupamentos de seguidores e número de *tweets*, são mais significativas. Na coluna da esquerda Número de seguidores (G2) fica com 30,54% e Número de *Tweets* (G1) com 30,81%. O motivo da prevalência destes agrupamentos é o filtro utilizado anteriormente. A remoção de usuários que não publicaram *tweets* em 2020 deixou este conjunto com uma maior proporção de usuários ativos e influentes na plataforma. Antes de se aprofundar nestes dados, é preciso de mais proximidade com a distribuição geográfica da Câmara dos Deputados.

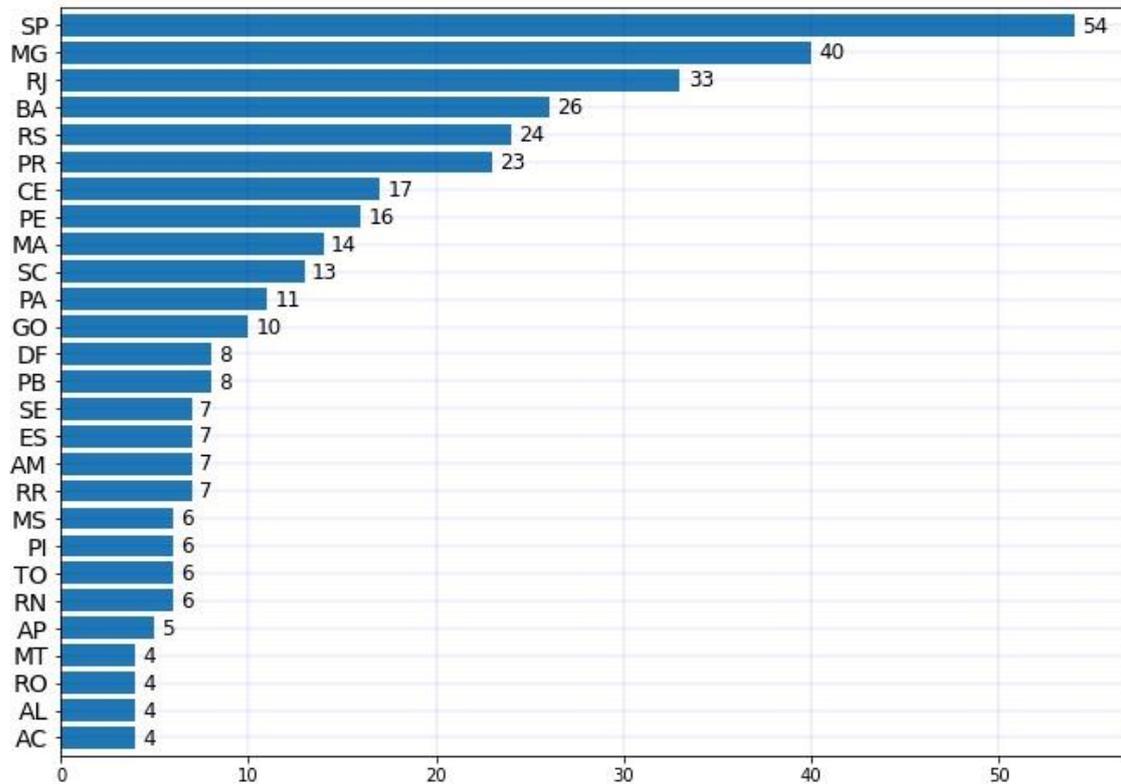
**Figura 8 – Quantidade de Deputados Federais por Estado**



Fonte: Câmara dos Deputados, TSE e Twitter

A revisão de literatura apresentada na seção “O uso do Twitter na Câmara dos Deputados” revelou que o número de deputados federais por Estado foi definido por um cálculo demográfico estipulado na Constituição de 1988. A Figura 8 foi gerada a partir da contagem de deputados na tabela “db-deputados.csv”, através do campo “ESTADO\_ELEICAO”, correspondendo a distribuição geográfica oficial. Em primeiro lugar, alcançando 70 cadeiras (número máximo por estado), está São Paulo. Em seguida, Minas Gerais e Rio de Janeiro, destacando a centralidade da região sudeste. Aqui o Rio Grande do Sul se encontra na quinta posição, com 31 deputados. As barras de Rondônia a Roraima dividem o último lugar com apenas 8 deputados, número mínimo de cadeiras conforme a Constituição.

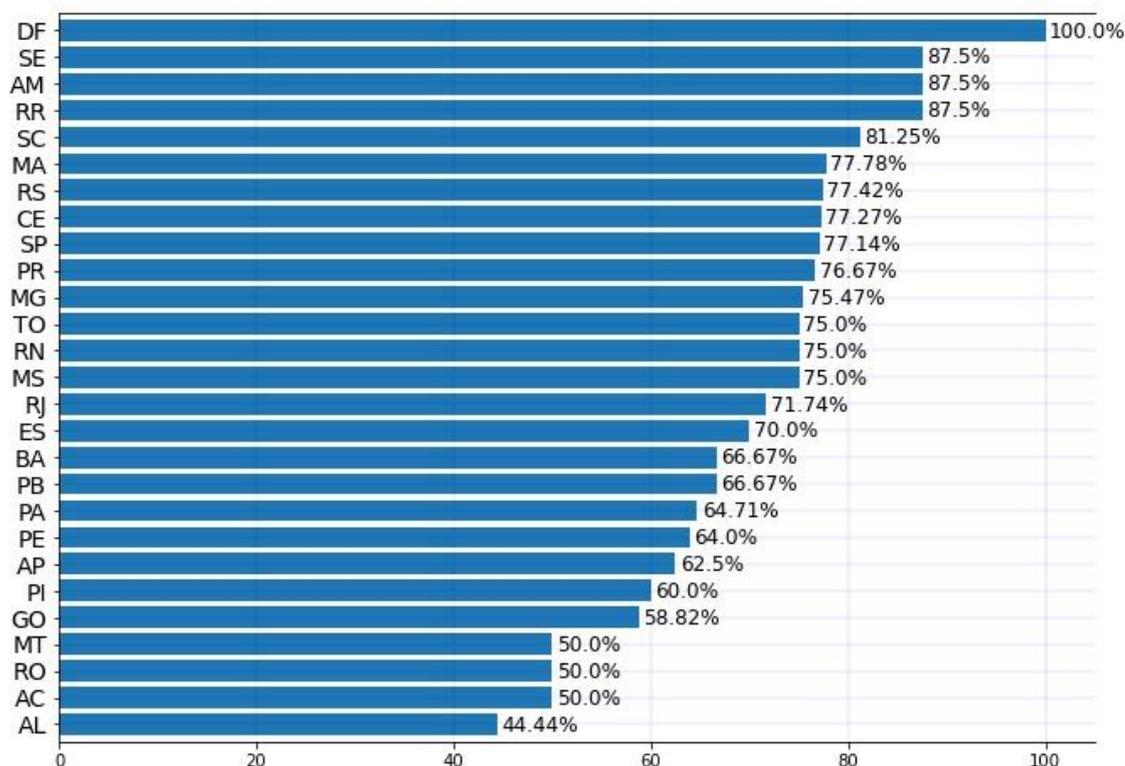
**Figura 9– Quantidade de Deputados Federais com proposições e tweets em 2020**



Fonte: Câmara dos Deputados, TSE e Twitter

A função de contagem por Estado é replicada na Figura 9. Desta vez, o método se restringe aos 370 deputados Federais que submeteram proposições e publicaram *tweets* em 2020, filtro exportada no arquivo “usuários-proposicoes.csv”. O fato de a ordem dos dados manter certa consistência coincide com uma observação feita anteriormente: características predominantes tendem a se repetir, mesmo diante de filtros e outras operações de agrupamento. Este é o caso para os seis primeiros Estados em número de deputados, que permaneceram em suas posições. A partir do sétimo lugar se evidenciam algumas diferenças. É entre os 11 Estados com o número mínimo de deputados, contudo, que a grande heterogeneidade da base de dados aparece. Nesta seleção, apenas dois (Distrito Federal e Paraíba) mantêm seu número original. Percebe-se, também, Mato Grosso, Rondônia, Alagoas e Acre nas últimas posições, apresentando apenas metade de suas bancadas. A fim de explorar melhor estas relações, foi gerada uma nova visualização de dados.

**Figura 10 – Proporção de Deputados Federais com proposições e tweets em 2020**



**Fonte: Câmara dos Deputados, TSE e Twitter**

A Figura 10 surgiu da necessidade de refinar o método apresentado anteriormente. Desta vez, a função de contagem foi seguida pela divisão dos seus resultados com o número original de deputados. Esta etapa permitiu comparar os Estados mais ativos no modelo de análise de modo proporcional, garantindo melhores indicadores de cultura política. Em primeiro lugar, atingindo 100%, está o Distrito Federal. Pode-se imaginar que o fato desta bancada se situar na capital do país incida em uma maior atividade parlamentar e sintonia com as mídias sociais. Em segundo e terceiro lugar estão Sergipe e Amazonas, respectivamente. Até a sétima posição (Rio grande do Sul) despontam somente estados das regiões Norte, Nordeste e Sul. A quebra da preponderância geográfica, acentuada na região Sudeste, dá indício as particularidades inerentes a cada Estado. É possível explorar os âmbitos legislativos e digitais de diversas maneiras.

A linha de pesquisa escolhida se baseia no índice de Consistência Temática – métrica criada para representar o grau de semelhança entre proposições parlamentares e *tweets*. A Tabela 9 exemplifica este procedimento, trazendo as colunas “KEYWORDS\_PLS”, “KEYWORDS\_TWEETS” e “INTERSEÇÃO” referentes ao Deputado Acácio Favacho, primeiro em ordem alfabética no conjunto.

**Tabela 9 – Exemplo de análise de consistência temática**

Deputado Acácio Favacho	
12 Proposições parlamentares 161 palavras	lei federal enfrentamento emergencia saude publica internacional espil autorizacao estado ente federado distrito municipio aplicacao recursos leilao petroleo regime cessao onerosa combate pandemia coronavirus social fundo fiscalizacao telecomunicacoes fistel financiamento medidas surto 2019 planos proibicao reajuste suspensao rescisao unilateral plano calamidade livre comercio vista alcbv roraima dispensa declaracao saida temporaria dst veiculo automotor isencao tributaria imposto produtos industrializados ipi desburocratacao tributacao concessao auxilio emergencial financeiro condutor auxiliar transporte escolar passageiros turista cobranca parcela consorcio aquisicao contrato segurosaude profissional deducao base caculo renda fisica irpf despesa equipamento protecao individual complementacao servicos qualquer natureza iss operacoes relativas circulacao mercadorias prestacoes interestadual intermunicipal comunicacao icms programa economica ppe emissao letra financeira tesouro guerra lftg especial nacional fetrn credito fce recuperacao fre nota ncre ministerio economia reducao atividade empresa movimentacao conta vinculada fgts garantia tempo servico disciplinamento procedimento interrupcao fornecimento energia eletrica falha sistema interligado sin direitos usuario prazo determinado estudante pagamento estudantil fies curso educacao avaliacao positiva saldo devedor aumento participacao uniao garantidor fgfies
24 Tweets 171 palavras	desafios luta assegurar melhorias amapa juntos povo amapaense viva democracia parabenizo especial macapaenses exerceram domingo desejo gestao conjunta bancada federal virtude apagao mincomunicacoes anatel determinaram lider pros camara deputados manifestamos apoio minluizramos trabalhado materia vivem lixao macapa solicitei prefeito clecio incluisse ultrapassamos triste marca 100 mil despedidas numero brasileiros morreram decorrencia trabalhamos incansavelmente busca acordo votacao novo fundeb sabador faco questao vir publico repudiar fake news anda circulando redes sociais respeito pl presidente jairbolsonaro estimamos plena rapida recuperacao volte logo combate reuni ministro comunicacoes fabio faria tratamos radio difusora internet esco estendemos auxilio emergencial 600 00 taxistas mototaxistas motoristas app manicures diari diaristas garco mandetta excelente trabalho frente min saude dr nelson teich ex 1 161 2020 garante arrecadacao fpm projeto beneficiara estados brasileiro protocolei ontem lei veda suspensao rescisao plano pandemia coronav entendo falencias perdas ganhos serao grandes bares restaurantes saloes beleza academias final sessao congresso varias pautas definidas veto 52 trata orcamento ldo aryoliver7 camilopsb briga ai amiga primeiro certo parabens estudantes linda cidade cidadao publica apaixonado venho fal
Interseção 13 palavras	plano publica recuperacao combate suspensao pandemia especial federal saude auxilio emergencial lei rescisao

**Fonte: Câmara dos Deputados, TSE e Twitter**

A coluna “KEYWORDS\_PLS”, é proveniente do campo “Keywords”, já mencionado previamente, e “KEYWORDS\_TWEETS”, relativo ao texto dos *tweets* dos deputados. Cabe destacar aqui que “KEYWORDS\_TWEETS” é uma fonte de dados

mais irregular que “KEYWORDS\_PLS”, uma vez que no momento da coleta a API do Twitter cortava os textos resgatados em 140 caracteres.

Dois métodos adicionais foram aplicados nestas colunas. A remoção de *stopwords*, isto é, palavras que não acrescentam nada a análise, como artigos e preposições (Célula). E o armazenamento das palavras únicas de cada linha em listas, estrutura de dados mais simples da linguagem Python, ambos na Célula 2.10. A coluna interseção, gerada a partir da Célula 2.11, é a junção de palavras presentes tanto no vetor de “KEYWORDS\_PLS” quanto no de “KEYWORDS\_TWEETS”.

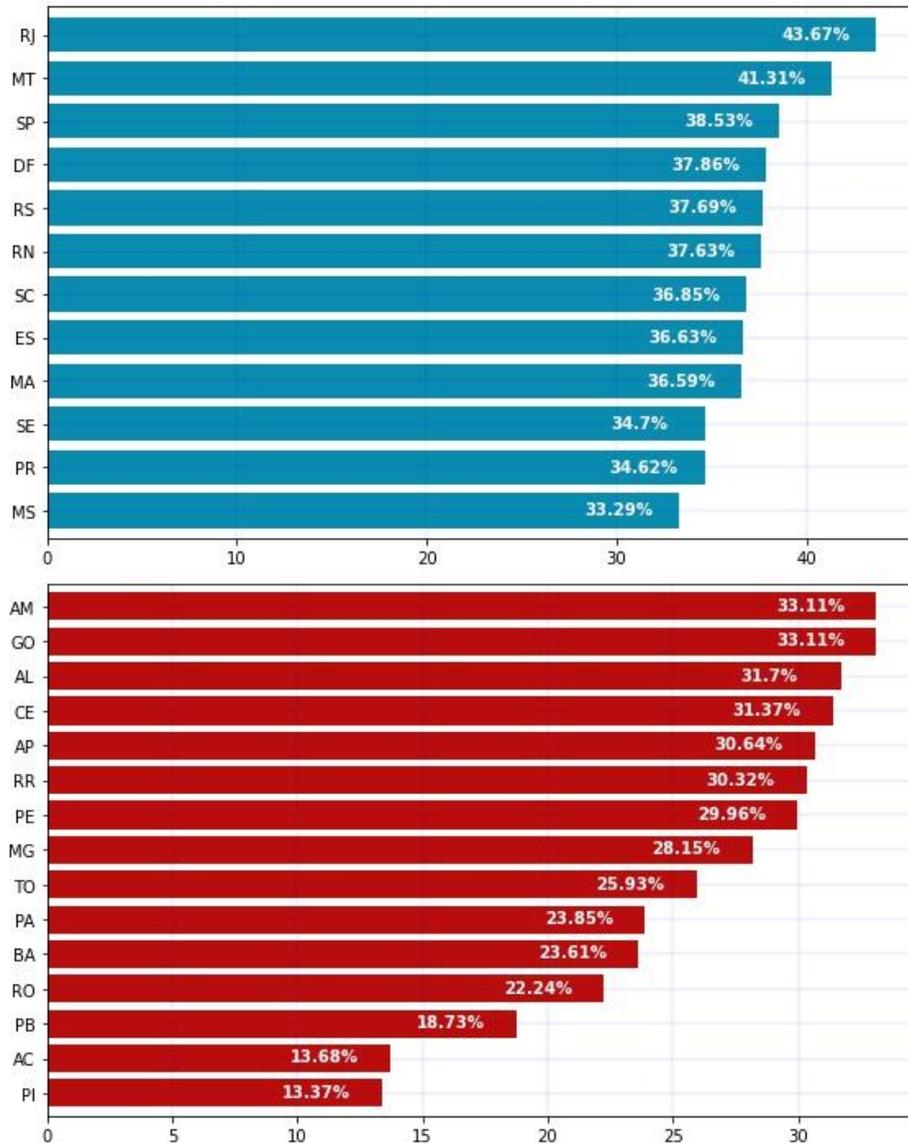
Na Tabela 6, “KEYWORDS\_PLS” aparece com 161 palavras, extraídas de 12 proposições. Em seguida, “KEYWORDS\_TWEETS”, com 171 palavras, extraídas de 26 *tweets*. Por fim, “INTERSEÇÃO”, com as 13 palavras presentes nos dois campos anteriores. A métrica de consistência temática é a porcentagem de “INTERSEÇÃO” com “KEYWORDS\_PLS” (Célula 2.12). Neste caso:  $(13/161) * 100 = 8.07\%$ . Uma nova coluna foi feita reproduzindo este cálculo para todos os deputados.

A média de todos os valores permitiu responder à pergunta de pesquisa “quão consistentes são os discursos dos deputados federais no Twitter, se comparados às suas proposições parlamentares?” O valor encontrado, em percentual é 31,97%. Ainda que a documentação dos métodos e abordagens, como foi feito anteriormente, seja suficiente para justificar esta resposta, esta, por si só, não é. Este trabalho se baseia primariamente em análise descritiva. Ou seja, não basta somente apresentar a resposta, é precisar mostrar o que ela significa sob os diferentes contextos do objeto de estudo escolhido.

Todos os procedimentos descritos até aqui foram armazenados em um *Dataframe*, estrutura de dados da biblioteca *pandas*, que em seguida foi exportada ao arquivo “usuariosProposicoes.csv” (Célula 2.17). Os gráficos e figuras a seguir foram produzidas a partir deste arquivo, por meio de funções das bibliotecas *matplotlib*, *seaborn* e *plotly*.

Para gerar a Figura 10, a coluna de consistência temática foi alvo de uma função de agrupamento por Estado (Célula 2.19). Deste modo, a média dos valores foi calculada para 27 perfis. Os 12 maiores perfis estão em azul, e os 15 menores em vermelho.

**Figura 11 - Índices de consistência temática por estado**

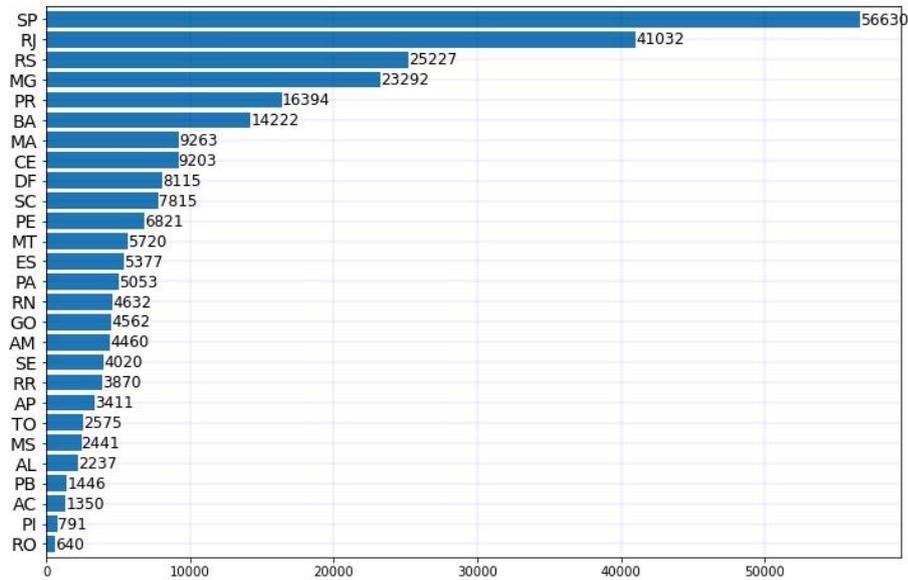


**Fonte: Câmara dos Deputados, TSE e Twitter**

O Estado com a maior métrica é o Rio de Janeiro, com 43,67%. Em segundo e terceiro lugar se encontram Mato Grosso e São Paulo. Distrito Federal se mantém entre o rol de bancadas ativas no Twitter, atingindo 37,86%. Assim como o Rio Grande do Sul, em quinto lugar, com 37,69%. Minas Gerais, segundo Estado em número de deputados (53), ficou em vigésima posição, com 28,15%.

Percebe-se que a quantidade de deputados não influi nos índices. O mesmo se aplica em relação aos *tweets*. Seguir o cálculo apresentado anteriormente pode levar a seguinte impressão: quanto maior o conjunto de publicações de determinado parlamentar, maiores as chances de uma interseção e, conseqüentemente, uma métrica mais robusta.

**Figura 12 - Quantidade de Tweets por Estado**



**Fonte: Câmara dos Deputados, TSE e Twitter**

É possível refutar esta argumentação a partir da quantidade de *tweets* por Estado, representada na Figura 11. O ordenamento dos dados aqui não difere significativamente ao da quantidade de deputados federais com proposições e *tweets* em 2020. Isto indica que a Consistência Temática diz mais sobre “como” do que o “quanto” a mídia social é utilizada. Ainda assim, o método não é perfeito, sendo preciso frisar seu alcance e limitações.

**Tabela 10 – Limitações do método de Consistência Temática**

Deputado Raimundo Costa	
1 Proposição Parlamentar	lei organica saude programa nacional leitos uti prout autorizacao uniao aquisicao leito unidade tratamento intensivo suplementar calamidade publica emergencia internacional espil espil pandemia coronavirus hospital privado adesao isencao tributaria beneficio fiscal tributacao
3 Tweets	domingo faltar moqueca alias nenhum obrigado amigo joao pedro certeza valenca orgulho imenso poder dizer lamento profundamente falecimento humorista baiano jose luiz almeida silva jotinha vitima covid 19
Interseção	

**Fonte: Câmara dos Deputados, TSE e Twitter**

Como se pode ver na Tabela 10, a proposição publicada pelo Deputado diz respeito a pandemia, vide as palavras-chave “saúde”, “uti”, “leito”, “tratamento”, “coronavírus” etc. Nos *tweets* não parece ser diferente, dada a presença de palavras

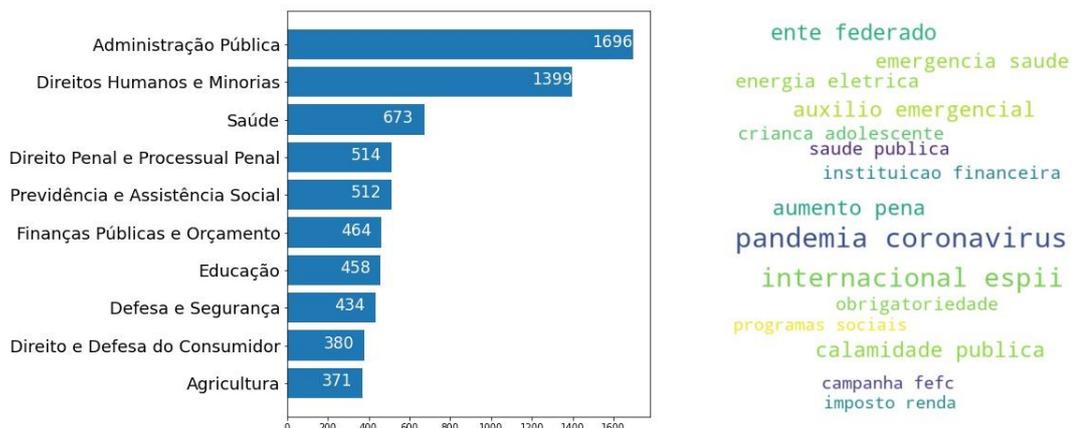
como “vítima” e “covid 19”. Ainda assim, não há correspondência direta então a intercessão fica com 0, e 0 dividido por 1 é 0.

Este caso evidencia um revés do método utilizado, já que há uma ligação “visível” entre o conteúdo dos *tweets* e do trabalho parlamentar. Uma possível resolução para este problema estaria em utilizar métodos de análise semântica mais sofisticados, à ribeira do Processamento de Linguagem Natural (PLN). No entanto, o aprofundamento destas técnicas foge do escopo desta dissertação.

### 5.3.1 ÂMBITO GERAL

A Figura 12 exhibe o conteúdo dos trabalhos parlamentares de 2020 segundo o sistema de classificação da Câmara, evidenciando a diferença entre os campos “Temas” e “Keywords” (Célula 2.20).

**Figura 13 – Temas e palavras-chave mais recorrentes nas proposições – Geral**



**Fonte: Câmara dos Deputados, TSE e Twitter**

No gráfico da esquerda, referente aos temas, “Administração Pública” está em primeiro lugar e “Direitos Humanos e Minorias” em segundo. “Saúde” só aparece em terceiro lugar. Contudo, ao observar a *Tagcloud* à direita percebe-se uma grande variedade de palavras-chaves relativas à pandemia, isso acontece porque estas não são relativas somente a saúde, mas transversais a outros temas. “Internacional” e “Espii” são outros termos com maior centralidade. O motivo disso é uma declaração oficial da Organização Mundial da Saúde (OMS) chamada “Emergência de Saúde Pública de Âmbito Internacional” acerca das atribuições e deveres legais dos países na pandemia.

**Figura 14 – Publicações mais compartilhadas no Twitter – Geral**

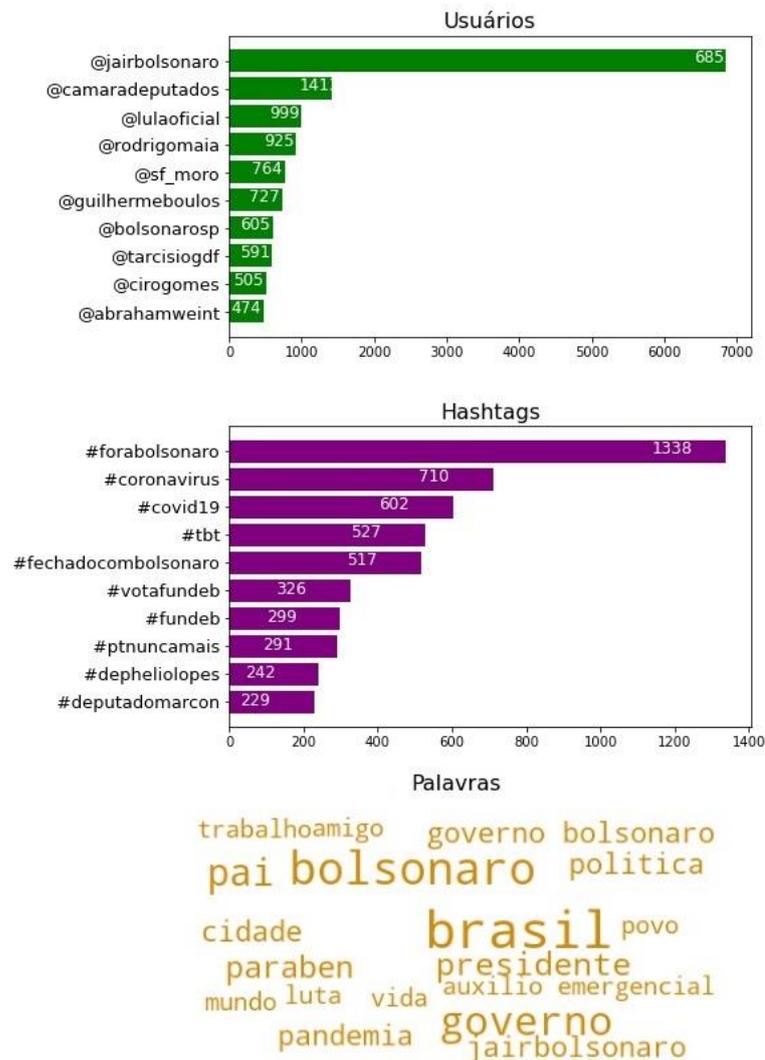


**Fonte: Câmara dos Deputados, TSE e Twitter**

Na figura 13 estão os *tweets* mais compartilhados dos deputados. Esta visualização de dados provém de um *script* que pega os identificadores únicos de cada *tweet*, gerando, a partir deles, códigos de HTML que renderizam as publicações no *layout* oficial da plataforma (Célula 2.21). O primeiro lugar, em número de *retweets*, é de uma publicação do deputado Túlio Gadêlha comentando a abdicação de Sérgio Moro ao cargo de Ministro da Justiça, em abril de 2020. Jandira Feghali (PCdoB) publicou o segundo maior *tweet* em alcance: um vídeo de uma brasileira na Itália alertando o perigo do coronavírus. O então presidente da Casa, deputado Rodrigo Maia (DEM), repudia manifestações contra o Congresso Nacional, em abril de 2020, na terceira publicação com o maior número de *retweets*. Cabe ressaltar que, apesar destas três mensagens pautarem temas de interesse público e pertinentes ao

processo democrático, nenhuma faz menção direta ou indireta a qualquer proposição parlamentar.

**Figura 15 – Usuários, hashtags e palavras mais frequentes no Twitter – Geral**



Fonte: Câmara dos Deputados, TSE e Twitter

É possível compreender melhor o uso do *Twitter* pelos deputados a partir da Figura 14, gerada a partir da Célula 2.22. O primeiro gráfico traz a quantidade de menções à usuários. Jair Bolsonaro está no topo, em seguida o perfil oficial da Câmara dos Deputados e o do ex-presidente Lula. O nome dos políticos foi utilizado em discussões polarizadas, que eventualmente especulavam sobre a eleição presidencial de 2022. Vale destacar que “menção” é um recurso de interação sem direcionamento específico, podendo ser tanto para apoio quanto para crítica.

Rodrigo Maia alcançou a quarta posição. Ele e Eduardo Bolsonaro, em sétimo lugar, foram os únicos deputados a entrarem neste ranking. Depois de Maia está Sérgio Moro, devido ao grande número de publicações comentando sua saída do governo. É notável também a presença de presidentes de 2018, além de Bolsonaro e Lula, estão Guilherme Boulos, em sexto, e Ciro Gomes em nono lugar. Os outros dois usuários desta lista são Tarcísio Gomes de Freitas, Ministro da Infraestrutura, e o ex-ministro da Educação, Abraham Weintraub.

No segundo gráfico se encontram as *hashtags* mais utilizadas pelos deputados federais em 2020. “ForaBolsonaro” é a mais proeminente. A segunda e terceira *hashtag* cobrem o tema da pandemia de coronavírus. A *hashtag* “tbt”, utilizada para marcar recordações, ficou em quarto lugar. Cabe lembrar que 2020 foi um ano de eleições municipais, onde muitos deputados fizeram campanhas de apoio para colegas que estavam concorrendo. O que consiste muitas vezes em recapitular ações e marcos de trajetórias políticas. “FechadoComBolsonaro” é a quinta *hashtag* mais utilizada, tendo menos da metade do alcance da de oposição. Na mesma linha está “PTnunca mais”, alcançando o oitavo lugar.

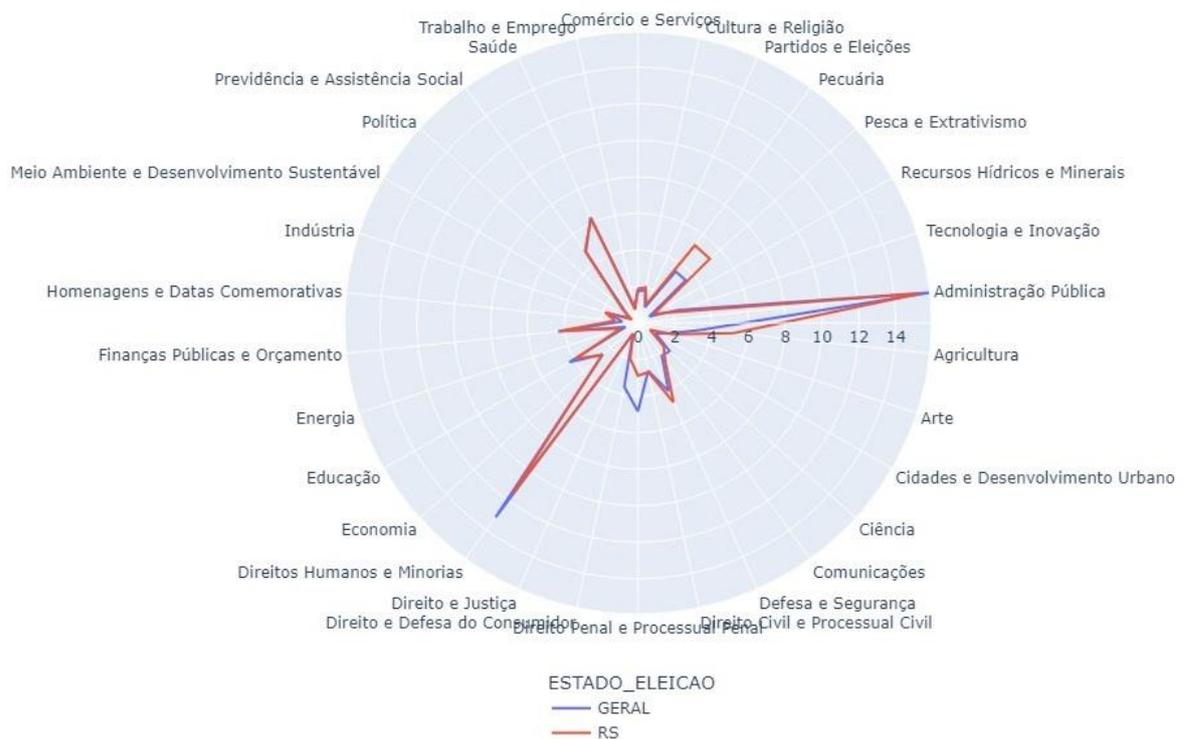
As *hashtags* mais relacionadas ao trabalho parlamentar são “votafundeb” e “fundeb” (sexto e sétimo lugar), que dizem respeito a votações sobre o Fundo de Manutenção e Desenvolvimento da Educação Básica. Já #dephiliopes e #deputadomarcon (oitavo e nono lugar) são *hashtags* autopromocionais dos deputados Dionilso Marcon e Helio Lopes.

Há poucas similaridades entre a nuvem de palavras do Twitter e da Câmara. O índice de consistência temática geral é 31,97%. A partir da última imagem da Figura 10, é possível entender o porquê. “Auxílio Emergencial” e “Pandemia” estão dentro desta interseção. As outras palavras pertencem a contextos diferentes. “Cidade”, “amigo” e “povo” muitas vezes remetem às eleições municipais. “Bolsonaro”, “presidente” e “jairbolsonaro” e “governo”, às incessantes manifestações de crítica e apoio ao presidente. “Mundo”, “luta”, “vida” e “pai” são utilizadas em *tweets* sobre a covid-19, muitas vezes lamentando novas vítimas da doença. Estas publicações mais pessoais não entram para a interseção, apesar de compartilharem dos mesmos temas, por motivos já explicados anteriormente. Isso conclui a análise de consistência temática em um âmbito geral.

### 5.3.2 ÂMBITO ESTADUAL

A seguir se encontram visualizações de dados voltadas ao conjunto de deputados do Rio Grande do Sul. O gráfico de radar na Figura 15 mostra o conjunto de temas mais frequente em suas proposições.

**Figura 16 - Temas mais recorrentes nas proposições - Rio Grande do Sul**



**Fonte: Câmara dos Deputados, TSE e Twitter**

A linha azul denota o quadro geral. A linha vermelha, o Rio Grande do Sul. Gerar esta imagem consistiu em executar funções de contagem no arquivo “db-proposicoes.csv” e, em seguida, normalizar os valores obtidos (Célula 2.27). Ou seja, tem-se aqui o que cada valor representa em relação ao todo, e não de modo absoluto. Isso foi feito devido a diferença de magnitude entre o âmbito geral e específico, e a alta frequência de proposições relativas à “Administração Pública”, que em muitos casos ofuscava as outras categorias.

A visualização de dados, em si, foi produzida a partir da biblioteca *seaborn*, apresentando os 15 temas mais recorrentes. Isto causou a sobreposição de alguns rótulos. O número de eixos do gráfico não foi reduzido, tendo em vista uma maior

diversidade de temas, por isso os que tiverem mais centralidade serão descritos no decorrer do texto.

Como visto anteriormente, prevalecem proposições sobre “Administração Pública”, “Direitos Humanos e Minorias” e “Saúde”. Entre os deputados gaúchos, “Direitos Humanos e Minorias” ficou com 11,62%, percentual ligeiramente menor que os 13,14% do âmbito geral. A partir da quarta posição, pronunciam-se as diferenças da bancada. O enfoque passa a ser “Pecuária”, “Pesca e ativismo” e “Agricultura”, ao invés de “Direito Penal e Processual Penal”, “Previdência e Assistência Social” e “Finanças Públicas e Orçamento”, como no quadro original. “Defesa e Segurança” fica na oitava posição, em ambos os campos, com menos de um ponto percentual a mais para os deputados do Rio Grande do Sul.

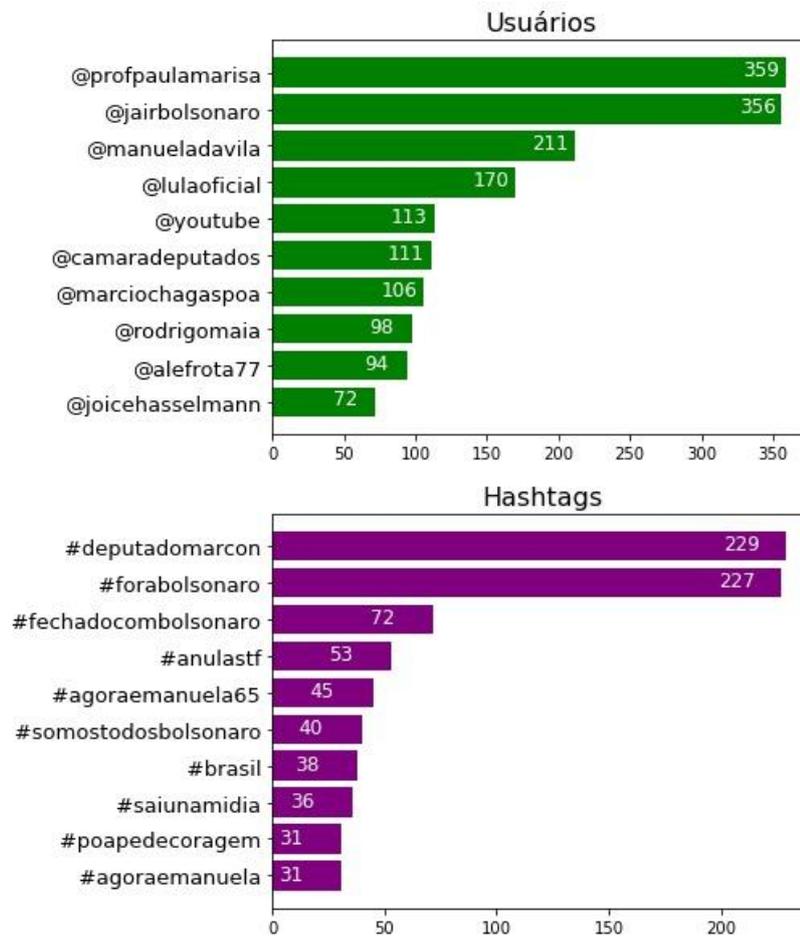
**Figura 17 – Publicações mais compartilhadas no Twitter – Rio Grande do Sul**



**Fonte: Câmara dos Deputados, TSE e Twitter**

Na Figura 16 se encontram os *tweets* de maior alcance da bancada gaúcha. O primeiro é uma crítica da deputada Maria do Rosário (PT) a uma nova contratação do goleiro Bruno, condenado pelo assassinato da ex-mulher em 2010. O deputado Marcel van Hattem (NOVO) é autor da segunda e terceira publicação mais compartilhada. Ambas trazem uma denúncia do deputado a inadmissão do então presidente da casa, deputado Rodrigo Maia (PP), a uma emenda do NOVO que visava destinar recursos do fundo eleitoral e partidário à Saúde.

**Figura 18 – Usuários e hashtags mais frequentes no Twitter - Rio Grande do Sul**



Fonte: Câmara dos Deputados, TSE e Twitter

A Figura 17 traz as menções e *hashtags* mais recorrentes no subconjunto de deputados do Rio Grande do Sul. Paula Marisa, professora e influenciadora digital de direita, foi a usuária mais mencionada pela bancada. Na segunda e quarta posição estão os perfis oficiais de Jair Bolsonaro e Lula, indicando tanto manifestações de apoio quanto de oposição a estas figuras políticas. Manuela D’Ávila, candidata à prefeita de Porto Alegre nas eleições municipais de 2020, aparece em terceiro lugar. Neste mesmo contexto, na sétima posição, se encontra o perfil de Marcio Chagas, vice na chapa da deputada Fernanda Melchionna (PSOL). O perfil do *site* Youtube, em 5º lugar, é decorrente das menções involuntárias que ocorrem ao compartilhar vídeos da plataforma. Menções ao perfil oficial da Câmara dos Deputados partem da divulgação de projetos e votações nominais. O deputado Rodrigo Maia (PP) é mencionado em discussões sobre a nova eleição à presidência da Câmara, em março de 2021. O nono e décimo lugar pertencem aos deputados ex-bolsonaristas Alexandre

Frota (PSDB) e Joice Hasselmann (PSL), respectivamente. Ambos são criticados, tanto pela base do governo quanto pela oposição.

Na parte inferior da imagem observa-se, em primeiro lugar, a *hashtag* autopromocional do deputado Dionilso Marcon (PT). Na segunda posição, #forabolsonaro, com 227 menções. Apesar desta ser a única *hashtag* contrária ao presidente, leva vantagem numérica sobre a soma de #fechadocombolsonaro (3ª) e #somostodosbolsonaro (6ª). Em quarto lugar, #anulastf traz pedidos para o Supremo Tribunal Federal (STF) anular as condenações de Lula na Lava Jato, tornando-o elegível nas eleições presidenciais de 2022. Deputados de esquerda manifestaram apoio a candidatura de Manuela D'Ávila através das *hashtags* #agoraemanuela65 (5ª), #poadecoragem (9ª) e #agoraemanuela (10ª). #brasil, em sétimo lugar, foi usado na divulgação de notícias sobre as eleições municipais de 2020. Bibó Nunes (PSL) foi o maior propulsor da *hashtag* #saiunamidia (8ª). Nas duas maiores publicações neste contexto, o deputado critica a decisão preliminar do STF e um projeto que possibilitaria a reeleição dos presidentes da Câmara e Senado.

**Figura 19 - Nuvens de palavras sobre proposições e tweets - Recorrência Parlamentar**



**Fonte: Câmara dos Deputados, TSE e Twitter**

As Nuvens de Palavras da Figura 18 foram feitas com a biblioteca *Wordcloud*, para contar e apresentar as palavras mais utilizadas nos conjuntos de textos, e *Matplotlib*, para organizar as imagens em formato de grade (Célula 2.32). Cabe destacar que o conteúdo da fileira “GERAL” é ligeiramente diferente do que foi

apresentado no começo da seção. Aqui foi necessário reduzir o número de palavras para os quadros seguirem o *layout* da imagem.

As palavras-chave das proposições do Rio Grande do Sul remetem aos temas discutidos na Figura 15. “Crime” “reintegração” “aumento”, “pena” estão dentro de “Defesa e Segurança”. Há uma correlação aqui com o âmbito geral, através das palavras “penal” e “mandado”. Existe outra conexão, ainda que menor. As palavras “saúde” e “suspensão” aparecem nas duas fileiras, mas há maior centralidade nos termos relativos a pandemia no quadro geral.

A coluna relativa ao Twitter indica autorreferência dos deputados do Rio Grande do Sul, visto a alta frequência das palavras “bancada” e “gaúcha”. É superlativo assumir um senso de união ou territorialíssimo a partir deste dado. A maioria dos *tweets* relacionados usou este termo para divulgar a destinação de verbas ou emendas aprovadas pela maioria deste conjunto. No contexto mais proeminente dos termos “parabéns”, “boas” e “notícias” se encontram congratulações a sucessos de campanha nas eleições municipais de 2020. O nome “Paulo” aparece em destaque na imagem por dois motivos: divulgação de atividades legislativas do deputado federal Paulo Pimenta e do senador Paulo Paim (PT), e celebração de novos leitos ao Hospital São Vicente de Paulo, na cidade de Osório. Maior proximidade com as bases e mensagens informais no período das festas de fim de ano são o contexto por trás das palavras “amigos” e “natal”.

Há pouca semelhança deste conjunto com o do âmbito geral. Evidencia-se a centralidade dos termos “pls” “assegurar” e “melhorias”, no mesmo contexto. E, de modo relacionado, “autora” e “coautora”, trazendo a divulgação de proposições de deputadas na Câmara. “Saúde”, “luta” e “desafios” são palavras utilizadas em *tweets* sobre a “Covid-19”. Também toma destaque, no âmbito geral, destinação de verbas e proposições relativas ao Estado do “Amapá”.

Na coluna “Interseção” se encontram representações visuais do índice de Consistência Temática. Os 31,07% da bancada gaúcha revelam que, apesar de não serem os termos mais utilizados, “sistema”, “sus” e “pandemia” estão presentes tanto nas proposições quanto nos *tweets*. Também partilham do contexto da Covid-19 as palavras “social”, “empesas” e “renda”, geralmente associadas a projetos e discussões sobre o Auxílio Emergencial. Menções a “alimentos” permanecem neste íterim, se associando ainda ao tema “Agricultura”. Um exemplo disso é a Indicação 282/2020, solicitando a então Ministra da Agricultura, Pecuária e Abastecimento, deputada

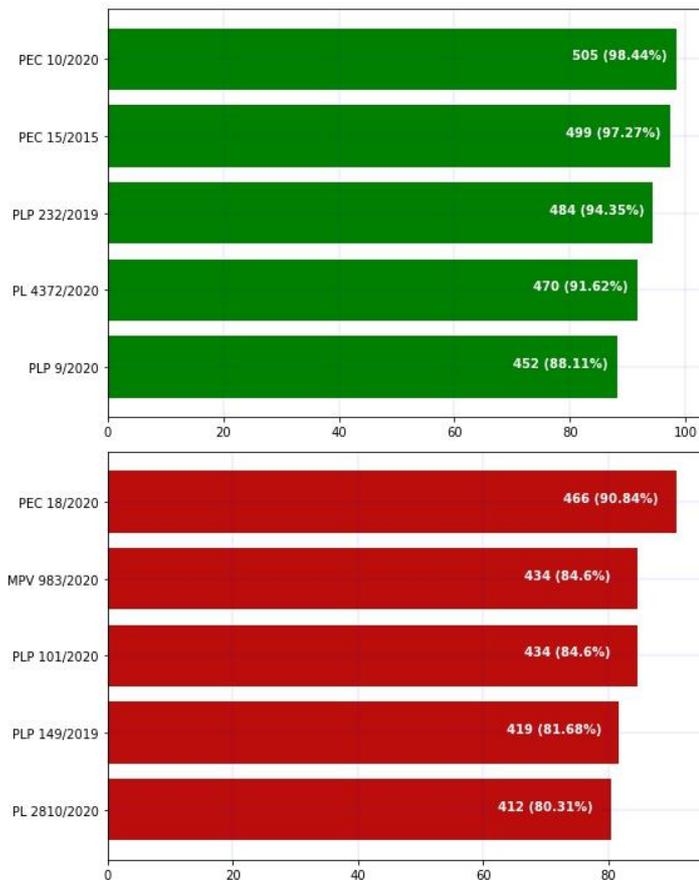
Tereza Cristina (DEM), formação de cestas básicas para afetados pela calamidade pública decorrente da pandemia de Covid-19. E o Projeto de Lei 1083/2020, que voga pela adaptação do Programa de Aquisição de Alimentos (PAA), voltado a agricultura familiar, para este fim. Ainda que o termo “gênero” esteja presente em proposições sobre “gênero alimentício”, o que mais se destacou no Twitter foram debates polarizado sobre Projetos de lei, como o 5587/2020, que institui aumento de pena nos crimes de violência política a mulheres, quando praticados no rádio, televisão ou internet. Nestes casos, deputados de esquerda usam a expressão “diversidade de gênero”, e de direita, “ideologia de gênero”.

Na última Nuvem de Palavras, referente à intercessão do âmbito geral (31,97%), “pandemia”, “saúde”, “suspensão” e “empresas” aparecem mais uma vez no contexto de enfrentamento a calamidade pública ocasionado pelo avanço da Covid-19. Finalizada a exploração dos dados através do conteúdo dos trabalhos legislativos e dos *tweets*, sintetizados pelo método de Consistência Temática, resta voltar-se a dinâmicas específicas às redes dos deputados.

#### **5.4 FORMAÇÃO DE REDES**

O modo como os deputados interagem no âmbito parlamentar e digital será avaliado aqui por meio de técnicas avançadas de filtragem, provenientes do campo das Expressões Regulares, e de visualizações de dados em rede, conhecidas como Grafos. Primeiramente foi preciso levantar quais as proposições que marcaram o recorte temporal estudado. O Parâmetro escolhido para isso partiu de mais uma revisão de literatura sobre o tema.

Se uma proposição atende todos os princípios e regras constitucionais, é levada para votação. (ANDRADE; COUTINHO, 2019). Segundo Vieira (2019), o resultado das votações nominais no Plenário da Câmara dos Deputados é o indicador mais importante do comportamento dos deputados em relação às orientações partidárias e a suas posições em relação ao governo. Seguir este indicador levou a ordenar as proposições de acordo com os maiores percentuais de aprovação e rejeição do ano de 2020.

**Figura 20 - Proposições com maiores percentuais de aprovação e rejeição**

Fonte: Câmara dos Deputados, TSE e Twitter

Na parte superior da Figura 19, gerada a partir do banco de dados “db-votacoes.csv” na Célula 3.3, estão as 5 proposições com maior número de votos “Sim”. Duas estão fora do recorte temporal estudado. Ou seja, foram votadas em 2020, mas submetidas em outros períodos. Isso pode ser considerado algo comum, já que 395 das 1257 proposições votadas (31,42%) são de anos anteriores. Uma possível explicação para isso está no fato do Regimento Interno da Câmara dos Deputados (RICD) possibilitar aos parlamentares formular recursos e estratégias para retardar a deliberação de proposições (VIEIRA, 2019).

Em primeiro lugar no percentual de aprovação, com 505 votos (98,44%), está a Proposta de Emenda à Constituição 10/2020, que institui um regime extraordinário nas contas públicas para garantir o enfrentamento a pandemia do coronavírus. A PEC 15/2015, em segundo lugar, com 499 votos (97,27%), torna o Fundo de Manutenção e Desenvolvimento da Educação Básica (FUNDEB) instrumento permanente da Constituição Federal. Na ementa do Projeto de Lei Complementar (PLP) 232/2019, terceiro em número de votos “Sim”, consta que este dispõe sobre a transferência de

saldos financeiros, provenientes de repasses federais, a fundos de Saúde dos Estados e Municípios. O Projeto de Lei (PL) 4372/2020, em quarto lugar, regulamenta o FUNDEB. A última proposição da lista, com 88,11% de aprovação, é o PLP 9/2020, que prorroga o prazo para micro e pequenas empresas em início de atividade se enquadrarem no regime do Simples Nacional, e faz com que a arrecadação de tributos e contribuições desta categoria seja tratada por lei federal.

As proposições com maior percentual de votos “Não” são apresentadas na parte inferior da Figura 19, em vermelho. Em primeiro lugar, com 90,84% de rejeição, está a PEC 18/2020, que propunha adiamento das eleições municipais em decorrência da pandemia de coronavírus. Este alto índice de reprovação não é nenhuma surpresa, se considerado que a estrutura política das campanhas está intrinsicamente associada às carreiras e redes de relações dos deputados, como visto na análise dos *tweets*. O segundo lugar traz um tipo de trabalho legislativo não visto até então. Se trata de uma Medida Provisória (MPV) – matéria submetida em caso de relevância e urgência pelo gabinete do presidente da República, tendo de ser votada no Congresso Nacional até sessenta dias após sua publicação (ANDRADE; COUTINHO, 2019).

Abranches (2018) discorre extensamente sobre estas medidas em seu livro “Presidencialismo de Coalizão”. Segundo o autor, este dispositivo deu a presidência, ainda no começo do novo regime constitucional, a capacidade de controlar a agenda legislativa. É relatado que diversos presidentes abusaram destas medidas, usando-as para coibir atividades parlamentares adversas a seus projetos de governo. O Legislativo tem o poder de regular a cessação de medidas provisórias, tendo que lidar, posteriormente, com as consequências políticas disso (ABRANCHES, 2018).

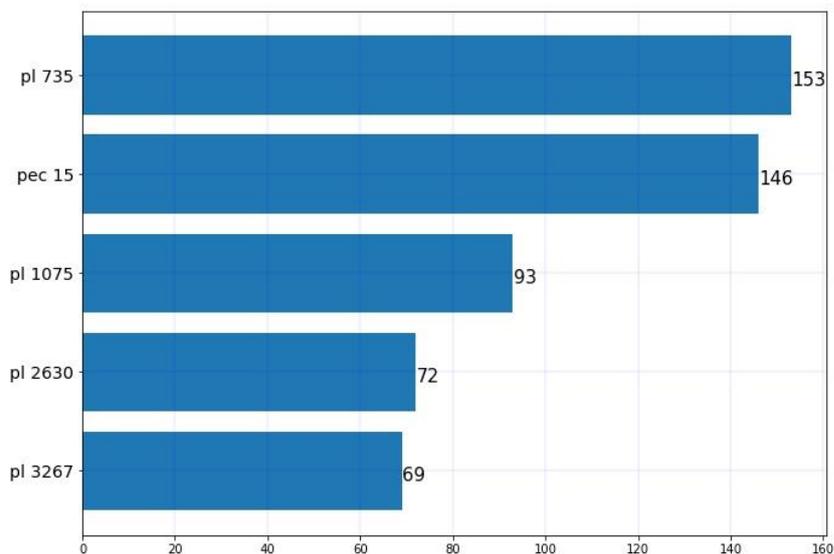
Foi o que aconteceu neste caso. A MPV 983/2020, rejeitada por 84,60% da Câmara, propunha conferir a assinaturas eletrônicas o mesmo valor legal das assinaturas tradicionais em papel. O site do Senado Federal (2020) cobriu o trâmite desta medida. Na matéria, deputados expressam preocupação com exclusão social, principalmente no caso de idosos e pessoas de baixa renda com difícil acesso às tecnologias digitais. 76 sugestões de emendas foram atreladas à medida provisória, que será revisada para uma nova submissão.

Os Projetos de Lei Complementar 101/2020 e PLP 149/2019, terceiro e quarto lugar, respectivamente, se voltavam ao estabelecimento de programas de acompanhamento, transparência e equilíbrio fiscal. Na quinta posição, com 80,31%,

está o PL 2810/2020, submetido pelo deputado Arthur Lira (PP), atual e então candidato à presidência da Câmara. O projeto visava alterações no Código Penal, propondo nova redação ao crime de denúncia caluniosa.

A busca pela menção de proposições no Twitter começou pela aplicação de um filtro na base de dados “tweets2020.csv” (Célula 3.9). Ao manter apenas publicações que contivessem os termos relativos aos tipos de proposições estudadas, o número de registro no âmbito geral passou de 279.131 para 5.891 (2,11%). Já o de usuários, de 370 para 339. Foi necessária outra etapa para obter as respectivas proposições da base filtrada. O método utilizado para este fim partiu de expressões regulares (Célula 3.9). Toda proposição tem um número seguido por uma barra e o ano em que foi submetida. Deste modo, o padrão encontrado para extrair os termos correspondentes tomou a seguinte forma “pl \s\d\*|ric \s\d\*| [...]”. Onde, após a o tipo de projeto (em sigla, ou escrito de forma literal), “\s” capta espaços em branco, “\d\*” todos os dígitos subsequentes, e “[...]” repete a operação para o próximo termo listado. Isso gerou uma lista com 4.065 elementos. Esta redução decorre de menções genéricas, onde não é especificado número ou ano da proposição.

**Figura 21 - Menção a proposições no Twitter- Geral**



**Fonte: Câmara dos Deputados, TSE e Twitter**

As cinco proposições mais mencionadas no âmbito geral são apresentadas na Figura 20, proveniente da Célula 4.6. O PL 735/2020, em primeiro lugar com 153 menções, abona atribuições legais de feirantes e agricultores familiares em quarentena devido a pandemia de COVID-19. Grande parte da reação no Twitter veio de deputados do MDB e PT, celebrando a aprovação desta matéria legislativa, que

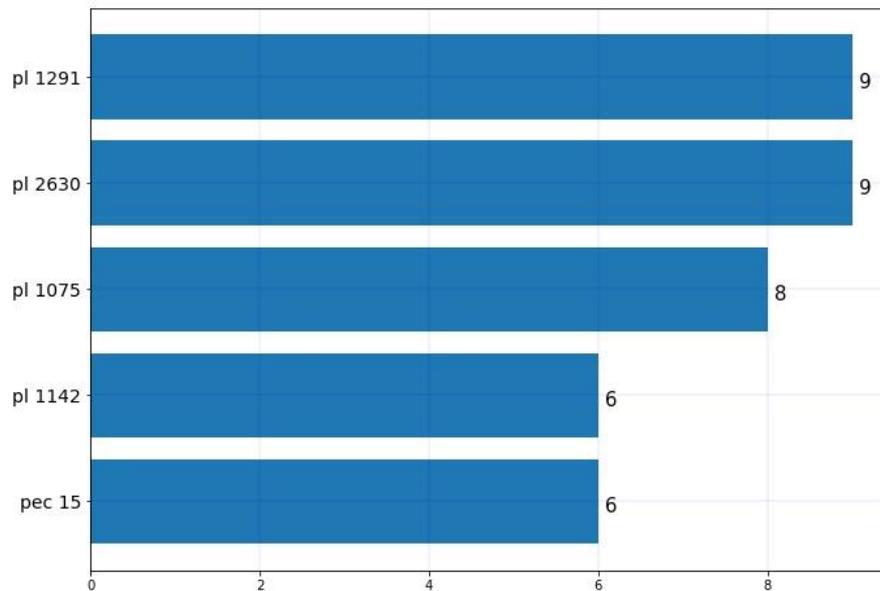
unificou 26 projetos sobre agricultura familiar, abrangendo implantação de acessos à água, condições de trabalho de mulheres agricultoras, entre outros temas relacionados. Em datas anteriores à votação, houve processos de mobilização de deputados de esquerda, antecipando vetos da base aliada do presidente.

Na segunda posição, com 146 menções, está a PEC 15/2015, que tornou o FUNDEB instrumento permanente da Constituição. Cabe destacar que o número de *tweets* relacionados a proposta seria significativamente maior, se considerado que muitos deputados se referiam a ela simplesmente usando a *hashtag* #Fundeb. Entre as mensagens, se menciona diversas vezes que foram 5 anos de discussão e preparo para submissão desta medida. Percebe-se também que o conjunto de *tweets* contendo celebrações a sua aprovação ultrapassou de visões ideológicas, partindo de deputados do PODE, PT, PP, PROS, PTB e REPUBLICANOS.

O PL 1075/2020, em terceiro lugar, dispõe sobre ações emergenciais destinadas ao setor cultural, enquanto as medidas de isolamento ou quarentena forem adotadas. Mobilizações, por meio da *hashtag* #AprovaEmergenciaCultural, e celebrações a aprovação do projeto partiram principalmente de deputados de Esquerda dos partidos PSB, PT e PSOL.

O contrário acontece nas reações ao Projeto de Lei 2630/2020, que institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, promovendo o combate a *fake news* e perfis automatizados nos sites de redes sociais. Deputados de direita, em grande parte dos partidos PSL e PROS, se mobilizaram contra e repudiaram a aprovação do que se referiam como “projeto da censura”.

Em último lugar da lista está o PL 3267/19, que modifica o Código de Trânsito Brasileiro, ampliando o prazo de validade das habilitações. As reações positivas a aprovação deste projeto foram em grande parte de deputados do DEM, REPUBLICANOS e PSL.

**Figura 22 - Menção a proposições no Twitter- Rio Grande do Sul**

**Fonte: Câmara dos Deputados, TSE e Twitter**

Após a aplicação do filtro explicado no início da seção, os 25.519 *tweets* publicados por deputados do Rio Grande do Sul foram reduzidos a 491 (1,92%). Deste conjunto foram extraídas 237 proposições. As primeiras, em número de menções, são os projetos de lei 1291/2020 e 2630/2020. O PL 1291/2020 assegura medidas de combate e prevenção à violência doméstica durante a declaração do estado de emergência em todo território nacional. Nas menções do Twitter, a *hashtag* #NãoÀViolênciaContraAMulher foi usada pela deputada Maria do Rosário (PT), autora do projeto, e outras colegas de bancada. A reação ao PL 2630/2020, relativo ao combate de *fake news*, foi semelhante à do âmbito nacional, havendo diversas críticas de deputados de direita. O mesmo pode ser dito em relação ao PL 1075/2020, sobre ações emergenciais ao setor cultural, que recebeu apoio da esquerda.

Em quarto lugar está o Projeto de Lei 1142/2020, que dispõe sobre medidas urgentes de apoio aos povos indígenas, comunidades ribeirinhas e quilombos em razão do novo coronavírus. Houve muita repercussão negativa a uma emenda do líder do governo no Senado que removia quilombos do projeto, excluindo estas comunidades do fornecimento de itens de higiene, água potável e leitos de hospitais. A PEC do Fundeb é a última da lista, com 6 menções. De forma semelhante ao âmbito geral, houve celebrações sobre sua aprovação em diferentes espectros ideológicos.

Destacadas as proposições mais relevantes do período estudado, pode-se explorar as dinâmicas incipientes a elas, tanto na Câmara quanto no Twitter. Estes

quadros de interação social são estudados aqui como redes. Nesta obra, todas as visualizações de dados em rede partiram da esquematização da Análise de Redes Sociais (ARS), definida por Recuero et al. (2015) como conjunto teórico de cunho estruturalista que descreve a relação entre atores sociais (nós) e suas conexões (arestas).

Nas últimas duas décadas, a união entre métodos da ARS e técnicas da Informática deu margem a novas linhas de pesquisa, ampliando a diversidade de objetos empíricos do campo (SILVA, 2012a). Nesta nova perspectiva, dados subjacentes às redes são analisados através de métricas que definem seus atributos em termos quantitativos (ARIF, 2015). A Tabela 10 reúne quatro destas métricas:

**Tabela 11 – Métricas usadas na Análise de Redes Sociais**

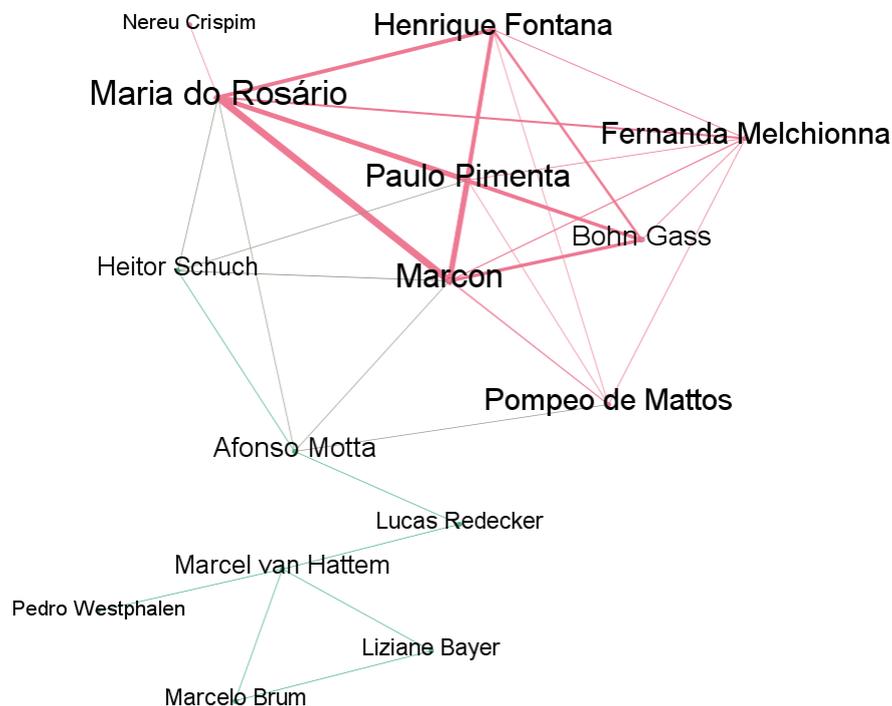
Densidade	Calcula a proporcionalidade das conexões de uma rede. O grafo é considerado denso se seu número de arestas se aproxima do valor máximo, e esparso caso contrário. Sendo o valor máximo de conexões definido pela fórmula: $n(n-1)/2$
Modularidade	Mede a divisão de uma rede em <i>clusters</i> , ou “comunidades”. Geralmente usada como método para detectar diferentes agrupamentos de um grafo.
Grau de centralidade	Quantifica o número de conexões incidente a um nó.
Grau de intermediação	Conta quantas vezes um nó é o “caminho mais curto” da rede. Nós com alto grau de intermediação desempenham um papel crucial no fluxo de informações e coesão dos seus meios.

**Fonte: Arif (2015), Bastian, Heymann e Jacomy (2009)**

A Figura 22 traz o grafo de coautoria das proposições submetidas por deputados gaúchos. Esta é a única do Gênero. Não foi possível fazer o mesmo ao âmbito geral, nem a uma “rede estendida” para ver a colaboração com parlamentares de outros estados, por falta de memória RAM na máquina usada para esta dissertação.

O processamento da visualização partiu de operações semânticas no arquivo “db-proposicoes.csv”, assim que filtrado pelo âmbito estadual. Após selecionar o nome dos autores, separados por vírgula, foi executado um código para criar a estrutura de dados do grafo (Célula 3.5). A partir disto, a biblioteca *networkx* foi usada para obter as métricas (Célula 3.7), e o programa *Gephi* para gerar a visualização.

**Figura 23 – Grafo de coautoria por deputados – Rio Grande do Sul**



**Fonte: Câmara dos Deputados, TSE e Twitter**

24 dos 31 deputados da bancada submeteram ao menos uma proposição no ano de 2020. Entretanto, o grafo apresentado na Figura 22 mostra apenas 15 nomes. A explicação para isto está na execução do filtro mencionado anteriormente: permaneceram apenas os deputados que compartilharam a autoria de proposições no âmbito Estadual. Bibó Nunes e Sanderson (PSL) foram removidos por ter apenas uma conexão entre si, possibilitando pouca fruição à análise. Deste modo, o número de nós chegou a 15, e de arestas 35. Aqui, o valor máximo de conexões é 105. Portanto, a densidade chega a 33,33%, uma rede parcialmente conectada. A métrica de modularidade, executada no *Gephi* com resolução de 1.0, retornou dois grupos. Na parte superior da imagem, em vermelho, está o maior em número de nós (8), representando 53,33% do grafo. Abaixo, em verde, está o grupo menor, com 7 nós configurando 46,67%.

A característica predominante desta divisão é a orientação ideológica, confirmando a observação de Da Silva (2012) sobre a formação das leis, apesar de tendências ao clientelismo e outras disformidades da política brasileira (PEREIRA,

2008), e da coesão nas votações a projetos de maior destaque, como visto na Figura 19.

A espessura das arestas representa a quantidade de conexões entre os deputados. Permanece a disparidade entre os dois *clusters*: no vermelho há 28, e no verde, 7. Já o tamanho dos rótulos traz o grau de centralidade. Se destacam deputados do grupo relativo à esquerda. Maria do Rosário e Marcon (PT) despontam o primeiro e segundo lugar, com 9 e 8 conexões.

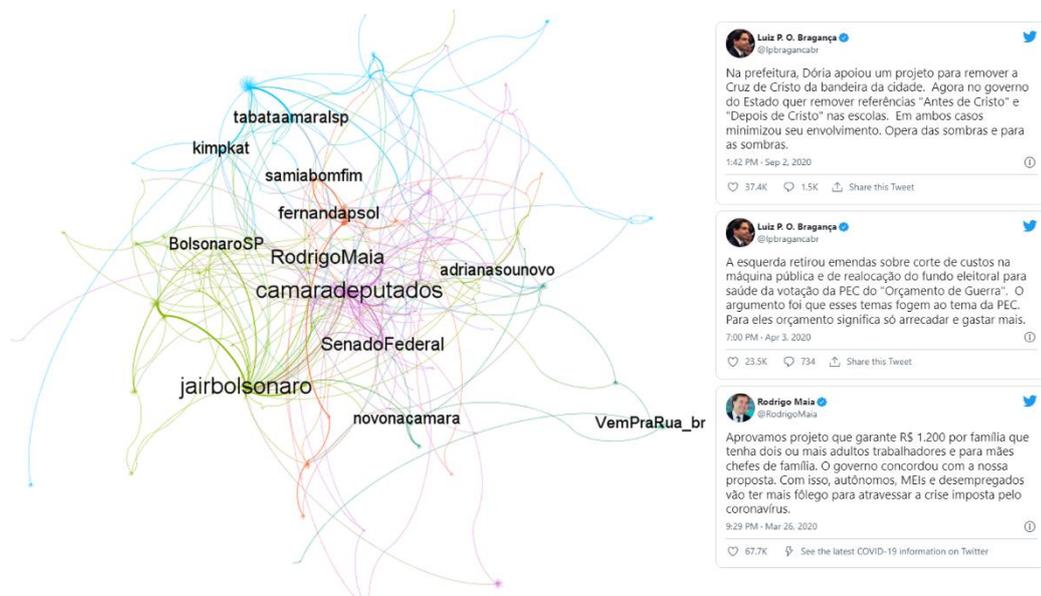
Há uma conexão latente entre o número de proposições submetidas em 2020 (dentro das categorias estudadas), e o ordenamento desta métrica. Primeiro e segundo lugar são os mesmos, com 91 e 81 proposições submetidas por Maria do Rosário e Marcon. Entre as exceções estão Fernanda Melchionna (PSOL), em sexto lugar no grau de centralidade, tendo 70 proposições, e Afonso Motta (PDT), em oitavo, com apenas 3. Outra inferência, ainda que indireta, é que até o quinto lugar todos os deputados fazem parte do maior agrupamento de seguidores (até 1 milhão 988 mil) e número de *tweets* (até 114 mil e 966). Apesar de Marcel van Hattem (NOVO) compartilhar destas características, e ter 27 proposições submetidas, situa-se no grupo menos concentrado, ficando na nona posição.

O grau de intermediação, não representado na Figura 22, traz resultados completamente diferentes. Os três primeiros lugares foram para Afonso Motta (PDF), Lucas Redecker (PSDB) e Marcel van Hattem (NOVO). Ou seja, estes foram os nós que mais serviram de “ponte” na rede. Isto pode ser notado na própria imagem. Afonso Motta (PDF) é o que mais faz conexão entre o grupo da esquerda e da direita. Lucas Redecker (PSDB) pontua nesta métrica por ser elo entre ele e Marcel van Hattem (NOVO) - a única “saída” do grupo mais isolado. Há poucas semelhanças no perfil destes deputados, sendo estes de Esquerda, Centro e Direita, respectivamente. O único fator em comum, ainda que não diretamente relacionado, está no fato de Afonso Motta (PDF) e Lucas Redecker (PSDB) serem do maior agrupamento de bens declarados (Até R\$38 milhões).

Concluída a exploração da rede parlamentar, cabe direcionar os métodos da ARS ao uso do Twitter. Para preparar as Figuras 23 e 24, foi feito um procedimento semelhante ao apresentado anteriormente (Célula 4.9). A diferença principal consistiu na extração de termos ligados à “@”, que denota compartilhamento de publicações ou menção a outros usuários no Twitter.

Neste grafo, referente a base de dados com menções aos trabalhos legislativos, não se distingue *retweets* de menções a usuários. O propósito destas figuras é abordar as interações na plataforma de modo geral. As conexões do subconjunto dos deputados do Rio Grande do Sul não foram suficientes para formar uma rede própria. Por este motivo, a Figura 24 traz o grafo de toda a Câmara, destacando o nome dos parlamentares da bancada gaúcha. As publicações com maior número de *retweets*, ao lado direito de cada imagem foram feitas com o mesmo método da Figura 13, onde os identificadores únicos dos *tweets* em destaque são pegos com Python, e depois renderizados no layout oficial da plataforma através de um código HTML.

**Figura 24 – Rede de Tweets mencionando proposições - Geral**



**Fonte: Câmara dos Deputados, TSE e Twitter**

Em termos de conectividade, o grafo da Figura 23 está extremamente aquém ao da anterior. Foram contadas 1248 interações. Se estivesse completamente conectado, o valor seria 471.906. Deste modo, a densidade aqui chega ao ínfimo 0,001%. Em relação ao número de nós, é consideravelmente maior. Ao todo, são listados 972 perfis. Esta informação pode parecer estranha, já que apenas 339 deputados falaram sobre proposições em seus *tweets*. Acontece que, apesar deste conjunto de usuários terem originado a rede, esta não se limita a eles.

Os parlamentares, assim como qualquer usuário da plataforma, tem a liberdade de compartilhar e mencionar quem quiser. Esta liberdade trouxe um quadro atípico:

apenas 225 dos perfis (23,14%) são de deputados federais. Esta considerável redução se explica pelo fato de existir quem, embora tenha publicado neste contexto, não foi compartilhado nem mencionado por seus colegas da Câmara.

Esta tendência pode ser observada nos *tweets* mais compartilhados. O primeiro e o segundo são de autoria de Luiz Philippe de Orléans e Bragança (PSL). No de maior destaque, o deputado denuncia "projetos" de João Dória, enquanto prefeito e governador de São Paulo, que visavam a laicidade das instituições públicas. O *tweet* subsequente traz uma crítica a deputados de esquerda que retiraram "emendas" sobre responsabilidade fiscal, e transferência de fundos eleitorais à saúde, da "PEC" do "Orçamento de Guerra". Este é o termo pelo qual ficou conhecido a Proposta de Emenda à Constituição 10/2020, que institui um regime extraordinário nas contas públicas frente a pandemia, e obteve o maior percentual de aprovação do ano.

O nome do deputado não aparece no grafo. Mesmo com o altíssimo alcance de seus *tweets*, as únicas conexões que estabelece dentro desta rede são três respostas a seus seguidores. Núcleos pequenos e isolados como este foram removidos do grafo para obter coesão nas métricas e uma visualização de dados adequada. Em terceiro lugar na lista de *retweets* está uma publicação de Rodrigo Maia (DEM), pela qual divulga a aprovação do "projeto" que garante R\$1200 a famílias em situação de vulnerabilidade social durante a pandemia.

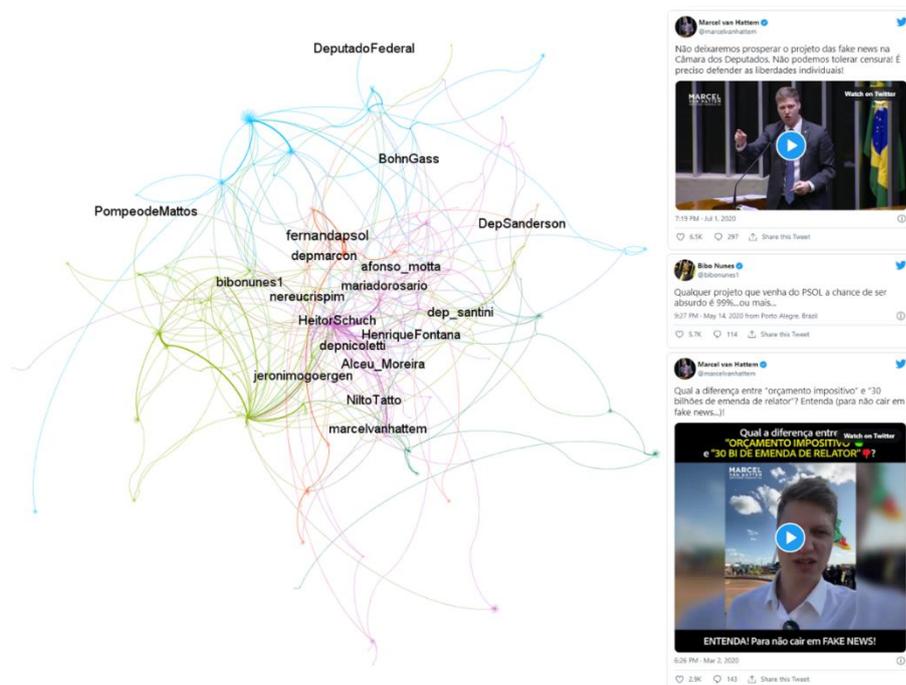
A execução do *script* de modularidade do *Gephi* na resolução 6.0 retornou cinco comunidades. O então presidente da Câmara, Rodrigo Maia, se encontra entre as duas maiores. A seção roxa, representando 33,33% da rede, é a que mais concentra perfis institucionais. No centro dela está "camaradosdeputados", com o segundo maior grau de centralidade em todo o grafo. Em seguida, neste nicho, se encontra, "SenadoFederal", "tvcamara", perfis de partido como "PSDBoficial" e "ptbrasil", e veículos de imprensa como "estadão" e "VEJA". No *cluster* verde claro, que compõe 25,97% da rede, Maia compartilha espaço com o presidente da república, Jair Bolsonaro (SEM PARTIDO), que possui o maior grau de centralidade do grafo.

Abranches (2018) considera comum certa tensão entre líderes do executivo e do legislativo, devido a características estruturais do sistema político brasileiro. Contudo, o histórico de interações destas duas autoridades extrapola quaisquer expectativas de conflitos e divergências (CNN BRASIL, 2020). A presença de diversos políticos da base aliada do presidente, como Eduardo Bolsonaro, Bia Kicis e Carla

Zambelli (PSL), indica que as menções à Maia neste *cluster* foram em sua maioria negativas.

Na comunidade azul claro, que representa 18,08% da rede, há muitos deputados de primeiro mandato e com alto engajamento no Twitter. Se destacam, neste conjunto os perfis de Tabata Amaral (PDT) e Kim Kataguirí (DEM), em primeiro e segundo lugar no grau de centralidade de intermediação. O *cluster* laranja, 12,70%, reúne diversos deputados de esquerda, prevalecendo em centralidade Fernanda Melchionna e Sâmia Bomfim (PSOL). A menor comunidade da rede, em verde escuro, evidencia parlamentares e perfis oficiais do Partido Novo, assim como movimentos sociais de direita como o “Vem pra Rua”.

**Figura 25 - Rede de Tweets mencionando proposições – Rio Grande do Sul**



**Fonte: Câmara dos Deputados, TSE e Twitter**

Na Figura 24, o mesmo grafo é apresentado, destacando os 18 deputados gaúchos que interagiram neste contexto. A lista de publicações mais compartilhadas põe em destaque Marcel Van Hattem (NOVO), no grupo relativo ao seu partido, e Bibi Nunes (PSL), entre o âmbito dos deputados de primeiro mandato e os favoráveis ao presidente. Na primeira e terceira publicação em número de *retweets*, Van Hattem se posiciona contra o “projeto” das *fake News* (em referência ao PL 2630/2020, que institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet), e explica ao seu público a diferença entre emendas de relator e Orçamento Impositivo.

Na segunda publicação de maior alcance, Bibo Nunes ridiculariza projetos do PSOL. Estes deputados também se destacam no grau de intermediação, atingindo respectivamente a terceira e quinta maior métrica.

A partir da modularidade do grafo, percebe-se que o *cluster* laranja traz perfis presentes do âmbito da esquerda da rede de coautoria de proposições. Entre estes, Fernanda Melchionna (PSOL), com maior grau de centralidade e de intermediação da bancada, e Marcon (PT). A deputada petista Maria do Rosário se situa na comunidade mais “institucional”, em roxo, tendo o segundo maior grau de intermediação. Afonso Motta não retoma o papel de “conector”, ficando em 13º lugar em intermediação.

Esta seção se distinguiu da anterior por não ter exaurido a prospecção dos dados. Em poucas palavras, sobrou possibilidades, mas faltou memória, processamento etc. A partir dos recursos computacionais disponíveis, tanto na máquina local quanto em instâncias do *Google Collab*, não foi possível explorar grandes redes. Ainda assim, a discussão sobre os dados atendeu o objetivo de pesquisa previamente estabelecido, uma vez que foi apontada a relação entre as redes de proposições legislativas e as formadas a partir de *tweets*. Encaminha-se, deste modo, as sínteses e reflexões na conclusão desta obra.

## CONSIDERAÇÕES FINAIS

Resultados obtidos nesta dissertação, por meio de revisões de literatura e ferramentas de análise, evidenciam a importância de acompanhar o processo parlamentar, peça fundamental ao funcionamento da democracia brasileira. É seguro dizer que isso acontece no Twitter, visto os milhões de usuários que seguem o conjunto de 513 deputados federais estudados. A pergunta de pesquisa "Quão consistentes são os discursos dos deputados federais no Twitter, se comparados às suas proposições parlamentares?" se voltou a eficácia deste fenômeno, buscando medir o quanto que as atividades legislativas transparecem nos discursos políticos do âmbito digital.

Encontrar o nível de consistência entre temas recorrentes nas proposições e nesta mídia social foi o primeiro passo para responder este questionamento. O percentual encontrado no quadro geral foi de 31,97%, e na bancada do Rio Grande do Sul, 31,07%. Como foi mencionado anteriormente, há muito espaço para o aperfeiçoamento dos procedimentos e códigos executados, como por exemplo, a aplicação de métodos de Processamento de Linguagem Natural (PLN).

Apesar destas limitações técnicas, não há como ignorar a baixa correspondência entre os dois campos. As publicações mais compartilhadas no Twitter indicam que, apesar desta plataforma ser um dos principais meios de discussão política do mundo digital, há pouca demanda para debates aprofundados sobre a formação de leis, ou criação de consenso sobre diretrizes sociais.

O recorte temporal desta dissertação foi completamente atípico. A cobertura midiática e a exposição a primeira pandemia global levou a uma adaptação compulsória dos trabalhos da Câmara, e portanto dos posicionamentos políticos nas mídias sociais, à temas ligados à saúde. Deste modo, é compreensível compartilhamento de notícias ou recomendações sanitárias da OMS pelos deputados, como foi visto na base de dados. Entretanto, outros assuntos tiveram mais visibilidade que as pautas do Congresso, tais como reviravoltas no quadro de alianças do presidente, escândalos cobertos pela imprensa, e questões identitárias.

Ao explorar a relação entre as redes de trabalhos legislativos e as formadas a partir de *tweets*, foi encontrada uma conjuntura semelhante. A divulgação de projetos de lei, propostas emendas à Constituição, entre outras proposições, representando cerca de 1% da base de dados, ocorreram em geral em três formas: Mobilização de

seguidores a período de votação, celebração de propostas aceitas no Congresso, e menções genéricas a natureza de trabalhos parlamentares de aliados ou oponentes de bancada. Foram poucos os casos em que se destacaram tentativas de aproximar os públicos a questões técnicas, e essenciais à função legislativa, ou de estabelecer consultas populares, que os integrassem a tomadas de decisão.

Neste ponto, pode-se valer do referencial teórico acumulado ao longo desta obra<sup>13</sup>. Existe uma dificuldade cada vez mais expressiva de manter ideais democráticos na era da informação. No começo do século, foi muito expresso um otimismo de que de que cada geração se tornará mais digital que a anterior, e portanto o futuro estaria “na mão dos jovens” (NEGROPONTE, 1996, p.231). Até mesmo Castells, empirista em seus primeiros diagnósticos sobre a internet, afirmou em certo momento a esperança de que “se pessoas forem esclarecidas, atuantes e se comunicarem em todo o mundo”, uma onda de solidariedade global aplacaria a defasagem entre o desenvolvimento tecnológico e humano, tornando o sonho do iluminismo uma realidade (Castells, 2009, p.429).

Este é um ponto interessante, e muito discutido nesta dissertação. A fundamentação teórica do estado moderno perpassa garantias mínimas à vida, à igualdade e à liberdade (VIEIRA, 2019). Contudo, em países com passado colonial, como o Brasil, isso sempre foi um blefe. Através do trabalho de Schwarcz e Starling (2015), se constata que a corte portuguesa compartilhava da mesma educação, costumes e referências da França pré-revolucionária, o que não a impediu de esmagar movimentos com a Conjuração Baiana e a Revolta dos Malês. É importante destacar que a corte não fez isto sozinha, o sistema escravocrata era protegido por um poder central, na figura do império, e mantido por poderes regionais, na forma de oligarquias e gerações de donos de terra. Ou seja, a desigualdade foi reproduzida de modo social, e não somente político.

Ainda que seja inegável o aumento da participação popular no governo e o enorme potencial criativo desvelado pelos meios de comunicação digital, é importante manter em mente que as relações de poder e o pensamento “tradicional” do Brasil não deixou de existir do dia para noite. Por este motivo, considerar muito “analógicas” críticas como a da escola de Frankfurt e de DeFleur e Ball-Rockeach (1993), que

---

<sup>13</sup> Em futuros estudos, buscaremos atualizar o referencial teórico com mais trabalhos interpretativos sobre a obra de Habermas, e com maior volume de materiais relativos a polarização dos debates digitais, ideologias, partidos e processo de formação política no Brasil.

denunciavam a repercussão dos ideais falidos da modernidade, é cair na falácia no determinismo tecnológico.

Nesta visão, novas tecnologias são a principal causa de mudança histórica, tanto em nível macrossocial quanto em processos sociais e psicológicos sutis (OXFORD REFERENCE, 2021). Um exemplo disso é esta crença de que as propriedades técnicas da internet tornariam a sociedade mais conectada, descentralizada e inclusiva. É verdade que a atual realidade midiática oferece possibilidades ilimitadas e muitas vezes gratuitas mas, como coloca Lanier (2018), para isso condiciona pessoas em produtos, sistematizando seus comportamentos a fim de monetizá-los.

Este modelo de negócios tomou proporções geopolíticas. Christopher Wylie (2018) conta em seu testemunho ao Congresso Americano que toda propriedade intelectual e ativos de dados da Cambridge Analítica, agência responsável pela campanha de Donald Trump em 2016, foi usado durante décadas para mitigar experiências democráticas, capitalizar o descontentamento e alimentar tensões étnicas em países da África e da Ásia. (WYLIE, 2018).

Santos e Silveira (2001) não desconsideram o papel da solidariedade na resolução de problemas como este. Contudo, os autores fazem uma distinção. De um lado, existe a solidariedade orgânica, que parte da interdependência entre ações e atores na construção da autonomia e do bem comum. Do outro, a solidariedade organizacional, que mecaniza este processo para atender a interesses particulares.

O mesmo pode ser dito sobre os processos de mobilização no Twitter. A maneira como os parlamentares estudados agiram na plataforma não diferiu muito a de influenciadores digitais. Enjolras (2014) encontra tendência semelhante em seu estudo sobre parlamentares noruegueses. Mesmo neste país, com baixíssimo índice de desigualdade social, os *tweets* dos políticos também dão maior destaque a aspectos de suas vidas privadas, ou tópicos que geram mais engajamento como esportes, música e cultura pop.

Aqui cabe uma contribuição da Análise de Redes Sociais: os nós centrais de uma rede tendem a concentrar o número de arestas, devido a um fenômeno chamado de “conexão preferencial” (SILVA; STABILE, 2016). Neste caso, quando um deputado obtém um público expressivo, e consegue antecipar razoavelmente suas reações a conteúdos amplamente difundidos na rede, a tendência é que continue crescendo independente se segue ou não as atribuições de transparência do cargo.

É preciso fazer um adendo sobre a natureza das discussões na plataforma. Devido aos seus perfis minimalistas, com foco na constante publicação de mensagens curtas e instantâneas (SILVA, 2012), é mais adequado pensar no Twitter como uma rede de promoção pessoal, do que em uma voltada para debates. Ao contrário do que se possa pensar, isso é bom para a democracia. De acordo com Habermas (1984), o ego pode influenciar visões de mundo informando, encorajando ou as advertindo criticamente. O autor desta dissertação estuda o Twitter desde 2014. Não existe espaço digital mais diverso em opiniões e aberto. O que falta nos circuitos desta rede, tanto na esquerda quanto na direita são dados factuais.

Há uma tendência perigosa de marginalizar fatos, enaltecendo interpretações e narrativas. Isso acontece porque o mundo está ficando cada vez mais complexo, sendo mais fácil escantear a realidade objetiva, a ajustando a perspectivas particulares. A predominância de dissensos irremediáveis nas mídias sociais decorre deste contexto, uma vez que um consenso mínimo sobre fatos é necessário para discutir temas de interesse público (MAGRANI; OLIVEIRA, 2018).

A crítica de Lyotard (1986) a Habermas indica que a busca por um consenso universal não é o parâmetro certo para avaliar o campo das discussões, pois dissensos também fazem parte de seu processo constitutivo. Deste modo, se torna pouco razoável esperar um cenário onde todos concordam, ou cair em um reducionismo onde nada se agrega no meio digital. Manovich (2018) acredita que a alfabetização digital hoje requer mais agência no uso das mídias sociais, de modo que seus usuários entendam por que cada anúncio e conteúdo é mostrado.

Pode-se acrescentar aí a necessidade de maior escrutínio a informações compartilhadas no âmbito político. De Faria (2012) e outros autores citados discorrem sobre o fracasso de iniciativas com esse objetivo. Sejam governamentais ou civis, estas geralmente erram ao esperar que políticos divulguem seu trabalho de modo espontâneo e pragmático, e que o público mude seu comportamento (sites que acessam ou como se expressam no meio digital) por motivos altruístas. Atualmente a Câmara dos Deputados e o Tribunal Superior de Contas aderem a uma política institucional de dados abertos. Esta deliberação mostra respaldo da democracia brasileira com as tecnologias digitais.

Entretanto, as dificuldades encontradas no processamento de dados públicos indicam falta de zelo das autoridades responsáveis. Sem conhecimentos prévios em informática, não seria possível cumprir os objetivos de pesquisa em tempo hábil. O

caso da declaração de bens dos candidatos no site do TSE é o exemplo perfeito. No arquivo, os candidatos são identificados somente por código, e os valores são divididos a cada bem declarado. Ainda que possa existir uma razão técnica para isso, envolvendo o banco de dados ou outros programas utilizados pelo Tribunal, não é justo delegar as informações somente a quem tenha conhecimento para decifrá-las. Especialmente no atual quadro da difusão da internet, onde a maioria da população a acessa através de dispositivos e redes móveis.

A forma como as entidades públicas vem cumprindo as normas de transparência não está completamente errada. O que falta é mais diálogo com a sociedade civil. O Congresso Nacional, ao menos, está se adaptando a novas realidades. Uma vez que a ideia de uma plataforma de participação ubíqua se mostrou inviável, investiu-se na descentralização dos serviços e informações da instituição.

Por meio desta e de outras medidas, imensas bases de dados tornaram-se passíveis de consulta. As possibilidades de pesquisa com estas amostras são ainda maiores, se somadas as consultas acadêmicas no Twitter. Esta dissertação representa uma fração mínima deste campo. O esforço à documentação das técnicas utilizadas esteve presente da conceitualização a conclusão do trabalho. Acredita-se que a reprodutibilidade dos resultados encontrados facilitará a validação, refutação ou apropriação destes em outros contextos, fomentando a análise científica de fenômenos comunicacionais.

A idoneidade dos dados não implica em "neutralidade ideológica". Segundo Said (2007), é impossível separar o erudito das circunstâncias da vida e um conjunto de crenças de sua posição social. Portanto, o ideal desta obra pode ser resumido em um único ponto: o bem comum deve pesar mais do que o interesse da classe política. Isto não implica em todas as pessoas pensando igual, com um único objetivo. Diferentes interesses de classe podem ser estruturados organicamente, de modo que se harmonizem suas diversidades.

Esta asserção leva ao conceito central desta obra. Apesar do referencial bibliográfico levantado, e da discussão sobre os dados dentro dos objetivos de pesquisa, ainda há margem para um último questionamento. Afinal, a Esfera Pública referida aqui se trata da Câmara dos Deputados ou do Twitter?

Como destacado nos primeiros capítulos, esta varia em contexto, história e tecnologia. Tanto nos meios que se dispõe, indo da ágora às plataformas digitais, ao seu condicionamento social, ou seja, se de fato dá voz a todos ou se restringe a um

seleto grupo de participantes. Na leitura de Rüdiger (2009) sobre a teoria da ação comunicativa, se destacam referências às “mídias sistêmicas”, que prescrevem a racionalização do mundo da vida, desonerando as pessoas da necessidade de se comprometerem com a própria comunicação.

Interpretar esta passagem pelo viés do determinismo tecnológico implica em atribuir a sistematicidade a meios tradicionais, tais quais a televisão e o rádio, e a descentralização e horizontalidade os meios digitais. Contudo, fenômenos abordados aqui, como a polarização do debate público e a possibilidade de manipular canais de informação muito acessados, invalidam este ponto de vista.

Consideramos então que as mídias sociais, epítome do mundo digital na atualidade, são alvo de duas tendências conflitantes: impulsos sistêmicos, pelos quais prevalecem relações de poder pré-determinadas pela concentração de capital político e econômico (LUBENOW, 2015), e comunidades ideais de comunicação, onde formação de consenso se dá através da racionalização coletiva de questões normativas, a fim de promover a tolerância mútua e convivência em comum entendimento (PERLATTO, 2012).

Mesmo se a “autonomia do mundo da vida”, ou a “solidariedade orgânica” se concretizassem em uma mídia social, esta não definiria inteiramente o estado da esfera pública na atualidade. A definição tradicional de Habermas ressalta que este espaço de comunicação não pode ser limitado ou restringido, desde que integre processos de deliberação pública, onde temas e contribuições de esferas autônomas sejam institucionalizados por sistemas políticos e administrativos estabelecidos democraticamente (LUBENOW, 2015).

Este aspecto multidimensional implica que a Esfera Pública na Era da Informação pode ser caracterizada pela interconexão das mídias sociais com outros sistemas sociotécnicos latentes, à medida que cultura, economia e política são digitalizadas. Portanto, a busca por transparência e maior participação política deve ser transversal a academia, imprensa e a sociedade civil, para desenvolver não só inteligência, mas sensibilidade e autoconhecimento coletivo que fortaleçam o laço social e o garantam funcionamento das instituições.

## REFERÊNCIAS

- ABRANCHES, Sérgio Henrique. **Presidencialismo de coalizão: raízes e evolução do modelo político brasileiro**. São Paulo, Brazil: Companhia das Letras, 2018.
- AGRESTI, Alan; FINLAY, Barbara. **Métodos estatísticos para as ciências sociais**. Porto Alegre, Brasil: Penso Editora, 2012.
- ANDRADE, Aparecida de Moura; COUTINHO, Robson Luiz Fialho. Processo Legislativo nas comissões da Câmara dos Deputados. **Brasília: Câmara dos Deputados, [S. l.], 2019.**
- ARAÚJO, Willian Fernandes. As narrativas sobre os algoritmos do Facebook: uma análise dos 10 anos do feed de notícias. *[S. l.]*, 2017.
- ARIF, Tasleem. The mathematics of social network analysis: metrics for academic social networks. **International Journal of Computer Applications Technology and Research, [S. l.]**, v. 4, n. 12, p. 889–93, 2015.
- BASTIAN, Mathieu; HEYMANN, Sebastien; JACOMY, Mathieu. Gephi: an open source software for exploring and manipulating networks. *In: 2009, Anais [...]. In: THIRD INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA. [s.l: s.n.]*
- BASTOS, Marco Toledo. Public Opinion Revisited: The propagation of opinions in digital networks. **Journal of Arab & Muslim Media Research, [S. l.]**, v. 4, n. 2–3, p. 185–201, 2012.
- BBC. Twitter “permanently suspends” Trump’s account. **BBC News, [S. l.]**, 2021. Disponível em: <https://www.bbc.com/news/world-us-canada-55597840>. Acesso em: 14 dez. 2021.
- BERTALAN, Vithor Gomes; SERON RUIZ, Evandro Eduardo. Using topic modeling to find main discussion topics in brazilian political websites. *In: PROCEEDINGS OF THE 25TH BRAZILLIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB, WEBMEDIA 2019 2019, Rio de Janeiro, Brazil. Anais [...]. Rio de Janeiro, Brazil: ACM Press, 2019. p. 245–248. DOI: 10.1145/3323503.3360644.*
- BESSA, Paula Alice Müller; DE AZEVEDO, Ione Galoza; DE SOUSA, Carlos Henrique Medeiros. Das Ruas às Redes: A informação e a comunicação no (re) pensar e (re) fazer política. *[S. l.]*, 2016.
- BUSH, Vannevar. As we may think. **The atlantic monthly**, Boston, p. 101–108, 1945.
- CARAZZA, Bruno. **Dinheiro, eleições e poder: as engrenagens do sistema político brasileiro**. São Paulo, Brazil: Companhia das Letras, 2018.
- CASTELLS, Manuel. **A sociedade em rede. A era da informação: economia, sociedade e cultura volume I**. São Paulo: Editora Paz e Terra, 1999. a.

CASTELLS, Manuel. **O poder da identidade. A era da informação: Economia, sociedade e cultura volume II.** São Paulo: Paz e Terra, 1999. b.

CASTELLS, Manuel. **O fim de milênio. A era da informação: Economia, sociedade e cultura volume III.** São Paulo: Paz e Terra, 1999. c.

CASTELLS, Manuel. The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance. **The ANNALS of the American Academy of Political and Social Science**, [S. l.], v. 616, n. 1, p. 78–93, 2008. DOI: 10.1177/0002716207311877.

CASTELLS, Manuel. **Redes de indignação e esperança: movimentos sociais na era da internet.** Rio de Janeiro: J. Zahar, 2013.

CGI.BR. **Pesquisa Sobre o Uso das Tecnologias de Informação e Comunicação nos Domicílios Brasileiros.** São Paulo.

CGI.BR. **Painel TIC Covid-19: Pesquisa sobre o uso da internet no Brasil durante a pandemia do novo Coronavírus.** São Paulo.

CHAGAS, Viktor; TOTH, Janderson. Monitorando memes me mídias sociais. **Monitoramento e pesquisa em mídias sociais: metodologias, aplicações e inovações.** São Paulo: Uva Limão, [S. l.], 2016.

CIDADÃ, AUDITORIA. **Twitter dos Deputados Federais 2020.** 2020. Disponível em: <https://auditoriacidadada.org.br/contato-dos-deputados-federais-2020-twitter/>.

CNN BRASIL. **Bolsonaro e Maia: ataques em entrevistas refletem histórico de relação tensa.** 2020. Disponível em: <https://www.cnnbrasil.com.br/politica/bolsonaro-e-maia-ataques-em-entrevistas-refletem-historico-de-relacao-tensa/>. Acesso em: 10 dez. 2021.

COSTA, Sérgio. A democracia e a dinâmica da esfera pública. **Lua Nova: Revista de Cultura e Política**, [S. l.], n. 36, p. 55–65, 1995. DOI: 10.1590/S0102-64451995000200004.

DA SILVA, Deborah Ramos; KERBAUY, Maria Teresa Miceli. Eleições 2018 e a forte influência das redes sociais. **Liberdade de Expressão Questões da atualidade**, [S. l.], p. 125, 2020.

DA SILVA, Felipe Carreira. Habermas e a Esfera Pública reconstruindo a História. **Sociologia, Problemas e Práticas**, [S. l.], v. 35, p. 117–138, 2001.

DA SILVA, José Afonso. Estrutura e funcionamento do Poder Legislativo. **Revista de informação legislativa**, [S. l.], v. 47, n. 187, p. 137–154, 2010.

DA SILVA RESENDE, Ciro Antônio; DA MATA, Janaina Ferreira; COSTA, Macdowell Batista; RAMOS, Vanessa Quirino; PEREIRA, Marcus Abílio. Gabinete digital: uma experiência de democracia na rede. *In*: CONFERENCE ODP, PORTO ALEGRE, BRAZIL 2012, **Anais [...]**. [s.l: s.n.]

DE FARIA, Cristiano Ferri Soares. **O parlamento aberto na era da internet: pode o povo colaborar com o Legislativo na elaboração das leis?** [s.l.] : Edições Câmara, 2012. v. 18

DE SOUSA JÚNIOR, João Henriques; RAASCH, Michele; SOARES, João Coelho; DE SOUSA, Letícia Virgínia Henriques Alves. Da Desinformação ao Caos: uma análise das Fake News frente à pandemia do Coronavírus (COVID-19) no Brasil. **Cadernos de Prospecção**, [S. l.], v. 13, n. 2 COVID-19, p. 331, 2020.

DEFLEUR, Melvin L.; BALL-ROCKACH, Sandra. **Teorias da Comunicação de Massa**. Rio de Janeiro: Zahar, 1993.

**Digital 2021**. [s.d.]. Disponível em: <https://wearesocial.com/digital-2021>. Acesso em: 20 jul. 2021.

**Digital in Brazil**. 2021. Disponível em: <https://datareportal.com/reports/digital-2021-brazil>. Acesso em: 19 jul. 2021.

ENJOLRAS, Bernard. How politicians use Twitter and does it matter? The case of Norwegian national politicians. [S. l.], 2014. DOI: 10.13140/2.1.3177.3123. Disponível em: <http://rgdoi.net/10.13140/2.1.3177.3123>. Acesso em: 13 dez. 2021.

FUCHS, Christian. Mídias Sociais e a Esfera Pública. **Revista Contracampo**, [S. l.], v. 34, n. 3, 2016. DOI: 10.20505/contracampo.v34i3.912. Disponível em: <http://www.contracampo.uff.br/index.php/revista/article/view/912>. Acesso em: 13 out. 2021.

GARCIA, David; MENDEZ, Fernando; SERDÜLT, Uwe; SCHWEITZER, Frank. Political polarization and popularity in online participatory media: An integrated approach. *In*: INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, PROCEEDINGS 2012, **Anais [...]**. [s.l.: s.n.] p. 3–10. DOI: 10.1145/2389661.2389665.

GARCIA, Leila Posenato; DUARTE, Elisete. **Infodemia: excesso de quantidade em detrimento da qualidade das informações sobre a COVID-19** SciELO Public Health, , 2020.

GILLESPIE, Tarleton. **Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media**. Londres: Yale University Press, 2018.

GOMES, Itania Maria Mota. O infotainment e a cultura televisiva." A TV em transição. **Tendências de programação no Brasil e no mundo**, [S. l.], n. 1, p. 195–221, 2009.

GONAWELA, A'Ndre; KUMAR, Reeshma; THAWANI, Udit; AHMAD, Dina; CHANDRASEKARAN, Ramgopal; PAL, Joyojeet. The Anointed Son, The Hired Gun, and the Chai Wala: Enemies and Insults in Politicians' Tweets in the Run-Up to the 2019 Indian General Elections. *In*: PROCEEDINGS OF THE 53RD HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES 2020, **Anais [...]**. [s.l.: s.n.] p. 10. DOI: 10.24251/hicss.2020.352.

GORKOVENKO, Katerina; TAYLOR, Nick. Politics at home: Second screen behaviours and motivations during TV Debates. *In: ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES 2016*, Gothenburg, Sweden. **Anais [...]**. Gothenburg, Sweden: ACM Press, 2016. p. 1–10. DOI: 10.1145/2971485.2971514.

GORKOVENKO, Katerina; TAYLOR, Nick. Audience and expert perspectives on second screen engagement with political debates. *In: TVX 2019 - PROCEEDINGS OF THE 2019 ACM INTERNATIONAL CONFERENCE ON INTERACTIVE EXPERIENCES FOR TV AND ONLINE VIDEO 2019*, Salford (Manchester), United Kingdom. **Anais [...]**. Salford (Manchester), United Kingdom: ACM Press, 2019. p. 70–82. DOI: 10.1145/3317697.3323352.

HABERMAS, Jürgen. **The Theory of Communicative Action, Volume 1**. Cambridge: Polity, 1984.

HABERMAS, Jürgen. **The structural transformation of the public sphere: an inquiry into a category of bourgeois society**. Cambridge: MIT press, 1989.

HABERMAS, Jürgen. **The Theory of Communicative Action: Lifeworld and Systems, a Critique of Functionalist Reason, Volume 2**. Hoboken: John Wiley & Sons, 2015. v. 2

HU, Mengdie; WONGSUPHASAWAT, Krist; STASKO, John. Visualizing social media content with sententree. **IEEE transactions on visualization and computer graphics**, [S. l.], v. 23, n. 1, p. 621–630, 2016.

JERÔNIMO, Luciana Saraiva de Oliveira. **Diálogo temático on-line na consulta pública digital: Um estudo sobre relações entre enunciados do governo e dos cidadãos**. 2016. Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brasil, 2016.

LAMOUNIER, Bolívar. O que é que se constrói quando se constrói a democracia. *In: AVELAR, Lúcia; CINTRA, Antônio Octávio (org.)*. **Sistema político brasileiro: uma introdução**. São Paulo: Unesp, 2015. v. 3.

LANIER, Jaron. **Ten arguments for deleting your social media accounts right now**. Nova York: Henry Holt, 2018.

LATOURETTE, Bruno. **Jamais fomos modernos**. São Paulo: Editora 34, 1994.

LE, Huyen; BOYNTON, G. R.; MEJOVA, Yelena; SHAFIQ, Zubair; SRINIVASAN, Padmini. Bumps and bruises: Mining presidential campaign announcements on Twitter. *In: HT 2017 - PROCEEDINGS OF THE 28TH ACM CONFERENCE ON HYPERTEXT AND SOCIAL MEDIA 2017*, Prague, Czech Republic. **Anais [...]**. Prague, Czech Republic: ACM Press, 2017. p. 215–224. DOI: 10.1145/3078714.3078736.

LIN, Yu-Ru; MARGOLIN, Drew; KEEGAN, Brian; LAZER, David. Voices of victory: A computational focus group framework for tracking opinion shift in real time. *In: 2013, Anais [...]*. *In: PROCEEDINGS OF THE 22ND INTERNATIONAL CONFERENCE ON WORLD WIDE WEB*. [s.l: s.n.] p. 737–748.

LOKOT, Tetyana; DIAKOPOULOS, Nicholas. News Bots: Automating news and information dissemination on Twitter. **Digital Journalism**, [S. l.], v. 4, n. 6, p. 682–699, 2016.

LUBENOW, Jorge Adriano. Esfera pública e democracia deliberativa em Habermas: modelo teórico e discursos críticos. **Kriterion: Revista de Filosofia**, [S. l.], v. 51, n. 121, p. 227–258, 2010. DOI: 10.1590/S0100-512X2010000100012.

LUBENOW, Jorge Adriano. **A Categoria de Esfera Pública em Jürgen Habermas: Para Uma Reconstrução da Autocrítica**. [s.l.] : EDITORA CRV, 2015. DOI: 10.24824/978854440627.4. Disponível em: <https://editoracriv.com.br/produtos/detalhes/31418-detalhes>. Acesso em: 13 out. 2021.

LYOTARD, Jean François. **O pós-moderno**. Rio de Janeiro: Jose Olympio, 1986.

MAGRANI, Eduardo; OLIVEIRA, Renan Medeiros De. A esfera pública (forjada) na era das fake news e dos filtros-bolha. **Cadernos Adenauer**, Fake news e as eleições. [S. l.], v. 19, Fake news e as eleições, p. 26, 2018.

MALINI, Fabio. O comunismo das redes: sistema midiático, p2p, colaboração em rede e novas políticas de comunicação na Internet. **Universidade Federal do Rio de Janeiro, Tese de Doutorado**, [S. l.], 2007.

MANOVICH, Lev. Cultural analytics: visualising cultural patterns in the era of “more media”. **Domus March**, [S. l.], 2009.

MANOVICH, Lev. **Software takes command**. Londres: A&C Black, 2013.

MANOVICH, Lev. 100 Billion Data Rows per Second: Culture Industry and Media Analytics in the Early 21st Century. **International Journal of Communication**, [S. l.], 2018.

MARAZZI, Christian. **La Place des chaussettes: Le tournant linguistique de l'économie et ses conséquences politiques**. [s.l.] : Éditions de l'Éclat, 1997.

MASCARO, Christopher; AGOSTO, Denise; GOGGINS, Sean P. One-sided conversations: The 2012 presidential election on twitter. *In*: ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES 2016, Shanghai, China. **Anais [...]**. Shanghai, China: ACM Press, 2016. p. 112–121. DOI: 10.1145/2912160.2912185.

MONTI, Corrado; ROZZA, Alessandro; ZAPPELLA, Giovanni; ZIGNANI, Matteo; ARVIDSSON, Adam; COLLEONI, Elanor. Modelling political disaffection from Twitter data. *In*: PROCEEDINGS OF THE 2ND INTERNATIONAL WORKSHOP ON ISSUES OF SENTIMENT DISCOVERY AND OPINION MINING, WISDOM 2013 - HELD IN CONJUNCTION WITH SIGKDD 2013 2013, Chicago, Illinois. **Anais [...]**. Chicago, Illinois: ACM Press, 2013. p. 1–9. DOI: 10.1145/2502069.2502072.

**MP sobre assinatura eletrônica em documentos públicos recebe 76 emendas**. Senado Federal. 2020. Disponível em: <https://www12.senado.leg.br/noticias/materias/2020/06/23/mp-sobre-assinatura-eletronica-em-documentos-publicos-recebe-76-emendas>. Acesso em: 5 dez. 2021.

NEGROPONTE, Nicholas. **Being digital**. Londres: Hodder & Stoughton, 1995.

NGUYEN, James. Politics and the Twitter Revolution: A Brief Literature Review and Implications for Future Research. **Social Networking**, [S. l.], v. 7, n. 4, p. 243–251, 2018.

NICOLAU, Jairo Marconi. **Representantes de quem? os (des)caminhos do seu voto da urna à Câmara dos Deputados**. Rio de Janeiro, RJ: Zahar, 2017.

NIEBORG, David B.; POELL, Thomas. The platformization of cultural production: Theorizing the contingent cultural commodity. **New Media and Society**, [S. l.], v. 20, n. 11, p. 4275–4292, 2018.

OLIVEIRA, Vânia Aparecida Rezende De; CANÇADO, Airton Cardoso; PEREIRA, José Roberto. Gestão social e esfera pública: aproximações teórico-conceituais. **Cadernos EBAPE.BR**, [S. l.], v. 8, n. 4, p. 613–626, 2010. DOI: 10.1590/S1679-39512010000400004.

OXFORD REFERENCE. **Technological Determinism**. 2021. DOI: 10.1093/oi/authority.20110803102813253. Disponível em: <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803102813253>. Acesso em: 26 dez. 2021.

PANDA, Anmol; CHAKRABORTY, Sunandan; RAVAL, Noopur; ZHANG, Han; MOHAPATRA, Mugdha; AKBAR, Syeda Zainab; PAL, Joyojeet. Affording extremes: Incivility, social media and democracy in the indian context. *In*: ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES 2020, Guayaquil Ecuador. **Anais [...]**. Guayaquil Ecuador: ACM, 2020. p. 1–12. DOI: 10.1145/3392561.3394637.

PEREIRA, Marisley. **Perspectivas dos Instrumentos de Accountability sobre a Supervisão da Atividade Parlamentar**. 2008. Universidade do Legislativo Brasileiro, [S. l.], 2008.

PERLATTO, Fernando. Habermas, a esfera pública e o Brasil. **Revista Estudos Políticos**, [S. l.], v. 3, n. 5, p. 78–94, 2012.

PINTO, Júlio Roberto de Souza. Processo legislativo no estado democrático de direito. **Revista de informação legislativa**, [S. l.], v. 42, n. 166, p. 193–203, 2005.

PORTAL DA CÂMARA DOS DEPUTADOS. **Número de deputados negros cresce quase 5%**. 2018a. Disponível em: <https://www.camara.leg.br/noticias/545913-numero-de-deputados-negros-cresce-quase-5/>. Acesso em: 23 jul. 2021.

PORTAL DA CÂMARA DOS DEPUTADOS. **Câmara tem 243 deputados novos e renovação de 47,3%**. 2018b. Disponível em: <https://www.camara.leg.br/noticias/545896-camara-tem-243-deputados-novos-e-renovacao-de-473/>. Acesso em: 23 jul. 2021.

PORTAL DA CÂMARA DOS DEPUTADOS. **Nova composição da Câmara ainda tem descompasso em relação ao perfil da população brasileira**. 2019. Disponível em: <https://www.camara.leg.br/noticias/550900-nova-composicao-da->

camara-ainda-tem-descompasso-em-relacao-ao-perfil-da-populacao-brasileira/. Acesso em: 23 jul. 2021.

PORTAL DA CÂMARA DOS DEPUTADOS. **Deputados**. 2021a. Disponível em: [https://www2.camara.leg.br/transparencia/aceso-a-informacao/copy\\_of\\_perguntas-frequentes/deputados](https://www2.camara.leg.br/transparencia/aceso-a-informacao/copy_of_perguntas-frequentes/deputados). Acesso em: 23 jul. 2021.

PORTAL DA CÂMARA DOS DEPUTADOS. **Suplentes em Exercício**. 2021b. Disponível em: <https://www.camara.leg.br/deputados/suplentes-em-exercicio>. Acesso em: 23 jul. 2021.

POSEGGA, Oliver; JUNGHER, Andreas. Characterizing Political Talk on Twitter: A Comparison Between Public Agenda, Media Agendas, and the Twitter Agenda with Regard to Topics and Dynamics. *In: PROCEEDINGS OF THE 52ND HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES 2019, Anais [...]*. [s.l.: s.n.] p. 10. DOI: 10.24251/hicss.2019.312.

RAY, Argha; GEORGE, Joey. Online Disinformation and the Psychological Bases of Prejudice and Political Conservatism. *In: PROCEEDINGS OF THE 52ND HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES 2019, Anais [...]*. [s.l.: s.n.] p. 11. DOI: 10.24251/hicss.2019.330.

RECUERO, Raquel; BASTOS, Marco; ZAGO, Gabriela. **Análise de redes para mídia social**. [s.l.] : Editora Sulina, 2015.

RÜDIGER, Francisco. **Martin Heidegger e a questão da técnica: prospectos acerca do futuro do homem**. Porto Alegre: Sulina, 2006.

RÜDIGER, Francisco. **As Teorias da Comunicação**. Porto Alegre: Penso Editora, 2009.

SAID, Edward W. **Orientalismo: o Oriente como invenção do Ocidente**. [s.l.] : Editora Companhia das Letras, 2007.

SANTOS, Milton; SILVEIRA, Maria Laura. **O Brasil: território e sociedade no início do século XXI**. 19. ed. Rio de Janeiro.

SARDINHA, Edison; COSTA, Sylvio. **Direita cresce e engole o centro no Congresso mais fragmentado da história**. 2019. Disponível em: <https://congressoemfoco.uol.com.br/legislativo/direita-cresce-e-engole-o-centro-no-congresso-mais-fragmentado-da-historia/>. Acesso em: 11 mar. 2021.

SCHWARCZ, Lilia Moritz; STARLING, Heloisa Maria Murgel. **Brasil: uma biografia**. São Paulo, Brazil: Companhia das Letras, 2015.

SIEBENEICHLER, Flavio Beno. O Conceito Esfera Pública No Pensamento Habermasiano. **Logeion: Filosofia da Informação**, [S. l.], v. 5, p. 84–96, 2018. a. DOI: 10.21728/logeion.2018v5n0.p84-96.

SIEBENEICHLER, Flavio Beno. Mundo Da Vida E Sistema Na Teoria Do Agir Comunicativo. **Logeion: Filosofia da Informação**, [S. l.], v. 5, p. 27–36, 2018. b. DOI: 10.21728/logeion.2018v5n0.p27-36.

SIEBENEICHLER, Flavio Beno. Jürgen Habermas: Uma Teoria Da Comunicação Humana. **Logeion: Filosofia da Informação**, [S. l.], v. 5, p. 8–26, 2018. c. DOI: 10.21728/logcion.2018v5n0.p8-26.

SILVA, Tarcízio Roberto Da. Aplicativos de análise de informações sociais: mapeamento e dinâmicas interacionais. [S. l.], 2012. a.

SILVA, Talita Victor. **O Uso Do Twitter Pelos Deputados Federais Brasileiros: Estudo Sobre Atuação E Tendências De Comportamento**. 2012b. Centro de Formação, Treinamento e Aperfeiçoamento da Câmara dos Deputados (Cefor), [S. l.], 2012.

SILVA, Tarcízio; STABILE, Max. Monitoramento e pesquisa em mídias sociais: metodologias, aplicações e inovações. **Análise de redes em Mídias Sociais. São Paulo: Uva Limão**, [S. l.], p. 237–260, 2016.

SMITH, Marc A. Conectando o poder das redes sociais. **RECUERO, Raquel; BASTOS, Marco; ZAGO, Gabriela. Análise de redes para mídia social. Porto Alegre: Sulinas**, [S. l.], p. 9–19, 2015.

SUL 21. **Site do Gabinete Digital é tirado do ar pelo governo do Estado**. 2015. Disponível em: <https://www.sul21.com.br/ultimas-noticias/geral/2015/08/site-do-gabinete-digital-e-tirado-do-ar-pelo-governo-do-estado/>. Acesso em: 13 nov. 2020.

TARNOFF, Ben; WEIGEL, Moira. **Why Silicon Valley can't fix itself**. 2018. Disponível em: <http://www.theguardian.com/news/2018/may/03/why-silicon-valley-cant-fix-itself-tech-humanism>. Acesso em: 19 jul. 2021.

VIEIRA, Fernando Sabóia. **Regras, instituições e decisões na Câmara dos Deputados do Brasil**. 1a. edição ed. Curitiba, PR: Appris Editora, 2019.

VILLA, Dana R. Postmodernism and the Public Sphere. **American Political Science Review**, [S. l.], v. 86, n. 3, p. 712–721, 1992. DOI: 10.2307/1964133.

WELLS, Chris; SHAH, Dhavan; LUKITO, Josephine; PELLED, Ayellet; PEVEHOUSE, Jon CW; YANG, JungHwan. Trump, Twitter, and news media responsiveness: A media systems approach. **New Media & Society**, [S. l.], v. 22, n. 4, p. 659–682, 2020.

WU, Vinicius Gomes. **A comunicação em experiências de e-participação: O caso da plataforma dialoga brasil**. 2018. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil, 2018. DOI: 10.17771/PUCRio.acad.36532. Disponível em: [http://www.maxwell.vrac.puc-rio.br/Busca\\_etds.php?strSecao=resultado&nrSeq=36532@1](http://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=36532@1). Acesso em: 13 out. 2021.

WYLIE, Christopher. 2018. Disponível em: <https://www.judiciary.senate.gov/download/05-16-18-wylie-testimony>. Acesso em: 19 jul. 2021.

YAQUB, Ussama; CHUN, Soon Ae; ATLURI, Vijayalakshmi; VAIDYA, Jaideep. Analysis of political discourse on twitter in the context of the 2016 US presidential

elections. *Government Information Quarterly*, [S. l.], v. 34, n. 4, p. 613–626, 2017. DOI: 10.1016/j.giq.2017.11.001.

## APÊNDICE A – Notebook de coleta e processamento de dados

Nesta dissertação, os *notebooks* voltados à coleta, processamento, análise e visualização de dados são considerados como apêndices – materiais elaborados pelo autor a fim de complementar sua argumentação.

```
import pandas as pd
import decimal
import numpy as np
from tabulate import tabulate
from fuzzywuzzy import process
import gspread
import tweepy
from tweepy import OAuthHandler
import time
from oauth2client.service_account import ServiceAccountCredentials
from df2gspread import df2gspread as d2g
import sys
import argparse
import json
import logging
import codecs
import os.path
from os.path import join as pjoin
import glob
import json
from io import StringIO
from io import open
from urllib.request import urlopen
from itertools import chain
import os
import csv
import re
from csv import DictReader
import itertools
import networkx as nx
from networkx.algorithms.community import girvan_newman
from networkx.algorithms.community import greedy_modularity_communities
from networkx.readwrite import json_graph
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

```
import sys
import os.path
import time
import argparse
import json
import tweepy
from tweepy import OAuthHandler
import logging
import codecs
from os.path import join as pjoin
```

## 0.4 - Declaração das chaves de API do Twitter

É preciso requisitá-las na [página de desenvolvedores da plataforma](#).

```
consumer_key = ''
consumer_secret = ''
access_token = ''
access_token_secret = ''

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)
```

## 0.5 - Download dos dados da Câmara dos Deputados e do Tribunal Superior Eleitoral

Câmara	TSE
deputados	candidatos2018
proposicoes	bensCandidatos2018
votacoesProposicoes	
proposicoesAutores	

## 0.6 - Significado dos sufixos

- Dataframe\_F: base proveniente de filtro
- Dataframe\_M: base proveniente de pareamento (merge)
- Dataframe\_G: base proveniente de agrupamento (grouby)

## 1 - Importação dos dados

### 1.1 - Carregamento de dados referentes à deputados, proposições, votações e frentes parlamentares

```
] : proposicoes = pd.read_csv("dados-camara/2020/originais/proposicoes-2020.csv", sep=";", encoding="utf-8")
proposicoesAutores = pd.read_csv("dados-camara/2020/originais/proposicoesAutores-2020.csv", sep=";", encoding="utf-8")
proposicoesTemas = pd.read_csv("dados-camara/2020/originais/proposicoesTemas-2020.csv", sep=";", encoding="utf-8")
votacoes = pd.read_csv("dados-camara/2020/originais/votacoes-2020.csv", sep=";", encoding="utf-8")
votacoesOrientacoes = pd.read_csv("dados-camara/2020/originais/votacoesOrientacoes-2020.csv", sep=";", encoding="utf-8")
votacoesProposicoes = pd.read_csv("dados-camara/2020/originais/votacoesProposicoes-2020.csv", sep=";", encoding="utf-8")
votacoesVotos = pd.read_csv("dados-camara/2020/originais/votacoesVotos-2020.csv", sep=";", encoding="utf-8")
frentesDeputados = pd.read_csv("dados-camara/2020/originais/frentesDeputados.csv", sep=";", encoding="utf-8")
deputados = pd.read_csv("dados-camara/2020/originais/deputados.csv", sep=";", encoding="utf-8")
```

### 1.2 - Listagem de colunas das bases

```
] : # proposicoes.columns
# list(proposicoesAutores.columns)
list(proposicoesTemas.columns)
# list(votacoes.columns)
# list(votacoesOrientacoes.columns)
# list(votacoesProposicoes.columns)
# list(votacoesVotos.columns)
# list(frentesDeputados.columns)
# list(deputados.columns)
```

### 1.3 - Remoção de colunas não usadas na análise

```
proposicoes_F1 = proposicoes.drop(['ultimoStatus_dataHora', 'ultimoStatus_uriRelator', 'ultimoStatus_idOrgao', 'ultimoStatus_siglaOrgao',
proposicoesAutores_F1 = proposicoesAutores.drop(['proponente', 'uriAutor', 'uriPartidoAutor', 'uriProposicao', 'ordemAssinatura'], axis
proposicoesTemas_F1 = proposicoesTemas2020.drop(['siglaTipo', 'numero', 'ano', 'codTema', 'relevancia'], axis = 1)
votacoes_F1 = votacoes.drop(['dataHoraRegistro', 'idOrgao', 'uriOrgao', 'siglaOrgao', 'idEvento', 'uriEvento', 'ultimaAberturaVotacao_data
votacoesOrientacoes_F1 = votacoesOrientacoes2020.drop(['uriVotacao', 'siglaOrgao'], axis = 1)
votacoesProposicoes_F1 = votacoesProposicoes2020.drop(['uriVotacao', 'proposicao_codTipo'], axis = 1)
votacoesVotos_F1 = votacoesVotos.drop(['uriVotacao', 'deputado_uri', 'deputado_uriPartido', 'deputado_urlFoto', 'dataHoraVoto'], axis =
frentesDeputados_F1 = frentesDeputados.drop(['deputado_uri', 'deputado_urlFoto', 'deputado_codTitulo', 'dataInicio', 'dataFim', 'depu
```

### 1.4 - Arquivamento de bases filtradas

```
proposicoes_F1.to_csv("dados-camara/2020/modificados/proposicoes-2020-F1.csv", sep=";", encoding="utf-8", index=False)
proposicoesAutores_F1.to_csv("dados-camara/2020/modificados/proposicoesAutores-2020-F1.csv", sep=";", encoding="utf-8", index=False)
proposicoesTemas_F1.to_csv("dados-camara/2020/modificados/proposicoesTemas-2020-F1.csv", sep=";", encoding="utf-8", index=False)
votacoes_F1.to_csv("dados-camara/2020/modificados/votacoes-2020-F1.csv", sep=";", encoding="utf-8", index=False)
votacoesOrientacoes_F1.to_csv("dados-camara/2020/modificados/votacoesOrientacoes-2020-F1.csv", sep=";", encoding="utf-8", index=False)
votacoesProposicoes_F1.to_csv("dados-camara/2020/modificados/votacoesProposicoes-2020-F1.csv", sep=";", encoding="utf-8", index=False)
votacoesVotos_F1.to_csv("dados-camara/2020/modificados/votacoesVotos-2020-F1.csv", sep=";", encoding="utf-8", index=False)
frentesDeputados_F1.to_csv("dados-camara/2020/modificados/frentesDeputados-2020-F1.csv", sep=";", encoding="utf-8", index=False)
```

### 1.5 - Carregamento das bases filtradas

```
proposicoes_F1 = pd.read_csv("dados-camara/2020/modificados/proposicoes-2020-F1.csv", sep=";", encoding="utf-8")
proposicoesAutores_F1 = pd.read_csv("dados-camara/2020/modificados/proposicoesAutores-2020-F1.csv", sep=";", encoding="utf-8")
proposicoesTemas_F1 = pd.read_csv("dados-camara/2020/modificados/proposicoesTemas-2020-F1.csv", sep=";", encoding="utf-8")
votacoes_F1 = pd.read_csv("dados-camara/2020/modificados/votacoes-2020-F1.csv", sep=";", encoding="utf-8")
votacoesOrientacoes_F1 = pd.read_csv("dados-camara/2020/modificados/votacoesOrientacoes-2020-F1.csv", sep=";", encoding="utf-8")
votacoesProposicoes_F1 = pd.read_csv("dados-camara/2020/modificados/votacoesProposicoes-2020-F1.csv", sep=";", encoding="utf-8")
votacoesVotos_F1 = pd.read_csv("dados-camara/2020/modificados/votacoesVotos-2020-F1.csv", sep=";", encoding="utf-8")
frentesDeputados_F1 = pd.read_csv("dados-camara/2020/modificados/frentesDeputados-2020-F1.csv", sep=";", encoding="utf-8")
deputados = pd.read_csv("dados-camara/2020/originais/deputados.csv", sep=";", encoding="utf-8")
```

### 1.6 - Listagem de colunas das bases filtradas

```
list(proposicoes2020_F1.columns)
list(proposicoesAutores2020_F1.columns)
list(proposicoesTemas2020_F1.columns)
list(votacoes2020_F1.columns)
list(votacoesOrientacoes2020_F1.columns)
list(votacoesProposicoes2020_F1.columns)
list(votacoesVotos2020_F1.columns)
list(frentesDeputados_F1.columns)
list(deputados.columns)
```

## 2 - Processamento da base de dados "listaDeputados"

### 2.1 - Filtragem da base do TSE para pegar apenas deputados federais

```
votacao = pd.read_csv("votacao_candidato_munzona_2018_BRASIL.csv", sep=";", encoding="iso-8859-1")
filtroVotacao = votacao[(votacao.DS_CARGO=='Deputado Federal') & (votacao.DS_SIT_TOT_TURNO.str.startswith("ELEITO"))]
filtroVotacao.to_csv("deputados-2018.csv", sep=";", encoding="utf-8")
```

### 2.2 - Agrupamento de informações do TSE para obter registros únicos e números totais de votos

```
votacaoFiltrado = pd.read_csv("deputados-2018.csv", sep=";", encoding="utf-8")
filtroVotacao2 = votacaoFiltrado.groupby(['NM_CANDIDATO']).agg({'NM_URNA_CANDIDATO': 'first', 'SG_PARTIDO': 'first', 'DS_COMPOSICAO_CO
filtroVotacao2.to_csv("2deputados-2018.csv", sep=";", encoding="utf-8")
```

### 2.3 - Filtragem da base do TSE para pegar apenas deputados federais suplentes

```
votacao = pd.read_csv("votacao_candidato_munzona_2018_BRASIL.csv", sep=";", encoding="iso-8859-1")
filtroVotacao = votacao[(votacao.DS_CARGO=="Deputado Federal") & (votacao.DS_SIT_TOT_TURNO.str.startswith("SUPLENTE"))]
filtroVotacao2 = filtroVotacao.groupby(['NM_CANDIDATO']).agg({'NM_URNA_CANDIDATO':'first', 'SG_PARTIDO':'first', 'DS_COMPOSICAO_COLI': 'first'})
filtroVotacao2.to_csv("suplentes-2018.csv", sep=";", encoding="utf-8")
```

### 2.4 - Exportação de informações sobre suplentes que assumiram titularidade

```
efetivados = pd.read_csv("suplentes-tse-2018.csv", sep=";", encoding="utf-8")
filtroEfetivados = efetivados[(efetivados.NM_CANDIDATO.isin(["AELTON JOSÉ DE FREITAS", "BEATRIZ ROSALIA RIBEIRO CAVASSA DE OLIVEIRA"])]
filtroEfetivados.to_csv("efetivados.csv", encoding="utf-8")
```

### 2.6 - Pareamento entre base de dados e lista da Auditoria Cidadã com função da biblioteca fuzzywuzzy

```
tse = pd.read_csv("deputados-eleitos-tse.csv", sep=";", encoding="utf-8")
auditoria = pd.read_csv("perfis-auditoria-cidada.csv", sep=";", encoding="utf-8")
deputados = pd.read_csv("deputados-camara.csv", sep=";", encoding="utf-8")

tse = tse.sort_values('Nome_Parlamentar')
auditoria = auditoria.sort_values('Nome_Parlamentar')

choices = auditoria['Nome_Parlamentar'].values.tolist()
tse['xNome_Parlamentar'] = tse['Nome_Parlamentar'].apply(lambda x: process.extractOne(x, choices)[0])

tabela1 = tse.merge(auditoria, left_on='xNome_Parlamentar', right_on='Nome_Parlamentar')
tabela2 = pd.merge(tabela1, deputados, how='left', left_on=['NM_CANDIDATO'], right_on=['nomeCivil'])
tabela2.to_csv("pareamento-perfis-5.csv", sep=";", encoding="utf-8")
```

### 2.7 - Busca de perfis oficiais no Twitter

```
count = 5

fileOut = open('results.txt', "w", encoding='utf-8')

with open('deputados.csv', "r", encoding='utf-8') as fileIn:

    nomes = fileIn.readlines()
    for n in nomes:
        # fileOut.write("GoL---"+n)
        usersI = api.search_users(n, count)
        usersM = [user.screen_name for user in usersI]
        print(str(usersM))
        print(str(usersM), file=fileOut)
        time.sleep(5)
```

### 2.8 - Coleta do número de seguidores e de publicações dos perfis oficiais

```
stats=open("results.txt", "w", encoding='utf-8', newline='')

with open('users.txt', "r", encoding='utf-8', newline='') as fileIn:
    users = fileIn.readlines()
    for user_name in users:
        user = api.get_user(screen_name = user_name)
        bio = str(user.description)
        bio = re.sub(r"\n|\r|\t|\n\r|\n\r\n", " ", bio)
        bio = re.sub(";", ",", bio)
        print(user.screen_name, ';', user.id, ';', user.followers_count, ';', user.statuses_count, ';', bio, file=stats)
        print(user.screen_name)
        time.sleep(5)
```

## 2.9 - Pareamento de dados da Câmara e do TSE com os do Twitter

```
tabela1 = pd.read_csv("db-deputados-1.csv", sep=";", encoding="utf-8")
tse2018 = pd.read_csv("tse-2018.csv", sep=";", encoding="utf-8")
tabela2 = pd.merge(tabela1, tse2018, how='left', left_on=['CPF'], right_on=['NR_CPF_CANDIDATO'])
tabela2.to_csv("db-deputados-2.csv")
```

## 2.10 - Agrupamento e soma de valores referentes a declaração de bens dos deputados em 2018.

Centavos precisaram ser arredondados para completar esta função

```
bensTse2018 = pd.read_csv("tse-bens-2018-1.csv", sep=";", encoding="utf-8")
# bensTse2018.shape[0]
# List(bensTse2018.columns)

bensTse2018['VR_BEM_CANDIDATO'] = bensTse2018['VR_BEM_CANDIDATO'].round(0).astype(int)

somaBens2018 = bensTse2018.groupby(['SQ_CANDIDATO']).agg({'VR_BEM_CANDIDATO': 'sum'})
somaBens2018.to_csv("soma-bens-2018.csv")
```

## 2.11 - Pareamento da base de dados com as declarações de bens do TSE

```
tabela1 = pd.read_csv("db-deputados-24-1.csv", sep=";", encoding="utf-8")
somaBens2018 = pd.read_csv("soma-bens-2018.csv", sep=";", encoding="utf-8")

tabela2 = pd.merge(tabela1, somaBens2018, how='left', left_on=['SQ_CANDIDATO'], right_on=['SQ_CANDIDATO'])

tabela2.to_csv("db-deputados-25.csv")
```

## 2.12 - Pareamento da base de dados com lista de ideologia dos partidos

```
tabela1 = pd.read_csv("db-deputados-25.csv", sep=";", encoding="utf-8")
ideologiaPartidos = pd.read_csv("ideologia-partidos.csv", sep=";", encoding="utf-8")

tabela2 = pd.merge(tabela1.astype("str"), ideologiaPartidos.astype("str"), how='left', left_on=['PARTIDO_ATUAL'], right_on=['PARTIDO'])

tabela2.to_csv("db-deputados-26.csv")
```

## 2.13 - Pareamento da base de dados com lista da divisão territorial do Brasil

```
tabela1 = pd.read_csv("db-deputados-26.csv", sep=";", encoding="utf-8")
regioes = pd.read_csv("regioes.csv", sep=";", encoding="utf-8")
tabela2 = pd.merge(tabela1, regioes, how='left', left_on=['ESTADO_ELEICAO'], right_on=['ESTADO'])
tabela2.to_csv("db-deputados-27.csv")
```

# 3 - Processamento da base de dados "proposicoes"

## 3.1 - Pareamento e filtragem das bases de proposições e autores de proposições

```
proposicoes_M = proposicoes_F1.merge(proposicoesAutores2020_F1, how='left', left_on='id', right_on='idProposicao').merge(proposicoesT)
proposicoes_MF = proposicoes_M.drop(['Unnamed: 0_x', 'Unnamed: 0_y', 'idProposicao_y', 'siglaTipo_y', 'numero_y', 'ano_y', 'Unnamed: 0',
```

### 3.2 - Agrupamento de informações referentes às proposições

```
proposicoesM.shape[0]
unico = proposicoesM['id'].unique()
unico.shape[0]
proposicoes_GMF = proposicoes_MF.astype('str').groupby('id').agg({'uri': 'first', 'siglaTipo_x': 'first', 'numero_x': 'first', 'ano_x':
# proposicoesG.shape[0]
proposicoes_GMF.to_csv("dados-camara/2020/modificados/proposicoes-G.csv", sep=";", encoding="utf-8")
```

```
proposicoes = pd.read_csv("db-proposicoes.csv")
proposicoes['NOME_AUTOR'] = proposicoes['NOME_AUTOR'].astype("str").apply(lambda x: ', '.join(set(x.split(','))))
proposicoes.to_csv("db-proposicoes.csv", sep=";", index=False)
```

## 4 - Processamento das bases de dados "votosDeputados" e "votacoes"

### 4.1 - Agrupamento das bases de orientações e votos a partir de coluna concatenada

```
votacoesOrientacoes_F1['idVotacaoPartido'] = votacoesOrientacoes_F1['idVotacao'] + votacoesOrientacoes_F1['siglaBancada']
# votacoesOrientacoes['idVotacaoPartido'].unique().shape[0]
votacoesVotos_F1['idVotacaoPartido'] = votacoesVotos_F1['idVotacao'] + votacoesVotos_F1['deputado_siglaPartido']
# votacoesVotos['idVotacaoPartido'].unique().shape[0]
votacoesVotos_F1['idProposicao'] = votacoesVotos_F1['idVotacao'].astype('str').replace("\-(.*)", "", regex=True)
votacoesOrientacoes_F1['idProposicao'] = votacoesOrientacoes_F1['idVotacao'].astype('str').replace("\-(.*)", "", regex=True)
votacoesProposicoes_F1['idProposicao'] = votacoesProposicoes_F1['idVotacao'].astype('str').replace("\-(.*)", "", regex=True)
votosOrientacoes_M = votacoesVotos_F1.merge(votacoesOrientacoes_F1, how='left', on='idVotacaoPartido')
# votosOrientacoes_MF = votosOrientacoes_M.drop(['Unnamed: 0_x', 'Unnamed: 0_y', 'idVotacao_y', 'uriVotacao_y', 'deputado_siglaPartido'])
print(votosOrientacoes_M.shape[0], votosOrientacoes_M['idProposicao_x'].unique().shape[0])

votacoesProposicoes_M = votacoes_F1.merge(votacoesProposicoes_F1, how='left', left_on='id', right_on='idVotacao')
votacoes_MF = votacoesProposicoes_M.drop(['Unnamed: 0_x', 'Unnamed: 0_y', 'data_y', 'descricao_y', 'dataHoraRegistro', 'idOrgao', 'idVotacao_y', 'uriVotacao_y', 'deputado_siglaPartido'])
```

### 4.2 - Pareamento da base "votosDeputados" com resultadosVotacoes" e deputados

```
v1resultadoVotacoes = pd.read_csv("planilha-google/v0-resultadoVotacoes.csv", sep=";", encoding="utf-8")
v0proposicoes = pd.read_csv("planilha-google/v0-proposicoes.csv", sep=";", encoding="utf-8")
v0deputados = pd.read_csv("planilha-google/v0-deputados.csv", sep=";", encoding="utf-8")
v0votosDeputados = pd.read_csv("planilha-google/v0-0-votosDeputados.csv", sep=";", encoding="utf-8")

v1votosDeputados = v0votosDeputados.merge(v1resultadoVotacoes, how='left', on='ID_VOTACAO')
v1votosDeputados = v1votosDeputados[v1votosDeputados['APROVACAO'].notna()]
v2votosDeputados = v1votosDeputados.astype('str').merge(v0deputados.astype('str'), how='left', left_on='ID_DEPUTADO', right_on='ID_CA')
v2votosDeputados['ID_PROPOSICAO'] = v2votosDeputados['ID_PROPOSICAO'].astype('str').str.replace('.0', '')

v3votosDeputados = pd.merge(v2votosDeputados.astype('str'), v0proposicoes[['ID', 'TEMA', 'PALAVRAS_CHAVE', 'EMENTA']].astype('str'), how='left', on='ID_PROPOSICAO')

v4votosDeputados = v3votosDeputados.drop(['EMENTA_PROPOSICAO', 'VOTOS_SIM', 'VOTOS_NAO', 'VOTOS_OUTROS', 'DESCRICAO', 'TIPO_PROPOSICAO', 'ID_DEPUTADO', 'ID_VOTACAO', 'URI_VOTACAO', 'DEPUTADO_SIGLA_PARTIDO'])

v5votosDeputados = v4votosDeputados[v4votosDeputados['TEMA'].notna()]
# list(v4votosDeputados.columns)
v5votosDeputados.to_csv("planilha-google/v0-4-votosDeputados.csv", sep=";", encoding="utf-8")
print(v4votosDeputados.shape[0], v4votosDeputados['ID_PROPOSICAO'].unique().shape[0])
```

```
votosDeputados = pd.read_csv("db-votosDeputados.csv")
votosDeputados = votosDeputados[votosDeputados['TEMA'].notna()]
votosDeputados.to_csv("db-votosDeputados.csv", index=False)
```

## 5 - Processamento da base de dados "tweetsDeputados"

### 5.1 - Coleta de publicações dos perfis oficiais

```

sys.path.append("../")

api = tweepy.API(auth, retry_count=5, retry_delay=5, timeout=150, wait_on_rate_limit=True, wait_on_rate_limit_notify=True)

def json_format(data_json, current_tweet):
    current_tweet['id'] = data_json.id
    current_tweet['created_at'] = data_json.created_at.strftime('%Y-%m-%dT%H:%M:%SZ')
    current_tweet['username'] = data_json.user.name
    try:
        #media = json.loads(data_json.entities['media'])
        current_tweet['media_type'] = data_json.entities['media'][0]['type']
        current_tweet['has_media'] = 'True'
        #print json.dumps(media)
    except:
        current_tweet['has_media'] = 'False'
        current_tweet['media_type'] = 'None'
    current_tweet['screen_name'] = data_json.user.screen_name
    current_tweet['retweet_count'] = data_json.retweet_count
    current_tweet['favorite_count'] = data_json.favorite_count
    current_tweet['full_text'] = data_json.text
    current_tweet['lang'] = data_json.lang
    return

def gather(api):
    try:
        user_id = api.get_user(perfil).id
    except tweepy.error.TweepError:
        print('User not found.')
        sys.exit(0)

```

```

if perfil:
    fname = perfil + '.json'
    dataset_file = open(fname, 'w', encoding='utf8')
else:
    fname = perfil + '.json'
    dataset_file = open(fname, 'w', encoding='utf8')

dataset_file.write(['')
current_tweet = {}
dados = api.home_timeline()
last_id = int(str(json.dumps(dados[0],_json))[55:74])
counter = 0

while True:
    try:
        new_tweets = api.user_timeline(id = user_id, count = 200, max_id=str(last_id-1))
        if not new_tweets:
            dataset_file.write('')
            break
        for tweet in new_tweets:
            if counter != 0:
                dataset_file.write(',')
            json_format(tweet, current_tweet)
            dataset_file.write(json.dumps(current_tweet, ensure_ascii=False))
            dataset_file.write('\n')
            counter+=1
            if sys.stdout.isatty():
                print("\rNumber of tweets collected so far...: %i"%counter, end='\r', flush=True)
            else:
                print(counter, end=' ', flush=True)

        last_id = new_tweets[-1].id

    except Exception as e:
        raise

```

```

print("\nAll done. Output written to " + fname)
dataset_file.close()
time.sleep(5.5)

with open('perfis-0.csv', "r",encoding='utf-8',newline='') as fileIn:
    lista_perfis = fileIn.read().split('\n')
    for perfil in lista_perfis:
        perfil = perfil.rstrip()
        gather(api)
        time.sleep(40)

```

## 5.2 - Processamento da primeira coleta de publicações oficiais

```

path_to_json = 'dados-twitter/publicacoes/A-29-12-2020\\'
json_pattern = os.path.join(path_to_json, '*.json')
file_list = glob.glob(json_pattern)

dfs = []
# print(chr(92))
for file in file_list:
    data = pd.read_json(file)
    dfs.append(data)

tweets_c1 = pd.concat(dfs, ignore_index=True)
tweets_c1['coleta'] = "C1"
# tweets_c1.head(5)
# tweets_c1.shape[0]
# tweets_c1['text'].unique().shape[0]
tweetsC1_F1 = tweets_c1[(tweets_c1['created_at'] > '2020-01-01 00:00:00')]
tweetsC1_F1.shape[0]
tweetsC1_F1.to_csv("dados-twitter/tweetsC1_F1.csv", sep=";", encoding="utf-8")

```

```

%%capture
arq = "dados-twitter/tweetsDeputados-F"
arq_csv = arq + ".csv"
arq_grafo = arq + "-grafo.csv"
arq_rt = arq + "-grafo-rt.csv"
arq_mt = arq + "-grafo-mt.csv"
arq_hashtag = arq + "-grafo-hashtags.csv"

outrow = []
with open(arq_csv, "r", encoding='utf-8') as fileIn: # input file location
    with open(arq_grafo, "w", encoding='utf-8') as fileOut: # output file location
        writer = csv.writer(fileOut)
        reader = csv.reader(fileIn, delimiter=';')
        writer.writerow(['id', 'created_at', 'media_type', 'retweet_count', 'favorite_count', 'lang', 'coleta', 'source', 'target', 'text'])
        for row in reader:
            idd = row[0]

```

```

            created_at = row[1]
            media_type = row[2]
            retweet_count = row[3]
            favorite_count = row[4]
            lang = row[5]
            coleta = row[6]
            screen_name = row[7]
            target = ''
            text = row[8]

            for cell in row:
                target = re.findall(r"@([\s]+)", cell)
                target = re.sub(r"\\[\s]+", '', str(target))
                outrow
            writer.writerow([idd, created_at, media_type, retweet_count, favorite_count, lang, coleta, screen_name, target, text])

df = pd.read_csv(arq_grafo)
# print(df.head(3))
df['target'].replace('', np.nan, inplace=True)
df.dropna(subset=['target'], inplace=True)
df.to_csv(arq_grafo, index=False)

df = pd.read_csv(arq_grafo)
rt = df['text'].str.startswith('RT @')

df[rt].to_csv(arq_rt, index=False)
df2 = pd.read_csv(arq_rt)
df2['target'].replace(r"[\s\S]*$", "", regex=True, inplace=True)
df2.to_csv(arq_rt, index=False)

def chainer(s):
    return list(chain.from_iterable(s.str.split(',')))

lens = df['target'].str.split(',').map(len)

```

```

res = pd.DataFrame({
    'source': np.repeat(df['source'], lens),
    'target': chainer(df['target']),
    'text': np.repeat(df['text'], lens),
    'id': np.repeat(df['id'], lens),
    'created_at': np.repeat(df['created_at'], lens),
    'media_type': np.repeat(df['media_type'], lens),
    'retweet_count': np.repeat(df['retweet_count'], lens),
    'favorite_count': np.repeat(df['favorite_count'], lens),
    'lang': np.repeat(df['lang'], lens),
    'coleta': np.repeat(df['coleta'], lens)
})

res['target'].replace('', np.nan, inplace=True)
res.dropna(subset=['target'], inplace=True)

mt = ~res['text'].str.startswith('RT @')
res[mt].to_csv(arq_mt, index=False)

outrow = []
with open(arq_csv, "r", encoding='utf-8') as fileIn: # input file location
    with open(arq_grafo, "w", encoding='utf-8') as fileOut: # output file location
        writer = csv.writer(fileOut)
        reader = csv.reader(fileIn, delimiter=',')
        writer.writerow(['hashtags'])
        for row in reader:
            # screen_name = row[0]
            hashtags = ''
            for cell in row:
                hashtags = re.findall(r"#([\s]+)", cell)
                hashtags = re.sub(r"[\s]|\:", '', str(hashtags))
                # hashtags = re.sub('[+string.punctuation+]', '', hashtags)
                hashtags = re.sub(r"^\P{P},", "", hashtags)
                hashtags = re.sub(r"^\s|\s|\s", "", hashtags)
            outrow

```

```

writer.writerow([hashtags])

dfrw = pd.read_csv(arq_grafo)
dfrw['hashtags'].replace('', np.nan, inplace=True)
dfrw.dropna(subset=['hashtags'], inplace=True)
dfrw.to_csv(arq_grafo, index=False)

with open(arq_grafo, encoding="utf-8") as f:
    vTweets = [row["hashtags"] for row in DictReader(f)]

col1=[]
col2=[]

for tweet in vTweets:
    # print(tweet)
    vPalavras = tweet.split(',')
    nPalavras = len(vPalavras)

    if nPalavras > 2:
        nLinhas = 0
        vNumeros = []
        for i in range(nPalavras):
            b = nPalavras-(i+1)
            if b>0:
                vNumeros.append(b)
            nLinhas = nLinhas + b

        nNumeros = len(vNumeros)
        invNumeros = vNumeros[::-1]

        c1 = []
        for i in range(nNumeros):
            for j in range(vNumeros[i]):
                c1.append(vPalavras[i])
            col1.extend(c1)

```

```

ordemC2 = []
for i in range(nNumeros):
    for j in invNumeros:
        ordemC2.append(j)
    invNumeros.pop(0)

Mposicao = []
posicaoC2 = []
for i in range(int(nLinhas/nPalavras)):
    for j in range(nPalavras):
        posicaoC2.append(vPalavras[j])
        Mposicao.append(j)

c2 = []
c2[:] = [posicaoC2[i] for i in ordemC2]
col2.extend(c2)

# usr = pd.DataFrame({'source': np.repeat(df['source'], lens),
# usr = dfRW['usuario'].values[0]
# n = len(col1)
# usrs = [usr]*n

matriz = np.c_[col1,col2]
df = pd.DataFrame(columns=["source","target"], data=matriz)
df['source'].replace('', np.nan, inplace=True)
df.dropna(subset=['source'], inplace=True)
df['target'].replace('', np.nan, inplace=True)
df.dropna(subset=['target'], inplace=True)

df.to_csv(arq_hashtag, index=False)
os.remove(arq_grafo)

```

```

tweetsDeputados = pd.read_csv("dados-twitter/tweetsDeputados-F.csv", sep=";", encoding="utf-8")
tweetsDeputados_T100 = tweetsDeputados.nlargest(100,"retweet_count")
tweetsDeputados_T100S = tweetsDeputados_T100['text']
tweetsDeputados_T100S.head(4)
# tweetsDeputados_T100.shape[0]
tweetsDeputados_T100S.to_csv("dados-twitter/tweetsDeputados-T100S.csv", sep=";", encoding="utf-8", index=False)

```

## APÊNDICE B – Notebook de análise e visualização de dados

### 0 - Setup do Sistema

#### 0.1 - Importa os módulos utilizados

```
import collections
import kaleido
import matplotlib
from unicode import unicode
import pandas as pd
import string
import decimal
import requests
import numpy as np
from tabulate import tabulate
from fuzzywuzzy import process
import gspread
import tweepy
from tweepy import OAuthHandler
import time
from oauth2client.service_account import ServiceAccountCredentials
from df2gspread import df2gspread as d2g
import sys
import argparse
```

```
import json
import logging
import codecs
import os.path
from os.path import join as pjoin
import glob
import json
from io import StringIO
from io import open
from urllib.request import urlopen
from itertools import chain
import os
import csv
import re
from collections import Counter
# import re as regex
# import regex as re
# import re
from tabulate import tabulate
from csv import DictReader
import itertools
import networkx as nx
from networkx.algorithms.community import girvan_newman
from networkx.algorithms.community import greedy_modularity_communities
from networkx.readwrite import json_graph
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
from collections import Counter
from PIL import Image
import matplotlib.pyplot as plt
import seaborn as sns
import chart_studio.plotly as chs
import cufflinks as cf
import plotly.express as px
from wordcloud import WordCloud, ImageColorGenerator
```

```

import time
from requests.exceptions import HTTPError
from urllib.request import urlopen
import urllib
import json

import ijson
from datetime import datetime
import glob
# import matplotlib.pyplot as plt, mpld3
from dateutil.parser import parse

%matplotlib inline

# Make Plotly work in your Jupyter Notebook
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
init_notebook_mode(connected=True)
# Use Plotly locally
cf.go_offline()

```

## 0.2 - Carrega a base de dados

```

listaDeputados = pd.read_csv("../1-analise-visualizacao/db-listaDeputados.csv")
print(listaDeputados.columns)
# pd.set_option('max_columns', None)
# listaDeputados.head(3)

```

## 0.3 - Formata rótulos das base de dados para as tabelas

```

keepCol = ['GENERO', 'COR_RACA', 'ESTADO_CIVIL', 'ESCOLARIDADE', 'OCUPACAO_ANTERIOR', 'FAIXA_ETARIA', 'PARTIDO_ATUAL',
listaDeputados_F = listaDeputados[keepCol]

listaDeputados_F = listaDeputados_F.rename(columns={'GENERO': 'Gênero', 'COR_RACA': 'Cor/Raça', 'ESTADO_CIVIL': 'Estado

```

# 1 - Análise exploratória

## 1.05 - Cria Faixas de dados

```

faixa_votos = ['VOTOS_G4', 'VOTOS_G3', 'VOTOS_G2', 'VOTOS_G1']
faixa_bens = ['BENS_G4', 'BENS_G3', 'BENS_G2', 'BENS_G1']
faixa_seguidores = ['SEG_G4', 'SEG_G3', 'SEG_G2', 'SEG_G1']
faixa_bseguidores = ['B_SEG_G4', 'B_SEG_G3', 'B_SEG_G2', 'B_SEG_G1']
faixa_tweets = ['NTWEETS_G4', 'NTWEETS_G3', 'NTWEETS_G2', 'NTWEETS_G1']
faixa_btweets = ['B_TWEETS_G4', 'B_TWEETS_G3', 'B_TWEETS_G2', 'B_TWEETS_G1']

listaDeputados['AGRUPAMENTO_VOTOS'] = pd.qcut(listaDeputados['QT_VOTOS'],4,precision=0,labels=faixa_votos)
listaDeputados['AGRUPAMENTO_BENS'] = pd.qcut(listaDeputados['VALOR_BENS'],4,precision=0,labels=faixa_bens)
listaDeputados['AGRUPAMENTO_SEGUIDORES'] = pd.qcut(listaDeputados['SEGUIDORES_E2'],4,precision=0,labels=faixa_seguidores)
listaDeputados['AGRUPAMENTO_BALANCO_SEGUIDORES'] = pd.qcut(listaDeputados['BALANÇO_SEGUIDORES'],4,precision=0,labels=faixa_bsegu
listaDeputados['AGRUPAMENTO_Número_TWEETS'] = pd.qcut(listaDeputados['TWEETS_E2'],4,precision=0,labels=faixa_tweets)
listaDeputados['AGRUPAMENTO_BALANCO_TWEETS'] = pd.qcut(listaDeputados['BALANÇO_TWEETS'],4,precision=0,labels=faixa_btweets)

listaDeputados.to_csv("db-listaDeputados.csv", index=False)

```

## 1.1 - Levanta coluna generica para as funções de agrupamento

```
listaDeputados_F['new'] = 'a'
```

## 1.2 - Filtra a base de dados por deputados gaúchos

```
listaDeputados_rs = listaDeputados_F[(listaDeputados_F['Estado']=='RS')]
```

## 1.3 - Cria tabela de características gerais

```
numero = listaDeputados.value_counts()
percentual = round(listaDeputados.value_counts(normalize=True)*100,2)
matriz = {'Número':numero,'Percentual':percentual}
stats_geral = pd.DataFrame(data=matriz)
moda_geral = listaDeputados_F.groupby('new').agg(lambda x: x.mode())
info_geral = listaDeputados_F.groupby('new').agg(lambda x: x.value_counts().index[0] + f"\n({x.value_counts().iloc[0]} || {round(x,2)}")
# moda_geral.transpose()
# info_geral.transpose().to_clipboard()
info_geral.transpose()
```

## 1.4 - Cria a tabela de características dos deputados gaúchos

```
numero = listaDeputados_rs.value_counts()
percentual = round(listaDeputados_rs.value_counts(normalize=True)*100,2)
matriz = {'Número':numero,'Percentual':percentual}
stats_rs = pd.DataFrame(data=matriz)
moda_rs = listaDeputados_rs.groupby('new').agg(lambda x: x.mode())

info_rs = listaDeputados_rs.groupby('new').agg(lambda x: x.value_counts().index[0] + f"\n({x.value_counts().iloc[0]} || {round(x,2)}")
# moda_rs.transpose()
# info_rs.transpose().to_clipboard()
info_rs.transpose()
```

## 1.5 - Gera gráfico de deputados por estado

```
deputados_estados = listaDeputados['ESTADO_ELEICAO'].value_counts().sort_values(ascending=True)

# deputados_estados = pd.DataFrame(list(zip(titulos,duracao)), index=titulos).sort_values(0, ascending=False)

top = deputados_estados.plot(kind='barh',
                             colormap='tab10',
                             figsize=(11, 8),width=0.8)

for i, v in enumerate(list(deputados_estados)):
    top.text(v+0.4, i-0.25, v, fontsize=12, color='black')

plt.tick_params(axis='y', which='major', labelsize=14)

# top.margins(x=1)
# top.spines['right'].set_visible(False)
# top.spines['top'].set_visible(False)
# top.spines['bottom'].set_visible(False)

# top.get_legend().remove()

# plt.rcParams.update(IPython_default);
# plt.style.use("fivethirtyeight")
plt.grid(color='b', linestyle='-', linewidth=1)
plt.savefig("novos-graficos/Figura x - Quantidade de Deputados Federais por Estado.jpg",bbox_inches = 'tight')
```

## 2 - Consistência Temática

### 2.1 - Carrega bases de dados

```
]: # pd.set_option('max_columns', None)
listaDeputados = pd.read_csv("db-listaDeputados.csv")
proposicoes = pd.read_csv("db-proposicoes.csv")
tweetsDeputados = pd.read_csv("tweetsDeputados.csv", sep=";")

# tweetsDeputados_100 = pd.read_csv("tweetsDeputados-T100.csv", sep=";")
# grafo_mt = pd.read_csv("grafo-mt.csv")
# grafo_rt = pd.read_csv("grafo-rt.csv")
# grafo_hashtags = pd.read_csv("grafo-hashtags.csv")
# grafo_semantico = pd.read_csv("grafo-semantico-T100.csv")

# print(listaDeputados.columns)
# listaDeputados.head(3)
# list(proposicoes.columns)
# # proposicoes.head(3)
```

### 2.2 - Cria base de dados da autoria das proposições

```
%%capture
PLs = pd.DataFrame(proposicoes[proposicoes['DESC_PROPOSICAO']=='Projeto de Lei'][proposicoes["TIPO_AUTOR"]=="Deputado"])
# proposicoes.shape[0]
# PLs.shape[0]
# List(PLs.columns)
# PLs.head(3)

def chainer(s):
    return list(chain.from_iterable(s.str.split(',')))

lens = PLs['NOME_AUTOR'].astype('str').str.split(',').map(len)
# PLs.shape[0]
# lens
autores = pd.DataFrame({
    'NOME_AUTOR': chainer(PLs['NOME_AUTOR'].astype("str")),
    'ID_AUTOR': np.repeat(PLs['ID_AUTOR'], lens),
    'TITULO_PROPOSICAO': np.repeat(PLs['TITULO_PROPOSICAO'], lens),
    'TEMA': np.repeat(PLs['TEMA'], lens),
    'PALAVRAS_CHAVE': np.repeat(PLs['PALAVRAS_CHAVE'], lens),
    'EMENTA': np.repeat(PLs['EMENTA'], lens),
    'DATA_APRESENTACAO': np.repeat(PLs['DATA_APRESENTACAO'], lens),
    'DESC_TRAMITACAO': np.repeat(PLs['DESC_TRAMITACAO'], lens),
    'DESC_SITUACAO': np.repeat(PLs['DESC_SITUACAO'], lens),
    'ID_PROPOSICAO': np.repeat(PLs['ID_PROPOSICAO'], lens),
    'PROPOSICAO_BASE': np.repeat(PLs['PROPOSICAO_BASE'], lens),
    'URL_PROPOSICOES': np.repeat(PLs['URL_PROPOSICOES'], lens),
    'TEOR_URL': np.repeat(PLs['TEOR_URL'], lens)
})
# autores.shape[0]
# autores.reset_index(level=0, inplace=True)
# autores.head(3)
```

### 2.3 - Conta e agrupa proposições

```
autores['N_PLS'] = 1

autores_G = autores.astype("str").groupby(["NOME_AUTOR"], as_index=False).agg({
    'NOME_AUTOR': "first",
    'N_PLS': 'count',
    'TITULO_PROPOSICAO': lambda tags: ', '.join(tags),
    'TEMA': lambda tags: ', '.join(tags),
    'PALAVRAS_CHAVE': lambda tags: ', '.join(tags),
    'EMENTA': lambda tags: ', '.join(tags),
    'DATA_APRESENTACAO': lambda tags: ', '.join(tags),
    'DESC_TRAMITACAO': lambda tags: ', '.join(tags),
    'DESC_SITUACAO': lambda tags: ', '.join(tags),
    'ID_AUTOR': "first",
    'ID_PROPOSICAO': lambda tags: ', '.join(tags),
    'PROPOSICAO_BASE': lambda tags: ', '.join(tags),
    'URL_PROPOSICOES': lambda tags: ', '.join(tags),
    'TEOR_URL': lambda tags: ', '.join(tags)
})
```

```

keepCol = ['NM_URNA', 'GENERO', 'COR_RACA', 'ESTADO_CIVIL', 'ESCOLARIDADE', 'OCUPACAO_ANTERIOR', 'FAIXA_ETARIA', 'PARTIDO_ATUAL', 'ID

listaDeputados_F = listaDeputados[keepCol]

deputados_f = pd.DataFrame(listaDeputados_F[listaDeputados_F['TEM_TWITTER']=='S'])
# twitter_f.shape[0]

autores_proposicoes = pd.merge(autores_G.astype("str"), deputados_f.astype("str"), how='inner', left_on=['NOME_AUTOR'], right_on=
# autores_proposicoes = autores_proposicoes.dropna()
# autores_proposicoes.shape[0]
# autores_proposicoes = autores_proposicoes.drop(['SG_PARTIDO', 'SG_UF'], axis=1)
# autores_proposicoes.head(3)
# autores_proposicoes.shape[0]
# autores_proposicoes.to_csv("autoresProposicoes.csv")

```

## 2.4 - Cria base de dados dos tweets

```

tweets2020 = pd.DataFrame(tweetsDeputados[tweetsDeputados['coleta']=='C1'])
rt = ~tweets2020['text'].str.startswith('RT @')

tweets2020 = tweets2020[rt]

# list(tweetsDeputados.columns)
# tweetsDeputados.head(3)
tweets2020.to_csv("tweets2020.csv")

# csv.field_size_limit(sys.maxsize)
arq_csv = "tweets2020.csv"

f = open("stopwords.txt", 'r', encoding='utf-8')
stopwords = [name.rstrip().lower() for name in f]

with open(arq_csv, encoding="utf-8") as f:
    vTweets = [row["text"] for row in DictReader(f)]

vFrases = []

for idx, tweet in enumerate(vTweets):
    tweet = re.sub(r"https?:\/\/\/(\S+)", "", unicode(tweet))
    # tira toda a pontuação exceto arraba e jogo da velha
    # tweet = re.sub(r"^[Pp]@#+", "", tweet)
    # tira toda a pontuação
    tweet = re.sub(r"^[w\s]", ' ', tweet)
    tweet = " ".join([x for x in tweet.split(' ') if x.lower() not in stopwords])
    tweet = tweet.rstrip()
    tweet = tweet.strip()
    tweet = tweet.lower()
    vFrases.append(tweet)
# vFrases

```

## 2.5 - Agrupa tweets com médias de rts e curtidas

```

tweets2020['TWEET_KEYWORDS'] = vFrases
tweets2020['media_retweets'] = tweets2020['retweet_count']
tweets2020['soma_retweets'] = tweets2020['retweet_count']
tweets2020['media_fav'] = tweets2020['favorite_count']
tweets2020['soma_fav'] = tweets2020['favorite_count']

tweets2020['count'] = 1

# tweets2020.tail(3)

tweets2020_G = tweets2020.groupby(["screen_name"], as_index=False).agg({
    'id': 'first',
    'screen_name': 'first',
    'text': lambda tags: ', '.join(tags),
    'TWEET_KEYWORDS': lambda tags: ', '.join(tags),
    'count': 'sum',
    'media_retweets': 'mean',
    'soma_retweets': 'sum',
    'media_fav': 'mean',
    'soma_fav': 'sum',
    'created_at': lambda tags: ', '.join(tags),
    'media_type': lambda tags: ', '.join(tags),
    'lang': lambda tags: ', '.join(tags),
    'coleta': 'first'
})

tweets2020_G['media_retweets'] = tweets2020_G['media_retweets'].round(0)
tweets2020_G['media_fav'] = tweets2020_G['media_fav'].round(0)

# tweets2020_G.shape[0]
# tweets2020_G.head(3)

```

## 2.6 - Combina bancos de dados de proposições e tweets e formata rótulos

```

1: usuarios_proposicoes = pd.merge(autores_proposicoes,tweets2020_G, how='inner', left_on=['PERFIL_TWITTER'], right_on=['screen_name']

# usuarios_proposicoes.reset_index(level=0, inplace=True)

usuarios_proposicoes = usuarios_proposicoes.rename(columns={'id':'TWEET_IDS', 'text':'TEXT', 'count':'TWEET_COUNTS', 'media_retwee
usuarios_proposicoes = usuarios_proposicoes.drop(['PERFIL_TWITTER'],axis=1)

# list(usuarios_proposicoes.columns)
# usuarios_proposicoes.shape[0]

# usuarios_proposicoes['PALAVRAS_CHAVE'] = usuarios_proposicoes['PALAVRAS_CHAVE'].astype("str").apply(lambda x: ', '.join(set(x.sp
# usuarios_proposicoes['TWEET_KEYWORDS'] = usuarios_proposicoes['TWEET_KEYWORDS'].astype("str").apply(lambda x: ', '.join(set(x.sp

usuarios_proposicoes['PLS_KEYWORDS'] = usuarios_proposicoes['PLS_KEYWORDS'].str.replace(", ", " ")
usuarios_proposicoes['TWEET_KEYWORDS'] = usuarios_proposicoes['TWEET_KEYWORDS'].str.replace(", ", " ")
# usuarios_proposicoes.head(3)

usuarios_proposicoes.to_csv("usuariosProposicoes.csv")

# usuarios_proposicoes = pd.read_csv("usuariosProposicoes.csv")
# usuarios_proposicoes.shape[0]

# usuarios_proposicoes.shape[0]
# usuarios_proposicoes = usuarios_proposicoes.dropna()
# usuarios_proposicoes.shape[0]

# usuarios_proposicoes.to_csv("usuariosProposicoes.csv")

# usuarios_proposicoes = usuarios_proposicoes.dropna()
# usuarios_proposicoes.shape[0]

```

## 2.7 - Gera gráfico da proporção de Deputados Federais que submeteram proposições e publicaram tweets em 2020

```

top = percentual.plot(kind='barh',
                    colormap='tab10',
                    figsize=(11, 8),width=0.8)

for i, v in enumerate(list(percentual)):
    top.text(v+0.4, i-0.25, str(v)+"%", fontsize=12, color='black')

plt.tick_params(axis='y', which='major', labelsize=14)

# top.margins(x=1)
top.spines['right'].set_visible(False)
# top.spines['top'].set_visible(False)
# top.spines['bottom'].set_visible(False)

# top.get_legend().remove()

# plt.rcParams.update(IPython_default);
# plt.style.use("fivethirtyeight")
plt.grid(color='b', linestyle='-', linewidth=.1)
plt.savefig("novos-graficos/Figura x - Proporção de Deputados Federais que submeteram proposições e publicaram tweets em 2020.jpg")

```

## 2.8 - Cria índice de consistência temática

### 2.9 - Remove stop words das proposições

```

keywordsCamara = usuarios_proposicoes['PLS_KEYWORDS'].astype("str").values.tolist()
# usuarios_proposicoes['PLS_KEYWORDS'].shape[0]
# Len(keywordsCamara)
# Len(keywordsTwitter)
# keywordsCamara[0]

f = open("stopwords.txt", 'r', encoding='utf-8')
stopwords = [name.rstrip().lower() for name in f]
vCamara = []
vCamara1 = []

for frase in keywordsCamara:
    frase = re.sub(r'^\w\s|', '', unicode(frase))
    frase = re.sub(r'\n_', ' ', frase)
    frase = re.sub(r'\n', ' ', frase)
    frase = re.sub(r'_', ' ', frase)
    # frase = re.sub(r' sobre | qualquer | das | da | dos | do | de | e | a | o | para | com | em | ou | à | na | no ', ' ', frase)
    frase = " ".join([x for x in frase.split(' ') if x.lower() not in stopwords])
    frase = frase.replace(" do ", "")
    frase = frase.lower()
    frase = frase.replace(" ,", ",")
    frase = frase.replace(" ", " ")
    frase = frase.replace(" ", " ")
    frase = frase.strip()
    vCamara1.append(frase)
    vfrases1 = frase.split(' ')
    vfrases2 = []
    for x in vfrases1:
        if x not in vfrases2 and x != '':
            vfrases2.append(x)
    vCamara.append(vfrases2)

```

## 2.10 - Remove stop words dos tweets

```

keywordsTwitter = usuarios_proposicoes['TWEET_KEYWORDS'].astype("str").values.tolist()

vTwitter = []

for frase in keywordsTwitter:
    # frase = re.sub(r'^\w\s',' ',frase)
    frase = re.sub(r'\n',' ',unicode(frase))
    frase = frase.replace(" ", " ")
    frase = frase.replace(" ", " ")
    # # frase = " ".join([x for x in tweet.split(' ') if x.lower() not in stopwords])
    # frase = frase.lower()
    frase = " ".join([x for x in frase.split(' ') if x.lower() not in stopwords])
    frase = frase.strip()
    # vTwitter.append(unicode(frase).split(' '))
    vfrases1 = unicode(frase).split(' ')
    vfrases2 = []
    for x in vfrases1:
        if x not in vfrases2:
            vfrases2.append(x)
    vTwitter.append(vfrases2)

# usuarios_proposicoes['TWEET_KEYWORDS'].head(3)
# usuarios_proposicoes['TWEET_KEYWORDS'].shape[0]
len(vTwitter[0])
# vTwitter[0]

```

## 2.11 - Faz a interseção entre tweets e propostas

```

somaLista = []
for x,y in zip(vCamara, vTwitter):
    somaLista.append(x+y)

# Len(vTwitter[0])
# Len(vCamara[0])
# Len(somaLista[0])

uniaoLista = []

for lista in somaLista:
    vUnicos = []
    for palavra in lista:
        if palavra not in vUnicos:
            vUnicos.append(palavra)
    uniaoLista.append(vUnicos)

# somaLista[0]

```

## 2.12 - Mede a semelhança entre conteúdo dos tweets e das proposições e cria rótulos a partir dos resultados

```

intersecao = [ (set.intersection(set(x), set(y))) for x,y in zip(vCamara, vTwitter)]
cTematica = [ (round(len(x)/len(y)*100,2)) for x,y in zip(intersecao,vCamara)]

# Len(cTematica)
# usuarios_proposicoes.shape[0]

# teste1 = cTematica.dropna()
# Len(teste1)

# print(type(cTematica))
# Len(cTematica2)

usuarios_proposicoes['CONSISTENCIA_TEMATICA'] = cTematica
usuarios_proposicoes['INTERSECAO'] = intersecao
usuarios_proposicoes.to_csv("usuariosProposicoes.csv", index=False)
# teste = usuarios_proposicoes[['NM_URNA', 'CONSISTENCIA_TEMATICA']]
# teste.shape[0]
# teste.sort_values(by='CONSISTENCIA_TEMATICA', ascending=False)
# teste = teste.dropna()
# usuarios_proposicoes

faixa = ['cTEMATICA_G1', 'cTEMATICA_G2', 'cTEMATICA_G3', 'cTEMATICA_G4']
usuarios_proposicoes['AGRUPAMENTO_cTEMATICA'] = pd.qcut(usuarios_proposicoes['CONSISTENCIA_TEMATICA'],4,precision=0,labels=faixa)
# qcut = usuarios_proposicoes.groupby('AGRUPAMENTO_cTEMATICA').agg(Lambda x: x.mode())
# qcut

```

### 2.13 - Calcula a média geral de consistência temática

```

ct_genero = usuarios_proposicoes.groupby(["GENERO"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_genero['CATEGORIA'] = "GENERO"
ct_raca = usuarios_proposicoes.groupby(["COR_RACA"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_raca['CATEGORIA'] = "COR_RACA"
ct_estadoCivil = usuarios_proposicoes.groupby(["ESTADO_CIVIL"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_estadoCivil['CATEGORIA'] = "ESTADO_CIVIL"
ct_escolaridade = usuarios_proposicoes.groupby(["ESCOLARIDADE"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_escolaridade['CATEGORIA'] = "ESCOLARIDADE"
ct_ocupacao = usuarios_proposicoes.groupby(["OCUPACAO_ANTERIOR"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ocupacao['CATEGORIA'] = "OCUPACAO_ANTERIOR"
ct_partido = usuarios_proposicoes.groupby(["PARTIDO_ATUAL"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_partido['CATEGORIA'] = "PARTIDO_ATUAL"
ct_ideologia = usuarios_proposicoes.groupby(["IDEOLOGIA_PARTIDO_ATUAL"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ideologia['CATEGORIA'] = "IDEOLOGIA_PARTIDO_ATUAL"
ct_milionario = usuarios_proposicoes.groupby(["MILIONARIO"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_milionario['CATEGORIA'] = "MILIONARIO"
ct_estado = usuarios_proposicoes.groupby(["ESTADO_ELEICAO"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_estado['CATEGORIA'] = "ESTADO_ELEICAO"
ct_regiao = usuarios_proposicoes.groupby(["REGIAO_ELEICAO"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_regiao['CATEGORIA'] = "REGIAO_ELEICAO"
ct_reeleito = usuarios_proposicoes.groupby(["ST_REELEICAO"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_reeleito['CATEGORIA'] = "ST_REELEICAO"
ct_ag_idade = usuarios_proposicoes.groupby(["AGRUPAMENTO_IDADE"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ag_idade['CATEGORIA'] = "AGRUPAMENTO_IDADE"
ct_ag_bens = usuarios_proposicoes.groupby(["AGRUPAMENTO_BENS"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ag_bens['CATEGORIA'] = "AGRUPAMENTO_BENS"
ct_ag_votos = usuarios_proposicoes.groupby(["AGRUPAMENTO_VOTOS"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ag_votos['CATEGORIA'] = "AGRUPAMENTO_VOTOS"
ct_ag_seguidores = usuarios_proposicoes.groupby(["AGRUPAMENTO_SEGUIDORES"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ag_seguidores['CATEGORIA'] = "AGRUPAMENTO_SEGUIDORES"
ct_ag_bseguidores = usuarios_proposicoes.groupby(["AGRUPAMENTO_BALANCO_SEGUIDORES"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ag_bseguidores['CATEGORIA'] = "AGRUPAMENTO_BALANCO_SEGUIDORES"
ct_ag_tweets = usuarios_proposicoes.groupby(["AGRUPAMENTO_NUMERO_TWEETS"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ag_tweets['CATEGORIA'] = "AGRUPAMENTO_NUMERO_TWEETS"
ct_ag_btweets = usuarios_proposicoes.groupby(["AGRUPAMENTO_BALANCO_TWEETS"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_ag_btweets['CATEGORIA'] = "AGRUPAMENTO_BALANCO_TWEETS"

```

```

consistencia_tematica = pd.concat([ct_genero, ct_raca, ct_estadoCivil, ct_escolaridade, ct_partido, ct_ideologia, ct_milionario,
consistencia_tematica['CONSISTENCIA_TEMATICA'].mean()
# consistencia_tematica

```

### 2.14 - Calcula a média de consistência temática por estado

```

ct_estado = usuarios_proposicoes.groupby(["ESTADO_ELEICAO"]).agg({'CONSISTENCIA_TEMATICA': 'mean'})
ct_estado['CONSISTENCIA_TEMATICA'] = ct_estado['CONSISTENCIA_TEMATICA'].round(2)

ct_estado['CONSISTENCIA_TEMATICA'].mean()

# ct_estado
# print(type(ct_estado))

```

```

# List(usuarios_proposicoes.columns)
usuarios_proposicoes['N_PLS']

```

## 2.15 - Gera gráfico com número de tweets por Estado

```

t_estado = usuarios_proposicoes.groupby(["ESTADO_ELEICAO"]).agg({'TWEET_COUNTS': 'sum'})

t_estado = t_estado.sort_values(by='TWEET_COUNTS', ascending=True)

total = t_estado['TWEET_COUNTS']
total

# ct_estado
# print(type(ct_estado))

top = total.plot(kind='barh',
                 colormap='tab10',
                 figsize=(12, 8),width=0.8)

for i, v in enumerate(list(total)):
    top.text(v+0.4, i-0.25, v, fontsize=12, color='black')

plt.tick_params(axis='y', which='major', labelsize=14)

# top.margins(x=1)
# top.spines['right'].set_visible(False)
# top.spines['top'].set_visible(False)
# top.spines['bottom'].set_visible(False)

# top.get_legend().remove()
top.set_ylabel('')
# plt.rcParams.update(IPython_default);
# plt.style.use("fivethirtyeight")
plt.grid(color='b', linestyle='-', linewidth=1)
plt.savefig("novos-graficos/Figura x - Quantidade de Tweets por Estado.jpg",bbox_inches = 'tight')
# plt.show()

```

## 2.16 - Cria tabela com as médias de consistência temática por estado

```

faixa = ['G4','G3','G2','G1']
ct_estado['AGRUPAMENTO'] = pd.qcut(ct_estado['CONSISTENCIA_TEMATICA'],4,precision=0,labels=faixa)

ct_estado = ct_estado.sort_values(['CONSISTENCIA_TEMATICA'], ascending=[False])

ct_estado.to_csv("consistenciaTematica-estados.csv", index=False)

ct_estado
# qcut = usuarios_proposicoes.groupby('AGRUPAMENTO_cTEMATICA').agg(Lambda x: x.mode())
# qcut

# d2g.upload(ct_estado, spreadsheet_key, 'consistenciaTematica', credentials=credentials, row_names=True)

```

## 2.17 - Salva resultados em arquivo csv

```

%%capture

# usuarios_proposicoes.shape[0]
# len(vCamara)
# len(vTwitter)

usuarios_proposicoes['PLS_KEYWORDS'] = vCamara
usuarios_proposicoes['TWEET_KEYWORDS'] = vTwitter

usuarios_proposicoes['PLS_KEYWORDS'] = usuarios_proposicoes['PLS_KEYWORDS'].astype(str).str.replace(',',' ')
usuarios_proposicoes['PLS_KEYWORDS'] = usuarios_proposicoes['PLS_KEYWORDS'].astype(str).str.replace('[^\w\s]','')

usuarios_proposicoes['TWEET_KEYWORDS'] = usuarios_proposicoes['TWEET_KEYWORDS'].astype(str).str.replace(',',' ')
usuarios_proposicoes['TWEET_KEYWORDS'] = usuarios_proposicoes['TWEET_KEYWORDS'].astype(str).str.replace('[^\w\s]','')

usuarios_proposicoes['INTERSECAO'] = usuarios_proposicoes['INTERSECAO'].astype(str).str.replace(',',' ')
usuarios_proposicoes['INTERSECAO'] = usuarios_proposicoes['INTERSECAO'].astype(str).str.replace('[^\w\s]','')

# usuarios_proposicoes.head(2)

usuarios_proposicoes.to_csv("usuarios_proposicoes-2.csv", index=False)

```

## 2.18 - Gera gráficos gerais

### 2.19 - Maiores e menores índices de consistência temática

```

ct_estado = ct_estado.drop('AGRUPAMENTO', 1)
ct_estado

```

```

ptop10 = ct_estado.nlargest(12,'CONSISTENCIA_TEMATICA').sort_values(by='CONSISTENCIA_TEMATICA', ascending=True)
indices1 = ptop10.apply(lambda row: str(row.CONSISTENCIA_TEMATICA) + "%", axis=1)
indices1 = list(indices1)
ptop10 = ptop10['CONSISTENCIA_TEMATICA']

pbottom10 = ct_estado.nsmallest(15,'CONSISTENCIA_TEMATICA').sort_values(by='CONSISTENCIA_TEMATICA', ascending=True)

indices2 = pbottom10.apply(lambda row: str(row.CONSISTENCIA_TEMATICA) + "%", axis=1)
# indices2 = list(pbottom10['CONSISTENCIA_TEMATICA'])
indices2 = list(indices2)
pbottom10 = pbottom10['CONSISTENCIA_TEMATICA']

```

```

axd = plt.figure(figsize=(8,10)).subplot_mosaic(
    """
    A
    B
    """
)

top10 = ptop10.plot(kind='barh',
                    color='#0A8CB1',
                    width=0.8,ax=axd['A'])

for i, v in enumerate(list(ptop10)):
    top10.text(v-5, i-0.1, indices1[i], color='white',weight='heavy')

```

```

bottom10 = pbottom10.plot(kind='barh',
                          color='#BC0D0D',
                          width=0.8,ax=axd['B'])

# bottom10 = sns.barplot(color='#BC0D0D',data=pbottom10,ax=axd['B'])

for i, v in enumerate(list(pbottom10)):
    bottom10.text(v-4, i-0.1, indices2[i], color='white',weight='heavy')

# for i, v in enumerate(list(votacoes_10)):
#     top.text(v-20, i-0.15, indices[i], fontsize=12, color='white',weight='heavy')

# top10.get_legend().remove()
# bottom10.get_legend().remove()

top10.set_ylabel('')
top10.set_xlabel('')

bottom10.set_ylabel('')
bottom10.set_xlabel('')

bottom10.grid(color='b', linestyle='-', linewidth=.1)
top10.grid(color='b', linestyle='-', linewidth=.1)

plt.tight_layout()
# axd

# plt.show()

plt.savefig("novos-graficos/Figura x - Índices de consistência temática por estado.jpg",bbox_inches = 'tight')
plt.clf()
# plt.show()

```

## 2.20 - Temas e palavras-chave mais frequentes nas proposições

```

vCamara2 = [item for sublist in vCamara for item in sublist]
vCamara3 = ' '.join([str(elem) for elem in vCamara2])

# if "codigo" in vCamara2:
#     print("é ta sim")
# else:
#     print("ta nao")

# len(vCamara2)
# len(vCamara1)

axd = plt.figure(figsize=(16,8)).subplot_mosaic(
    """
    AB
    """,
    gridspec_kw={
        'width_ratios': [5, 6],
        'wspace': 0.1
    })

lang_responses = usuarios_proposicoes['TEMA']

language_counter = Counter()

for response in lang_responses:
    language_counter.update(response.split(','))

languages = []
popularity = []

```

```

for item in language_counter.most_common(10):
    languages.append(item[0])
    popularity.append(item[1])

# n30 = popularity[:10]
# n30 = popularity[-30:]

languages.reverse()
popularity.reverse()

axd['A'].barh(languages, popularity)

# axd['A'].yticks(fontsize=28)
axd['A'].set_yticklabels(languages,fontsize=18)

for i, v in enumerate(popularity):
    axd['A'].text(v-200, i-.07, str(v), color='white',size=17)

# -----

wordcloud = WordCloud(
    # max_font_size=80,
    min_font_size=22,
    max_words=15,
    prefer_horizontal=1,
    # scale=10,
    height=640,
    width=550,
    background_color = 'white',
    stopwords=stopwords).generate(str(vCamera3))

# axd['A'].set_title('Temas',fontsize=20)
# axd['B'].set_title('Palavras-chave',fontsize=20)

```

```

axd['B'].axis('off')
axd['B'].imshow(wordcloud, interpolation = 'bilinear')

# plt.set_title('Palavras-chave',fontsize=20)
# plt.axis('off')
# plt.imshow(wordcloud, interpolation = 'bilinear')

# plt.tight_layout(pad=0)
# plt.show()

plt.savefig("2 - Figura 8 - Temas e palavras-chave mais recorrentes nas proposições.jpg",bbox_inches = 'tight')
# plt.figure(figsize=(5,10))
# bottom10 = sns.barplot(x=popularityr,y=Languagesr, color='#0A8CB1')

# plt.barh(Languages, popularity)
# # plt.title("Most popular languages")
# # plt.ylabel("Programming Languages")
# # plt.xlabel("Number of People who use")

# plt.tight_layout()
# plt.show()

# quase tagcloud
# y = [(x, vCamera.count(x)) for x in set(vCamera)]

```

## 2.21 - Publicações mais compartilhadas no Twitter

```
# tweets2020.head(2)
topTweets = tweets2020.nlargest(3,'retweet_count')
# tweetsTop10
topIds = topTweets['id'].astype("str").values.tolist()

string = "https://twitter.com/OReillyMedia/status/"

topIds = [string + s for s in topIds]
# topIds

class Tweet(object):
    def __init__(self, s, embed_str=False):
        if not embed_str:
            # Use Twitter's oEmbed API
            # https://dev.twitter.com/web/embedded-tweets
            api = 'https://publish.twitter.com/oembed?url={}'.format(s)
            response = requests.get(api)
            self.text = response.json()["html"]
        else:
            self.text = s

    def _repr_html_(self):
        return self.text

# Tweet("https://twitter.com/OReillyMedia/status/1253710046405500928")

for x in topIds:
    Tweet(x)
# vTweets(topIds)
```

## 2.22 - Usuários, hashtags e palavras mais utilizadas no Twitter

```
keywordsTwitter = usuarios_proposicoes['TWEET_KEYWORDS'].astype("str").values.tolist()

vTwitter1 = []

f = open("stopwords.txt", 'r', encoding='utf-8')
stopwords = [name.rstrip().lower() for name in f]

for tweet in keywordsTwitter:
    tweet = re.sub(r'\n', ' ', unicode(tweet))
    tweet = tweet.replace(" ", " ")
    # tweet = tweet.replace("paraben", "parabens")
    tweet = " ".join([x for x in tweet.split(' ') if x.lower() not in stopwords])
    vTwitter1.append(tweet)
# vFrases

# tweets2020['TWEET_KEYWORDS'] = vFrases
# len(vTwitter)
# len(vTwitter1)
# vTwitter[:10]
```

## 2.23 - Extrai as informações dos tweets

```

words_repetition = 1
hashtags_repetition = 1
users_repetition = 1

#variables for how many words, hashtags, users to display
word_limit = 10
hashtag_limit = 10
user_limit = 10

#file path for stopwords and json exports
stopwords_file_path = 'stopwords.txt'
# json_export_file_path = '1-total.json'

with open(stopwords_file_path, encoding='utf-8') as f:
    stopwords = f.read().splitlines()

items = tweets2020.to_dict('records')
# items[0]

#taken from https://stackoverflow.com/a/49146722
def remove_emoji(string):
    emoji_pattern = re.compile("[
        "\U0001F600-\U0001F64F" # emoticons
        "\U0001F300-\U0001F5FF" # symbols & pictographs
        "\U0001F680-\U0001F6FF" # transport & map symbols
        "\U0001F1E0-\U0001F1FF" # flags (iOS)
        "\U00002702-\U000027B0"
        "\U000024C2-\U0001F251"
    ]+", flags=re.UNICODE)
    return emoji_pattern.sub(r'', string)

```

```

#taken from https://gist.github.com/glenbot/4684356
def sanitize_3(user_input, stop_words):
    """Sanitize using standard lists"""
    new_list = []
    for w in user_input:
        if w not in stop_words:
            new_list.append(w)
    return new_list

#remove accents, emoji, convert to all lowercase
for idx, item in enumerate(items):
    items[idx]['text'] = remove_emoji(items[idx]['text'])
    items[idx]['text'] = unicode(items[idx]['text'])
    items[idx]['text'] = items[idx]['text'].lower()
    #print items[idx]['text']

#removing stopwords
#convert every tweet into a List of words
#call sanitize_3 for every List of words
current_string = ''
placeholder_list = [None] * len(items)
master_word_list = []
master_hashtag_list = []
master_user_list = []

for idx, item in enumerate(items):
    #word_list = re.sub("[^\w]", " ", items[idx]['text']).split()
    word_list = re.sub(r'[:,;:~]', ' ', items[idx]['text']).split()
    word_list = sanitize_3(word_list, stopwords)
    for idx2, word in enumerate(word_list):
        if word_list[idx2][:1] == '#':
            current_string += ' ' + word_list[idx2]
            master_hashtag_list.append(word_list[idx2])
        elif word_list[idx2][:1] == '@':

```

```

        current_string += ' ' + word_list[idx2]
        master_user_list.append(word_list[idx2])
    elif word_list[idx2][:3] == '//t':          #dirty workaround
        pass
    elif 'co/' not in word_list[idx2]:
        current_string += ' ' + word_list[idx2]
        master_word_list.append(word_list[idx2])
placeholder_list[idx] = current_string
current_string = ''

#release memory immediately
del placeholder_list[:]

#rank words
#adapted from https://github.com/kevinschau/Word-Rank/blob/master/wordRank.py
l = {}
hashtag_dict = {}
user_dict = {}

for word in master_word_list:
    # if word is in dictionary, increment the value
    # otherwise add the word to dictionary with value 1
    if word in l:
        l[word] += 1
    else:
        l[word] = 1

for hashtag in master_hashtag_list:
    if hashtag in hashtag_dict:
        hashtag_dict[hashtag] += 1
    else:
        hashtag_dict[hashtag] = 1

```

```

for user in master_user_list:
    if user in user_dict:
        user_dict[user] += 1
    else:
        user_dict[user] = 1

wordsV1 = []
wordsV2 = []

# this prints the dict out d by value in descending order
count = 0
# print('\n##### Printing ranking of words #####\n')
for key, value in sorted(iter(l.items()), reverse=True, key=lambda k_v3: (k_v3[1],k_v3[0])):
    if count < word_limit:
        # print('%s: %s' % (key, value))
        wordsV1.append(key)
        wordsV2.append(value)
    count+=1

wordsD1 = dict(zip(wordsV1, wordsV2))
# print(wordsD1)

hashtagV1 = []
hashtagV2 = []

count = 0
# print('\n##### Printing ranking of hashtags #####\n')
for key, value in sorted(iter(hashtag_dict.items()), reverse=True, key=lambda k_v4: (k_v4[1],k_v4[0])):
    if count < hashtag_limit:
        # print('%s: %s' % (key, value))
        hashtagV1.append(key)
        hashtagV2.append(value)

    count +=1

hashtagD1 = dict(zip(hashtagV1, hashtagV2))

```

```

usersV1 = []
usersV2 = []

count = 0
# print('\n##### Printing ranking of most mentioned users #####\n')
for key, value in sorted(iter(user_dict.items()), reverse=True, key=lambda k_v5: (k_v5[1],k_v5[0])):
    if count < user_limit:
        # print('%s: %s' % (key, value))
        usersV1.append(key)
        usersV2.append(value)
        count += 1

usersD1 = dict(zip(usersV1, usersV2))
# print(usersD1)

most_retweeted_list = []
list_is_full = False

for tweet in items:
    if 'rt @' in tweet['text']:
        pass
    elif tweet['retweet_count'] > 10:
        most_retweeted_list.append(tweet)

count = 0

tweets = []
rts = []
ids = []

```

```

for tweet in sorted(most_retweeted_list, key = lambda x: x['retweet_count'], reverse=True):
    current_date = datetime.strptime(tweet['created_at'],'%Y-%m-%d %H:%M:%S+00:00')
    # if (current_date.year == 2020) and (current_date.month == 11) and (current_date.day < 17):
    tweets.append(tweet['screen_name'] + " - " + tweet['text'])
    rts.append(tweet['retweet_count'])
    ids.append(tweet['id'])
    if count == 20:
        break
    count+=1

# rtKeys1 = ['tweet', 'rts', 'ids']
# rtKeys2 = rtKeys1 *Len(ids)
# print(rtKeys2)

retweetsD1 = [{'tweet': tweet, 'rts': rt, 'ids': idd} for tweet,rt,idd in zip(tweets,rts,ids)]

day_list = []
count = 0
for tweet in items:
    current_date = datetime.strptime(tweet['created_at'],'%Y-%m-%d %H:%M:%S+00:00')
    # if (current_date.year == 2020) and (current_date.month == 11) and (current_date.day < 17):
    date_time_str = str(current_date.day) + "/" + str(current_date.month) + "/" + str(current_date.year)
    # print(date_time_str)
    # day_list.append(str(current_date.day) + "/" + str(current_date.month))
    date_time_obj = datetime.strptime(date_time_str, '%d/%m/%Y')
    # print(date_time_obj)
    day_list.append(date_time_obj)

day_dict = {}

for day in day_list:
    if day in day_dict:
        day_dict[day] += 1
    else:
        day_dict[day] = 1

```

```

DatasV1 = []
DatasV2 = []

# #going to break on cases that go from 23 to midnight or backwards
# print('\n##### Printing hourly traffic #####\n')
for key, value in sorted(iter(day_dict.items()), reverse=False, key=lambda k_v: (k_v[0],k_v[1])):
    chave = str(key)
    chave = re.sub(" 00:00:00","", chave)
    DatasV1.append(chave)
    DatasV2.append(value)
    # print('%s: %s' % (key, value))
DatasD1 = dict(zip(DatasV1, DatasV2))

# json.dump({report:[{'words':wordsD1},{'hashtags':hashtagD1},{'users':usersD1},{'retweets':retweetsD1},{'dates':DatasD1}],pre_j

```

## 2.24 - Gera a visualização

```

axd = plt.figure(figsize=(7,15)).subplot_mosaic(
    """
    A
    B
    C
    """,
    gridspec_kw={
        # 'width_ratios': [5, 5, 7],
        # 'height_ratios': [5, 5, 7],
        # 'right': 1.2,
        # 'left': 1,
        'hspace': 0.3
    }
)

```

```

axd['A'].barh(usersV1, usersV2, color='green')

for i, v in enumerate(usersV2):
    axd['A'].text(v-450, i-.07, str(v), color='white',size=12)

axd['B'].barh(hashtagV1, hashtagV2, color='purple')

for i, v in enumerate(hashtagV2):
    axd['B'].text(v-200, i-0.1, str(v), color='white',size=12)

axd['A'].set_yticklabels(usersV1,fontsize=13)
axd['B'].set_yticklabels(hashtagV1,fontsize=13)

```

```

wordcloud = WordCloud(
    # width = 280,
    # height = 2800,
    prefer_horizontal=1,
    max_font_size=40,
    min_font_size=10,
    max_words=18,
    background_color = 'white',
    color_func=lambda *args, **kwargs: "#D38B07",
    stopwords=stopwords).generate(str(vTwitter1))

axd['A'].set_title('Usuários',fontsize=16)
axd['B'].set_title('Hashtags',fontsize=16)
axd['C'].set_title('Palavras',fontsize=16)

axd['C'].axis('off')
axd['C'].imshow(wordcloud)

# axd['A'].add_gridspec({'right': 100, 'wspace': 30})
plt.tight_layout()

# axd['C'].imshow(wordcloud, interpolation = 'bilinear')

plt.savefig("4 - Figura 10 - Usuários, hashtags e palavras mais utilizadas no Twitter.jpg",bbox_inches = 'tight')

```

## 2.25 - Gera gráficos do Rio Grande do Sul

### 2.26 - Carrega base de dados

```

usuarios_proposicoes = pd.read_csv("usuarios_proposicoes-2.csv")
tweets2020 = pd.read_csv("tweets2020.csv")
# usuarios_proposicoes.head(2)
usuarios_proposicoes_rs = usuarios_proposicoes[(usuarios_proposicoes['ESTADO_ELEICAO']=='RS')]

```

### 2.27 - Temas mais recorrentes nas proposições

```

%%capture

perfil = 'ESTADO_ELEICAO'

filtro = usuarios_proposicoes[usuarios_proposicoes['ESTADO_ELEICAO'].isin(['RS'])]

lang_responses = usuarios_proposicoes['TEMA']

language_counter = Counter()

for response in lang_responses:
    language_counter.update(response.split(','))

languages = []
popularity = []

for item in language_counter.most_common(100):
    languages.append(item[0])
    popularity.append(item[1])

# language_counter

contagem_geral = dict(language_counter)
df = pd.DataFrame.from_dict(contagem_geral, orient='index')
# df.columns['TEMA', 'COUNT']

```

```

df.rename(columns = {0:'count'},inplace = True)

df.index.name = 'TEMA'
df['ESTADO_ELEICAO'] = 'GERAL'

df.reset_index()
df.columns
# df
temas = df[['ESTADO_ELEICAO', 'count']]
temas = temas.reset_index()

```

```

soma = temas.groupby([perfil],as_index=False).agg({
    perfil: "first",
    'count': 'sum'
})

temasM_geral = pd.merge(temas, soma, how="left", on=perfil)
temasM_geral['proporcao'] = round(temasM_geral['count_x']/temasM_geral['count_y']*100,2)

# rd = px.line_polar(temasM_geral, r="proporcao", theta="TEMA", color=perfil,line_close=True, width=wd, height=ht)

# rd.update_layout(
#     margin=dict(l=ll, r=rr, t=tt, b=bb),
# )

```

```
temasM_geral
```

```

%%capture
perfil = 'ESTADO_ELEICAO'
filtro = usuarios_proposicoes[usuarios_proposicoes['ESTADO_ELEICAO'].isin(['RS'])]

filtro = usuarios_proposicoes[usuarios_proposicoes['ESTADO_ELEICAO'].isin(['RS'])]

lang_responses = filtro['TEMA']

language_counter = Counter()

for response in lang_responses:
    language_counter.update(response.split(','))

languages = []
popularity = []

for item in language_counter.most_common(100):
    languages.append(item[0])
    popularity.append(item[1])

# language_counter

contagem_geral = dict(language_counter)
df = pd.DataFrame.from_dict(contagem_geral, orient='index')
# df.columns['TEMA', 'COUNT']

```

```

df.rename(columns = {0:'count'}, inplace = True)

df.index.name = 'TEMA'
df['ESTADO_ELEICAO'] = 'RS'

df.reset_index()
df.columns
# df
temas = df[['ESTADO_ELEICAO', 'count']]
temas = temas.reset_index()

```

```

soma = temas.groupby([perfil], as_index=False).agg({
    perfil: "first",
    'count': 'sum'
})

temasM_rs = pd.merge(temas, soma, how="left", on=perfil)
temasM_rs['proporcao'] = round(temasM_rs['count_x']/temasM_rs['count_y']*100,2)

```

```

temasM_rs.sort_values(by='count_x', ascending=False)

```

```

temasM_geral = temasM_geral.nlargest(30, 'count_x')
temasM_geral = temasM_geral.sort_values(by='TEMA')

temasM_rs = temasM_rs.nlargest(30, 'count_x')
temasM_rs = temasM_rs.sort_values(by='TEMA')

TEMAS_C = pd.concat([temasM_geral, temasM_rs])
TEMAS_C.to_clipboard()

```

```

wd = 1000
ht = 800
ll = 300
rr = 250
tt = 5
bb = 5

arq = "novos-graficos/Gráfico x - Temas mais recorrentes nas proposições - Rio Grande do Sul.jpg"

rd = px.line_polar(TEMAS_C, r="proporcao", theta="TEMA", color=perfil, line_close=True, width=wd, height=ht)

rd.update_layout(
    margin=dict(l=ll, r=rr, t=tt, b=bb),
)

# rd.show()

rd.write_image(arq)

```

## 2.28 - Publicações mais compartilhadas no Twitter

```

from IPython.core.display import display, HTML
# perfil = 'ST_REELEICAO'
def topRTs2(perfil):
    unik = list(filtro[perfil].unique())
    # len(unik)
    # listaLista = [[] for i in range(len(unik))]
    listas = []
    # lista

    for idx, k in enumerate(unik):
        # listaLista[idx].append(filtro[filtro[perfil]==k]['SCREEN_NAME'].astype("str").values.tolist())
        listas.append(filtro[filtro[perfil]==k]['SCREEN_NAME'].astype("str").values.tolist())

    # len(lista)
    # lista[1]

    for idx, s in enumerate(listas):
        print(unik[idx])
        tweets2020F = tweets2020[tweets2020['screen_name'].isin(listas[idx])]
        # tweets2020F

        #
        # tweets2020.head(2)
        topTweets = tweets2020F.nlargest(5, 'retweet_count')
        # tweetsTop10
        topIds = topTweets['id'].astype("str").values.tolist()

        string = "https://twitter.com/OReillyMedia/status/"

        topIds = [string + s for s in topIds]
        # topIds

```

```

class Tweet(object):
    def __init__(self, s, embed_str=False):
        try:
            if not embed_str:
                # Use Twitter's oEmbed API
                # https://dev.twitter.com/web/embedded-tweets
                api = 'https://publish.twitter.com/oembed?url={}'.format(s)
                response = requests.get(api)
                self.text = response.json()["html"]
                display(HTML(self.text))
            else:
                self.text = s
        except Exception:
            pass

    def _repr_html_(self):
        try:
            return self.text
        except Exception:
            pass

# Tweet("https://twitter.com/OReillyMedia/status/1253710046405500928")

for x in topIds:
    Tweet(x)
# vTweets(topIds)

```

```

! perfil = 'ESTADO_ELEICAO'
filtro = usuarios_proposicoes[usuarios_proposicoes['ESTADO_ELEICAO'].isin(['RS'])]
topRTs2(perfil)

```

## 2.29 - Usuários e hashtags e palavras mais frequentes no Twitter

```

usuarios_proposicoes_rs = usuarios_proposicoes[(usuarios_proposicoes['ESTADO_ELEICAO']=='RS')]

```

```

listaDeputados = pd.read_csv("db-listaDeputados.csv")

```

```

keywordsTwitter = usuarios_proposicoes_rs['TWEET_KEYWORDS'].astype("str").values.tolist()

vTwitter1 = []

f = open("stopwords.txt", 'r', encoding='utf-8')
stopwords = [name.rstrip().lower() for name in f]

for tweet in keywordsTwitter:
    tweet = re.sub(r'\n', ' ', unicode(tweet))
    tweet = tweet.replace(" ", "")
    # tweet = tweet.replace("paraben", "parabens")
    tweet = " ".join([x for x in tweet.split(' ') if x.lower() not in stopwords])
    vTwitter1.append(tweet)
# vFrases

# tweets2020['TWEET_KEYWORDS'] = vFrases
# len(vTwitter)
# len(vTwitter1)
# vTwitter[:10]

```

## 2.30 - Extrai as informações dos tweets

```
listaDeputados.columns
```

```
# Len(tweets2020)

tweets2020_info = tweets2020.merge(listaDeputados, left_on='screen_name', right_on='PERFIL_TWITTER')
# Len(tweets2020_info)
# tweets2020_rs =
# tweets2020_info

tweets2020_info_rs = tweets2020_info[(tweets2020_info['ESTADO_ELEICAO']=='RS')]
len(tweets2020_info_rs)
```

```
words_repetition = 1
hashtags_repetition = 1
users_repetition = 1

#variables for how many words, hashtags, users to display
word_limit = 10
hashtag_limit = 10
user_limit = 10

#file path for stopwords and json exports
stopwords_file_path = 'stopwords.txt'
# json_export_file_path = '1-total.json'

with open(stopwords_file_path, encoding='utf-8') as f:
    stopwords = f.read().splitlines()

items = tweets2020_info_rs.to_dict('records')
# items[0]
```

```
#taken from https://stackoverflow.com/a/49146722
def remove_emoji(string):
    emoji_pattern = re.compile("[
        '\U0001F600-\U0001F64F' # emoticons
        '\U0001F300-\U0001F5FF' # symbols & pictographs
        '\U0001F680-\U0001F6FF' # transport & map symbols
        '\U0001F1E0-\U0001F1FF' # flags (i05)
        '\U00002702-\U000027B0'
        '\U000024C2-\U0001F251'
    ]+", flags=re.UNICODE)
    return emoji_pattern.sub(r'', string)

#taken from https://gist.github.com/glenbot/4684356
def sanitize_3(user_input, stop_words):
    """Sanitize using standard lists"""
    new_list = []
    for w in user_input:
        if w not in stop_words:
            new_list.append(w)
    return new_list

#remove accents, emoji, convert to all lowercase
for idx, item in enumerate(items):
    items[idx]['text'] = remove_emoji(items[idx]['text'])
    items[idx]['text'] = unidecode(items[idx]['text'])
    items[idx]['text'] = items[idx]['text'].lower()
    #print items[idx]['text']

#removing stopwords
#convert every tweet into a list of words
#call sanitize_3 for every list of words
current_string = ''
placeholder_list = [None] * len(items)
master_word_list = []
master_hashtag_list = []
master_user_list = []
```

```

for idx, item in enumerate(items):
    #word_list = re.sub("[^\w]", " ", items[idx]['text']).split()
    word_list = re.sub(r'[.!,;:~?]', ' ', items[idx]['text']).split()
    word_list = sanitize_3(word_list, stopwords)
    for idx2, word in enumerate(word_list):
        if word_list[idx2][:1] == '#':
            current_string += ' ' + word_list[idx2]
            master_hashtag_list.append(word_list[idx2])
        elif word_list[idx2][:1] == '@':
            current_string += ' ' + word_list[idx2]
            master_user_list.append(word_list[idx2])
        elif word_list[idx2][:3] == '//t':          #dirty workaround
            pass
        elif 'co/' not in word_list[idx2]:
            current_string += ' ' + word_list[idx2]
            master_word_list.append(word_list[idx2])
    placeholder_list[idx] = current_string
    current_string = ''

#release memory immediately
del placeholder_list[:]

#rank words
#adapted from https://github.com/kevinschau/Word-Rank/blob/master/wordRank.py
l = {}
hashtag_dict = {}
user_dict = {}

for word in master_word_list:
    # if word is in dictionary, increment the value
    # otherwise add the word to dictionary with value 1
    if word in l:
        l[word] += 1
    else:
        l[word] = 1

```

```

for hashtag in master_hashtag_list:
    if hashtag in hashtag_dict:
        hashtag_dict[hashtag] += 1
    else:
        hashtag_dict[hashtag] = 1

for user in master_user_list:
    if user in user_dict:
        user_dict[user] += 1
    else:
        user_dict[user] = 1

wordsV1 = []
wordsV2 = []

# this prints the dict out sorted by value in descending order
count = 0
# print('\n##### Printing ranking of words #####\n')
for key, value in sorted(iter(l.items()), reverse=True, key=lambda k_v3: (k_v3[1], k_v3[0])):
    if count < word_limit:
        # print('%s: %s' % (key, value))
        wordsV1.append(key)
        wordsV2.append(value)
    count+=1

wordsD1 = dict(zip(wordsV1, wordsV2))
# print(wordsD1)

hashtagV1 = []
hashtagV2 = []

count = 0

```

```

for key, value in sorted(iter(hashtag_dict.items()), reverse=True, key=lambda k_v4: (k_v4[1],k_v4[0]]):
    if count < hashtag_limit:
        # print('%s: %s' % (key, value))
        hashtagV1.append(key)
        hashtagV2.append(value)

    count +=1

hashtagD1 = dict(zip(hashtagV1, hashtagV2))

# print(hashtagD1)

# plt.rcParams()
# fig, ax = plt.subplots()

# y_pos = np.arange(len(hashtagV1))
# error = np.random.rand(len(hashtagV1))

# ax.barh(y_pos, hashtagV2, xerr=error, align='center')
# ax.set_yticks(y_pos)
# ax.set_yticklabels(hashtagV1)
# ax.invert_yaxis() # Labels read top-to-bottom
# ax.set_xlabel('Frequência')
# ax.set_title('Hashtags mais utilizadas')

# plt.show()

usersV1 = []
usersV2 = []

count = 0

```

```

for key, value in sorted(iter(user_dict.items()), reverse=True, key=lambda k_v5: (k_v5[1],k_v5[0]]):
    if count < user_limit:
        # print('%s: %s' % (key, value))
        usersV1.append(key)
        usersV2.append(value)
        count += 1

usersD1 = dict(zip(usersV1, usersV2))
# print(usersD1)

most_retweeted_list = []
list_is_full = False

for tweet in items:
    if 'rt @' in tweet['text']:
        pass
    elif tweet['retweet_count'] > 10:
        most_retweeted_list.append(tweet)

count = 0

tweets = []
rts = []
ids = []

# print('\n##### Printing most retweeted tweets #####\n')
for tweet in sorted(most_retweeted_list, key = lambda x: x['retweet_count'], reverse=True):
    current_date = datetime.strptime(tweet['created_at'], '%Y-%m-%d %H:%M:%S+00:00')
    # if (current_date.year == 2020) and (current_date.month == 11) and (current_date.day < 17):
    tweets.append(tweet['screen_name'] + " - " + tweet['text'])
    rts.append(tweet['retweet_count'])
    ids.append(tweet['id'])
    if count == 20:
        break
    count+=1

```

```

retweetsD1 = [{'tweet': tweet, 'rts': rt, 'ids': idd} for tweet,rt,idd in zip(tweets,rts,ids)]

day_list = []
count = 0
for tweet in items:
    current_date = datetime.strptime(tweet['created_at'],'%Y-%m-%d %H:%M:%S+00:00')
    # if (current_date.year == 2020) and (current_date.month == 11) and (current_date.day < 17):
    date_time_str = str(current_date.day) + "/" + str(current_date.month) + "/" + str(current_date.year)
    # print(date_time_str)
    # day_list.append(str(current_date.day) + "/" + str(current_date.month))
    date_time_obj = datetime.strptime(date_time_str, '%d/%m/%Y')
    # print(date_time_obj)
    day_list.append(date_time_obj)

day_dict = {}

for day in day_list:
    if day in day_dict:
        day_dict[day] += 1
    else:
        day_dict[day] = 1

DatasV1 = []
DatasV2 = []

# #going to break on cases that go from 23 to midnight or backwards
# print('\n##### Printing hourly traffic #####\n')
for key, value in sorted(iter(day_dict.items()), reverse=False, key=lambda k_v: (k_v[0],k_v[1])):
    chave = str(key)
    chave = re.sub(" 00:00:00","", chave)
    DatasV1.append(chave)
    DatasV2.append(value)
    # print('%s: %s' % (key, value))
DatasD1 = dict(zip(DatasV1, DatasV2))

```

### 2.31 - Gera a visualização

```

axd = plt.figure(figsize=(6,10)).subplot_mosaic(
    """
    A
    B
    """,
    gridspec_kw={
        # 'width_ratios': [5, 5, 7],
        # 'height_ratios': [5, 5, 7],
        # 'right': 1.2,
        # 'left': 1,
        # 'hspace': 0.3
    })

# usersV1.reverse()
# usersV2.reverse()

# hashtagV1.reverse()
# hashtagV2.reverse()

axd['A'].barh(usersV1, usersV2, color='green')

for i, v in enumerate(usersV2):
    axd['A'].text(v-30, i-.07, str(v), color='white',size=12)

axd['B'].barh(hashtagV1, hashtagV2, color='purple')

```

```

for i, v in enumerate(hashtagV2):
    axd['B'].text(v-28, i-0.1, str(v), color='white',size=12)

axd['A'].set_yticklabels(usersV1,fontsize=13)
axd['B'].set_yticklabels(hashtagV1,fontsize=13)

# -----

axd['A'].set_title('Usuários',fontsize=16)
axd['B'].set_title('Hashtags',fontsize=16)

# axd['A'].add_gridspec({'right': 100,'wspace': 30})
# plt.tight_layout()

# axd['C'].imshow(wordcloud, interpolation = 'bilinear')

plt.savefig("novos-graficos/Figura x - Usuários, hashtags e palavras mais frequentes no Twitter - Rio Grande do Sul.jpg",bbox_inc

```

## 2.32 - Palavras mais utilizadas nas proposições e tweets

```

def tagclouds(perfil,ws,hs,anq):
    unik = list(usuario_proposicoes[perfil].unique())
    tags = ["PLS_KEYWORDS", "TWEET_KEYWORDS", "INTERSECAO"]

    nomes = ['Projetos de Lei', 'Tweets', 'Interseção']

    axs1 = []

    fig = plt.figure(figsize=(20,20))
    # gs0 = matplotlib.gridspec.GridSpec(len(unik),len(tags), figure=fig)
    gs0 = matplotlib.gridspec.GridSpec(len(unik),len(tags), figure=fig,wspace=ws,hspace=hs)

    for i1,t in enumerate(unik):
        for i2,k in enumerate(tags):
            axs1.append(fig.add_subplot(gs0[i1,i2]))

    cores = ['#D38B07', '#3399FF','#ABAB5B']
    mxCores = [cores for k in unik]

    matrix = [tags for k in unik]
    indMatrix1 = [[i for i in range(len(tags))] for value in matrix]
    contagem = [i for i in range(len(tags)*len(unik))]
    indMatrix2 = [contagem[x:x+3] for x in range(0, len(contagem), 3)]

    f = open("stopwords.txt", 'r', encoding='utf-8')
    stopwords = [name.rstrip().lower() for name in f]

```

```

# for idx1, t in enumerate(tags):
for idx1, k in enumerate(unik):

    for idx2, t in enumerate(tags):
    # for idx2, k in enumerate(unik):
        wordcloud = WordCloud(
            # width = 50,
            # height = 50,
            prefer_horizontal=1,
            max_font_size=40,
            min_font_size=10,
            max_words=10,
            background_color = 'white',
            color_func=lambda *args, **kwargs: mxCores[idx1][idx2],
            # color_func=lambda *args, **kwargs: cores[idx2],
            stopwords=stopwords).generate(str(usuario_proposicoes[usuario_proposicoes[perfil]==k][t]))

logro = []
axs1[indMatrix2[idx1][idx2]].set_yticklabels(logro)
axs1[indMatrix2[idx1][0]].set_ylabel(unik[idx1],fontsize=17)
axs1[idx2].set_title(nomes[idx2],fontsize=17)

axs1[indMatrix2[idx1][idx2]].tick_params(
    axis='both',
    which='both',
    bottom=False,
    top=False,
    # left=False,
    # LabelLeft=False,
    labelbottom=False)

axs1[indMatrix2[idx1][idx2]].imshow(wordcloud, interpolation = 'bilinear')
# axs[idx].plot([-1, 1.2], [0, 0], color='black', lw=1, transform=axs[idx].transAxes, clip_on=False)

```

```
plt.savefig(arq,bbox_inches = 'tight')
```

```

usuarios_proposicoes['GERAL'] = 'GERAL'
perfil = 'GERAL'
arq = "novos-graficos/Figura x - Nuvens de palavras sobre proposições e tweets - Geral.jpg"
tagclouds(perfil,0.1,-0.73,arq)

```

```

def tagclouds2(perfil,ws,hs,arq):
    unik = list(filtro[perfil].unique())
    tags = ["PLS_KEYWORDS", "TWEET_KEYWORDS", "INTERSECAO"]

    nomes = ['Projetos de Lei', 'Tweets', 'Interseção']

    axs1 = []

    fig = plt.figure(figsize=(20,20))
    # gs0 = matplotlib.gridspec.GridSpec(len(unik),len(tags), figure=fig)
    gs0 = matplotlib.gridspec.GridSpec(len(unik),len(tags), figure=fig,wspace=ws,hspace=hs)

    for i1,t in enumerate(unik):
        for i2,k in enumerate(tags):
            axs1.append(fig.add_subplot(gs0[i1,i2]))

    cores = ['#D38B07', '#3399FF','#ABAB5B']
    mxCores = [cores for k in unik]

```

```

matrix = [tags for k in unik]
indMatrix1 = [[i for i in range(len(tags))] for value in matrix]
contagem = [i for i in range(len(tags)*len(unik))]
indMatrix2 = [contagem[x:x+3] for x in range(0, len(contagem), 3)]

f = open("stopwords.txt", 'r', encoding='utf-8')
stopwords = [name.rstrip().lower() for name in f]

# for idx1, t in enumerate(tags):
for idx1, k in enumerate(unik):

    for idx2, t in enumerate(tags):
    # for idx2, k in enumerate(unik):
        wordcloud = WordCloud(
            # width = 50,
            # height = 50,
            prefer_horizontal=1,
            max_font_size=40,
            min_font_size=10,
            max_words=10,
            background_color = 'white',
            color_func=lambda *args, **kwargs: mxCores[idx1][idx2],
            # color_func=lambda *args, **kwargs: cores[idx2],
            stopwords=stopwords).generate(str(filtro[filtro[perfil]==k][t]))

logro = []
axs1[indMatrix2[idx1][idx2]].set_yticklabels(logro)
axs1[indMatrix2[idx1][0]].set_ylabel(unik[idx1],fontsize=17)
axs1[idx2].set_title(nomes[idx2],fontsize=17)

```

```

axs1[indMatrix2[idx1][idx2]].tick_params(
    axis='both',
    which='both',
    bottom=False,
    top=False,
    # left=False,
    # labelleft=False,
    labelbottom=False)

axs1[indMatrix2[idx1][idx2]].imshow(wordcloud, interpolation = 'bilinear')
# axs[idx].plot([-1, 1.2], [0, 0], color='black', lw=1, transform=axs[idx].transAxes, clip_on=False)

# axd['A'].add_gridspec({'right': 100, 'wspace': 30})
# plt.tight_layout()

plt.savefig(arq,bbox_inches = 'tight')

```

```

arq = "novos-graficos/Figura x - Nuvens de palavras sobre proposições e tweets - Rio Grande do Sul.jpg"
filtro = usuarios_proposicoes[usuarios_proposicoes['ESTADO_ELEICAO'].isin(['RS'])]
perfil = 'ESTADO_ELEICAO'

tagclouds2(perfil,0.1,-0.5,arq)

```

### 3 - Formação de redes

### 3.1 - Carrega bases de dados

```
listaDeputados = pd.read_csv("db-listaDeputados.csv")

usuarios_proposicoes = pd.read_csv("usuarios_proposicoes-2.csv")
tweets2020 = pd.read_csv("tweets2020.csv")
# usuarios_proposicoes.head(2)
usuarios_proposicoes_rs = usuarios_proposicoes[(usuarios_proposicoes['ESTADO_ELEICAO']=='RS')]

votacoes = pd.read_csv("db-votacoes.csv")

usuarios_proposicoes_rs.to_csv("db-usuarios-rs.csv", index=False)

# Len(tweets2020)
proposicoes = pd.read_csv("db-proposicoes.csv")
```

### 3.2 - Junta os tweets com as informações demográficos dos seus autores

```
tweets2020_info = tweets2020.merge(listaDeputados, left_on='screen_name', right_on='PERFIL_TWITTER')
tweets2020_info_rs = tweets2020_info[(tweets2020_info['ESTADO_ELEICAO']=='RS')]
# Len(tweets2020_info_rs)
```

### 3.3 - Proposições com maior percentual de aprovação e rejeição

```
votacoes["TITULO_PROPOSICAO"] = votacoes["TITULO_PROPOSICAO"].str.replace("\\((.*?)\\)", "")
votacoes["TITULO_PROPOSICAO"] = votacoes["TITULO_PROPOSICAO"].str.strip()
```

```
votacoes_c = votacoes[["TITULO_PROPOSICAO", "VOTOS_SIM"]]
votacoes_c = votacoes_c.drop_duplicates(subset='TITULO_PROPOSICAO')

votacoes_c10 = votacoes_c.nlargest(5, 'VOTOS_SIM').sort_values(by='VOTOS_SIM', ascending=True)

votacoes_c10['percentual'] = votacoes_c10.apply(lambda row: round(row.VOTOS_SIM/513*100,2), axis=1)
votacoes_c10 = votacoes_c10.set_index('TITULO_PROPOSICAO')

votacoes_c10['rotulos'] = votacoes_c10.apply(lambda row: str(int(row.VOTOS_SIM)) + " (" + str(row.percentual)+"%" ) , axis=1)

votacoes_s10 = votacoes_c10['percentual']
# type(votacoes_c10)
```

```
votacoes_b = votacoes[["TITULO_PROPOSICAO", "VOTOS_NAO"]]
# votacoes_b = votacoes_b.drop_duplicates(subset='TITULO_PROPOSICAO')

votacoes_b10 = votacoes_b.nlargest(5, 'VOTOS_NAO').sort_values(by='VOTOS_NAO', ascending=True)

votacoes_b10['percentual'] = votacoes_b10.apply(lambda row: round(row.VOTOS_NAO/513*100,2), axis=1)
votacoes_b10 = votacoes_b10.set_index('TITULO_PROPOSICAO')

votacoes_b10['rotulos'] = votacoes_b10.apply(lambda row: str(int(row.VOTOS_NAO)) + " (" + str(row.percentual)+"%" ) , axis=1)

votacoes_n10 = votacoes_b10['percentual']
# type(votacoes_b10)
```

```

axd = plt.figure(figsize=(8,10)).subplot_mosaic(
    """
    A
    B
    """
)

top10 = votacoes_s10.plot(kind='barh',
                          color='green',
                          width=0.8,ax=axd['A'])

indices1 = list(votacoes_c10['rotulos'])

for i, v in enumerate(list(votacoes_s10)):
    top10.text(v-18, i, indices1[i], color='white',weight='heavy')

bottom10 = votacoes_n10.plot(kind='barh',
                              color='#BC000D',
                              width=0.8,ax=axd['B'])

indices2 = list(votacoes_b10['rotulos'])

for i, v in enumerate(list(votacoes_n10)):
    bottom10.text(v-18, i, indices2[i], color='white',weight='heavy')

top10.set_ylabel('')
top10.set_xlabel('')

bottom10.set_ylabel('')
bottom10.set_xlabel('')

```

```

bottom10.grid(color='b', linestyle='-', linewidth=.1)
top10.grid(color='b', linestyle='-', linewidth=.1)

plt.tight_layout()
# axd

# plt.show()

# plt.savefig("novos-graficos/Figura x - Índices de consistência temática por estado.jpg",bbox_inches = 'tight')
# plt.clf()
# plt.show()

plt.savefig("novos-graficos/Figura x - Proposições com maiores percentuais de aprovação e rejeição.jpg",bbox_inches = 'tight')

```

## 3.4 - Rede de coautoria de proposições parlamentares

### 3.5 - Criação do Grafo

```

%%capture

arq_grafo = "db-proposicoes.csv"
arq_semantico = "grafo-coautoria.csv"
df_col3 = pd.read_csv(arq_grafo)
# writer.writerow(['TITULO_FRENTE', 'DEPUTADOS_INTEGRANTES'])
# writer = csv.writer(fileOut)
# reader = csv.reader(fileIn, delimiter=',')
with open(arq_grafo, encoding="utf-8") as f:
    vTweets = [row["NOME_AUTOR"] for row in DictReader(f)]

col1 = []
col2 = []
vcol3 = []

```

```

for tweet in vTweets:
    # print(tweet)
    vPalavras = tweet.split(',')
    nPalavras = len(vPalavras)

    if nPalavras > 2:
        nLinhas = 0
        vNumeros = []
        for i in range(nPalavras):
            b = nPalavras-(i+1)
            if b>0:
                vNumeros.append(b)
                nLinhas = nLinhas + b
            vcol3.append(nLinhas)

        nNumeros = len(vNumeros)
        invNumeros = vNumeros[::-1]

        c1 = []
        for i in range(nNumeros):
            for j in range(vNumeros[i]):
                c1.append(vPalavras[i])
            col1.extend(c1)

        ordemC2 = []
        for i in range(nNumeros):
            for j in invNumeros:
                ordemC2.append(j)
            invNumeros.pop(0)

        Mposicao = []
        posicaoC2 = []
        for i in range(int(nLinhas/nPalavras)):
            for j in range(nPalavras):
                posicaoC2.append(vPalavras[j])
                Mposicao.append(j)

```

```

c2 = []
c2[:] = [posicaoC2[i] for i in ordemC2]
col2.extend(c2)

scol3 = df_col3[['TITULO_PROPOSICAO','TEMA', 'PALAVRAS_CHAVE', 'TIPO_AUTOR']].values.tolist()

col3 = [c for c, i in zip(scol3, vcol3) for _ in range(i)]

# print(len(col3))
# print(len(scol3))
matriz = np.c_[col1,col2,col3]
df = pd.DataFrame(columns=["SOURCE","TARGET","TITULO_PROPOSICAO","TEMA", "PALAVRAS_CHAVE", "TIPO_AUTOR"], data=matriz)
df['SOURCE'].replace('', np.nan, inplace=True)
df.dropna(subset=['SOURCE'], inplace=True)
df['TARGET'].replace('', np.nan, inplace=True)
df.dropna(subset=['TARGET'], inplace=True)

df.to_csv(arq_semantico, index=False)

```

### 3.6 - Filtro do Grafo

```

coautoria_r = pd.read_csv("grafo-coautoria.csv")
# Len(proposicoes)

# Len(coautoria_r)
coautoria_f = coautoria_r[(coautoria_r['TIPO_AUTOR']=='Deputado')]
# Len(coautoria_f)

coautoria_info = coautoria_f.merge(usuario_proposicoes, left_on='SOURCE', right_on='NOME_AUTOR', how='inner')

# coautoria_info.shape[0]
# list(coautoria_info.columns)

keepcol = ['SOURCE', 'TARGET', 'TITULO_PROPOSICAO_x', 'TEMA_x', 'PALAVRAS_CHAVE', 'TIPO_AUTOR', 'NOME_AUTOR', 'N_PLS', 'GENERO', 'COR_RACA

```

```

coautoria_info = coautoria_info[keepcol]

coautoria_info = coautoria_info.rename(columns={'SOURCE':'source','TARGET':'target'})

coautoria_info.to_csv("grafo-coautoria.csv", index=False)
arq_semantico = "grafo-coautoria.csv"
# Len(tweets2020_info_rs)

coautoria_info_rs= coautoria_info[(coautoria_info['ESTADO_ELEICAO']=='RS')]
# Len(coautoria_info_rs)

coautoria_info_rs.to_csv("grafo-coautoria-rs.csv", index=False)
arq_semantico = "grafo-coautoria-rs.csv"

coautoria_rs = pd.read_csv("grafo-coautoria-rs.csv")
lista_rs = list(coautoria_rs['source'].unique())

coautoria_f_rs = coautoria_rs.query('target in @lista_rs')
coautoria_f_rs.to_csv("grafo-coautoria-f-rs.csv", index=False)

```

```
len(lista_rs)
```

### 3.7 - Métricas

```

grafo = 'grafo-coautoria-f-rs.csv'

df = pd.read_csv(grafo)
Graphtype = nx.Graph()
G = nx.from_pandas_edgelist(df, create_using=Graphtype)

```

```

Gcc = sorted(nx.connected_components(G), key=len, reverse=True)
G = G.subgraph(Gcc[0])

```

```

G.number_of_nodes()
G.number_of_edges()

```

```
nx.density(G)
```

```

c = list(greedy_modularity_communities(G))
# Len(c)
# c

```

```

degree_centrality = dict(nx.degree(G))
# dict(degree_centrality)
df = pd.DataFrame.from_dict(degree_centrality, orient='index')

df.index.name = 'deputado'
df = df.reset_index()
df.rename(columns = {0:'grau'},inplace = True)

df_info = pd.merge(df, usuarios_proposicoes_rs[['NOME_AUTOR','PARTIDO_ATUAL','IDEOLOGIA_PARTIDO_ATUAL','ST_REELEICAO','ESTADO_ELEICAO']], on='deputado', how='left')
df_info = df_info.drop(['NOME_AUTOR'],axis=1)

df_info = df_info.sort_values(by='grau', ascending=False)

df_info = df_info.reset_index()

df_info = df_info.drop(['index'],axis=1)
# df_info[:10]
df_info

```

```

# df = pd.DataFrame.from_dict(betweenness centrality, orient='index')

betweenness centrality = dict(nx.betweenness centrality(G,normalized=False))
# degree centrality = dict(nx.degree(G))
# dict(degree centrality)
df = pd.DataFrame.from_dict(betweenness centrality, orient='index')

df.index.name = 'deputado'
df = df.reset_index()
df.rename(columns = {'intermediacao'},inplace = True)

df_info = pd.merge(df, usuarios_proposicoes_rs[['NOME_AUTOR','PARTIDO_ATUAL','IDEOLOGIA_PARTIDO_ATUAL','ST_REELEICAO','ESTADO_ELE
df_info = df_info.drop(['NOME_AUTOR'],axis=1)

df_info = df_info.sort_values(by='intermediacao', ascending=False)

df_info = df_info.reset_index()

df_info = df_info.drop(['index'],axis=1)
# df_info[:10]
df_info

```

### 3.8 - Menção a proposições no Twitter

#### 3.9 - Filtra tweets por menções de proposições

```

pat = 'projeto de lei|requerimento de informação|projeto de decreto legislativo|projeto de lei complementar|medida provisória|pro

filtro_geral =tweets2020_info[tweets2020_info['text'].str.contains(pat)]

filtro_rs =tweets2020_info_rs[tweets2020_info_rs['text'].str.contains(pat)]

# tweets2020

```

```

# filtro_rs.to_csv("consulta_rs.csv")
# filtro_geral.to_csv("consulta_geral.csv")

```

#### 4.0 - Calcula a porcentagem para o âmbito geral

```

len(tweets2020_info)
len(filtro_geral)
uno = len(tweets2020_info)
dos = len(filtro_geral)
round(dos/uno*100,2)

```

```

len(filtro_geral['NM_URNA'].unique())

```

#### 4.5 - Calcula a porcentagem para deputados gaúchos

```
len(tweets2020_info_rs)
len(filtro_rs)
uno = len(tweets2020_info_rs)
dos = len(filtro_rs)
round(dos/uno*100,2)
```

```
len(filtro_rs['NM_URNA'].unique())
```

```
rgx = 'projeto de lei\s\d*\requerimento de informação\s\d*\indicação\s\d*\projeto de decreto legislativo\s\d*\projeto de lei comp
lista_rs = list(tweets2020_info_rs['text'].apply(unidecode).str.lower())
lista_geral = list(tweets2020_info['text'].apply(unidecode).str.lower())
```

```
# newlist = list(filter(r.match, lista_rs))
# r = re.compile(rgx2)

str_list_rs = " ".join(lista_rs)
str_list_geral = " ".join(lista_geral)

etapa_rs = re.findall(rgx,str_list_rs)
etapa_geral = re.findall(rgx,str_list_geral)

resultados_rs = [i for i in etapa_rs if any(map(str.isdigit, i))]
resultados_geral = [i for i in etapa_geral if any(map(str.isdigit, i))]
```

```
len(resultados_geral)
# resultados_geral
len(resultados_rs)
```

#### 4.6 - Gera Gráfico com menções a proposições - Geral

```
resultados_rs = [i.strip() for i in resultados_rs]
resultados_geral = [i.strip() for i in resultados_geral]

contagem_geral = Counter(resultados_geral)
contagem_geral = dict(contagem_geral)
df = pd.DataFrame.from_dict(contagem_geral, orient='index')

df = df.nlargest(6,[0])
# df.nlargest(12,df[0]).sort_values(by=df[0], ascending=False)

contagem_geral = df[0]
# contagem_geral.drop([0])
# contagem_geral.drop(contagem_geral.index[0])
# type(contagem_geral)
contagem_geral = contagem_geral.iloc[1:]
contagem_geral = contagem_geral.sort_values(ascending=True)
# contagem_geral
```

```

top = contagem_geral.plot(kind='barh',
                        colormap='tab10',
                        figsize=(11, 8),width=0.8)

for i, v in enumerate(list(contagem_geral)):
    top.text(v+1, i-0.1, v ,fontsize=15, color='black')

plt.tick_params(axis='y', which='major', labelsize=14)

# top.margins(x=1)
# top.spines['right'].set_visible(False)
# top.spines['top'].set_visible(False)
# top.spines['bottom'].set_visible(False)

# top.get_legend().remove()

# plt.rcParams.update(IPython_default);
# plt.style.use("fivethirtyeight")
plt.grid(color='b', linestyle='-', linewidth=.1)
plt.savefig("novos-graficos/Figura x - Menção a proposições no Twitter- Geral.jpg",bbox_inches = 'tight')
# plt.show()

```

#### 4.7 - Gera Gráfico com menções a proposições - Rio Grande do Sul

```

contagem_rs = Counter(resultados_rs)
contagem_rs = dict(contagem_rs)
df = pd.DataFrame.from_dict(contagem_rs, orient='index')

df = df.nlargest(7,[0])
# df.nlargest(12,df[0]).sort_values(by=df[0], ascending=False)

contagem_rs = df[0]
# contagem_rs.drop([0])
# contagem_rs.drop(contagem_rs.index[0])
# type(contagem_rs)
contagem_rs = contagem_rs.iloc[1:]
contagem_rs = contagem_rs.sort_values(ascending=True)
contagem_rs = contagem_rs.drop(['pl 22'])
contagem_rs
# contagem_rs

```

```

top = contagem_rs.plot(kind='barh',
                    colormap='tab10',
                    figsize=(11, 8),width=0.8)

for i, v in enumerate(list(contagem_rs)):
    top.text(v+0.08, i-0.1, v ,fontsize=15, color='black')

plt.tick_params(axis='y', which='major', labelsize=14)

# top.margins(x=1)
# top.spines['right'].set_visible(False)
# top.spines['top'].set_visible(False)
# top.spines['bottom'].set_visible(False)

# top.get_legend().remove()

# plt.rcParams.update(IPython_default);
# plt.style.use("fivethirtyeight")
plt.grid(color='b', linestyle='-', linewidth=.1)
plt.savefig("novos-graficos/Figura x - Menção a proposições no Twitter- Rio Grande do Sul.jpg",bbox_inches = 'tight')
# plt.show()

```

```

### Grafo de conexão do twitter filtrando por PLS - geral

```

```

from collections import Counter

```

```
filtro_geral.to_csv("filtro-geral.csv", encoding="utf-8", index=False)
filtro_rs.to_csv("filtro-rs.csv", encoding="utf-8", index=False)
```

```
list(filtro_geral.columns)
```

## 4.8 - Rede de Tweets mencionando proposições

### 4.9 - Grafo

```
%%capture
arq = "filtro-geral"

arq_json = arq + ".json"
arq_csv = arq + ".csv"
arq_grafo = arq + "-grafo.csv"
arq_rt = arq + "-grafo-rt.csv"
arq_mt = arq + "-grafo-mt.csv"

outrow = []
with open(arq_csv, "r", encoding="utf-8") as fileIn: # input file location
    with open(arq_grafo, "w") as fileOut: # output file location
        writer = csv.writer(fileOut)
        reader = csv.reader(fileIn, delimiter=',')
        writer.writerow(['source', 'target', 'text'])
        for row in reader:
            screen_name = row[0]
            text = row[1]
            target = ''
            for cell in row:
                target = re.findall(r"@([\s]+)", cell)
                target = re.sub(r"[\|\\|:|:]", '', str(target))
            outrow
            writer.writerow([screen_name, target, text])

df = pd.read_csv(arq_grafo)
df['target'].replace('', np.nan, inplace=True)
df.dropna(subset=['target'], inplace=True)
df.to_csv(arq_grafo, index=False)

df = pd.read_csv(arq_grafo)
rt = df['text'].str.startswith('RT @')
df[rt].to_csv(arq_rt, index=False)
df2 = pd.read_csv(arq_rt)
df2['target'].replace(r"[\s\S]*", "", regex=True, inplace=True)
df2.to_csv(arq_rt, index=False)
```

```
def chainer(s):
    return list(chain.from_iterable(s.str.split(',')))

lens = df['target'].str.split(',').map(len)

res = pd.DataFrame({'source': np.repeat(df['source'], lens),
                  'target': chainer(df['target']),
                  'text': np.repeat(df['text'], lens)
                  })

res['target'].replace('', np.nan, inplace=True)
res.dropna(subset=['target'], inplace=True)

mt = ~res['text'].str.startswith('RT @')
res[mt].to_csv(arq_mt, index=False)
# os.remove(arq_grafo)
```

```
# tweets2020[tweets2020['screen_name']=='lpbragancabr']
# len(filtro_geral['NM_URNA'].unique())
```

### 4.10 - Métricas

```

grafo = 'filtro-geral-grafo.csv'

df = pd.read_csv(grafo)
Graphtype = nx.DiGraph()
G = nx.from_pandas_edgelist(df, create_using=Graphtype)

# Gcc = sorted(nx.connected_components(G), key=len, reverse=True)
# G = G.subgraph(Gcc[0])
G.remove_nodes_from(list(nx.isolates(G)))

G.number_of_nodes()
G.number_of_edges()

```

```
nx.density(G)
```

```

degree_centrality = dict(G.in_degree())
# dict(degree_centrality)
df = pd.DataFrame.from_dict(degree_centrality, orient='index')

df.index.name = 'deputado'

df = df.reset_index()
df.rename(columns = {'grau'}, inplace = True)

df_info = pd.merge(df, usuarios_proposicoes[['SCREEN_NAME', 'NOME_AUTOR', 'PARTIDO_ATUAL', 'IDEOLOGIA_PARTIDO_ATUAL', 'ST_REELEICAO'],
# df_info

df_info = df_info.drop(['SCREEN_NAME'], axis=1)

df_info = df_info.sort_values(by='grau', ascending=False)
# df_info.shape[0]
df_info

```

```

df_info_rs = df_info[(df_info['ESTADO_ELEICAO']=='RS')]
df_info_rs

```

```

# df = pd.DataFrame.from_dict(betweenness_centrality, orient='index')

betweenness_centrality = dict(nx.betweenness_centrality(G, normalized=False))
# degree_centrality = dict(nx.degree(G))
# dict(degree_centrality)
df = pd.DataFrame.from_dict(betweenness_centrality, orient='index')

df.index.name = 'deputado'
df = df.reset_index()
df.rename(columns = {'intermediacao'}, inplace = True)

df_info = pd.merge(df, usuarios_proposicoes[['SCREEN_NAME', 'NOME_AUTOR', 'PARTIDO_ATUAL', 'IDEOLOGIA_PARTIDO_ATUAL', 'ST_REELEICAO'],
# df_info

df_info = df_info.drop(['SCREEN_NAME'], axis=1)

df_info = df_info.sort_values(by='intermediacao', ascending=False)
# df_info.shape[0]
df_info

```

```

df_info_rs = df_info[(df_info['ESTADO_ELEICAO']=='RS')]
df_info_rs

```

## 4.11 - Publicações mais compartilhadas

### 4.12 - Geral

```
# filtro_geral.head(2)
topTweets = filtro_geral.nlargest(3,'retweet_count')
# tweetsTop10
topIds = topTweets['id'].astype("str").values.tolist()

string = "https://twitter.com/OReillyMedia/status/"

topIds = [string + s for s in topIds]
# topIds

class Tweet(object):
    def __init__(self, s, embed_str=False):
        if not embed_str:
            # Use Twitter's oEmbed API
            # https://dev.twitter.com/web/embedded-tweets
            api = 'https://publish.twitter.com/oembed?url={}'.format(s)
            response = requests.get(api)
            self.text = response.json()["html"]
        else:
            self.text = s

    def _repr_html_(self):
        return self.text

# Tweet("https://twitter.com/OReillyMedia/status/1253710046405500928")

for x in topIds:
    Tweet(x)
# vTweets(topIds)
```

### 4.13 - Rio Grande do Sul

```
topTweets = filtro_rs.nlargest(3,'retweet_count')
# tweetsTop10
topIds = topTweets['id'].astype("str").values.tolist()

string = "https://twitter.com/OReillyMedia/status/"

topIds = [string + s for s in topIds]
# topIds

class Tweet(object):
    def __init__(self, s, embed_str=False):
        if not embed_str:
            # Use Twitter's oEmbed API
            # https://dev.twitter.com/web/embedded-tweets
            api = 'https://publish.twitter.com/oembed?url={}'.format(s)
            response = requests.get(api)
            self.text = response.json()["html"]
        else:
            self.text = s

    def _repr_html_(self):
        return self.text

# Tweet("https://twitter.com/OReillyMedia/status/1253710046405500928")

for x in topIds:
    Tweet(x)
# vTweets(topIds)
```



Pontifícia Universidade Católica do Rio Grande do Sul  
Pró-Reitoria de Graduação  
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar  
Porto Alegre - RS - Brasil  
Fone: (51) 3320-3500 - Fax: (51) 3339-1564  
E-mail: [prograd@pucrs.br](mailto:prograd@pucrs.br)  
Site: [www.pucrs.br](http://www.pucrs.br)