

ARTIGO ORIGINAL

Análise longitudinal do discurso dos políticos brasileiros: uma abordagem de tópicos

Longitudinal analysis of brazilian politicians' discourse: a topic approach

Lucas Santos de Oliveira ¹, Ronald dos Santos Matos¹, Eudes Diônatas Silva Souza¹

¹Universidade Estadual do Sudoeste da Bahia

*lsoliveira@uesb.edu.br; ronaldmatos76@gmail.com; eudes@uesb.edu.br

Recebido: 31/10/2022. Revisado: 09/07/2023. Aceito: 01/11/2023.

Resumo

A exposição de escândalos de corrupção, bem como a polarização política no Brasil desde as eleições de 2018, instalou uma aguda e profunda crise política e social no país. Esses conflitos também estiveram presentes nas redes sociais, que abriram caminho para discussões acaloradas entre políticos e o público em geral. Nesse contexto, nosso trabalho visa utilizar métodos computacionais para analisar dados de mídias sociais, a fim de identificar tópicos de mensagens políticas postadas por deputados brasileiros no período de 2013 a 2019. Para esta tarefa, adaptamos o modelo de tópico *biterm* (BTM), um algoritmo que encontra modelos de tópicos em textos curtos, para permitir a análise de tópicos ao longo do tempo. A partir desse modelo, investigamos o comportamento dos políticos nas redes sociais (Twitter), identificando os principais assuntos discutidos ao longo do tempo. Nossa abordagem divide o período estudado em segmentos anuais e compara os modelos de tópicos de diferentes intervalos de tempo, construindo um gráfico de similaridade de tópicos. No que diz respeito à investigação da evolução dos temas temporais, nossos resultados mostraram que temas relacionados à crise política e atividades relacionadas à Câmara dos Deputados foram os mais discutidos pelos deputados no período estudado.

Palavras-Chave: Brasil; Discurso; Modelo de tópico; Política; Propagação.

Abstract

The exposure of corruption scandals, as well as the political polarization in Brazil since the 2018 elections, installed a sharp and deep political and social crisis in the country. These conflicts were also present on social media, which set the stage for heated discussions between politicians and the general public. In this context, our work aims to use computational methods to analyze social media data, in order to identify topics of political messages posted by Brazilian deputies from 2013 to 2019. For this task, we adapted the biterm topic model (BTM), an algorithm that finds topic models in short texts, to enable topic analysis over time. From this model, we investigated the behavior of politicians on social networks (Twitter), identifying the main issues discussed over time. Our approach divides the studied period into annual segments and compares the topic models from different time intervals, building a topic similarity graph. Regarding the investigation of the evolution of temporal topics, our results showed that themes related to the political crisis and activities related to the Chamber of Deputies were the most discussed by deputies during the period studied.

Keywords: Brazil; Discourse; Political; Propagation; Topic model.

1 Introdução

De acordo com [Adamic and Glance \(2005\)](#), a web possibilita que os partidos políticos e a população divulguem suas opiniões de forma rápida para o público. Essa forma recente de os políticos alcançarem seus eleitores possibilitou uma maior participação da população na política, facilitando que expressem suas opiniões, participem de campanhas e manifestem-se. No entanto, também evidenciou e ampliou outros problemas como a polarização ([Recuero et al., 2020](#)), desinformação ([Reis and Benevenuto, 2021](#)) e as notícias falsas (*Fake news*) ([Bertalan and Ruiz, 2019](#); [Silva et al., 2020](#)).

Por esses motivos e pela grande quantidade de dados disponível, sociólogos e cientistas da computação veem um grande potencial nas redes sociais online para compreender a sociedade e as suas implicações na política ([Farrell, 2012](#)). Com a atual crise política do país que se iniciou em 2013 com as manifestações populares, passando pelo *impeachment* da ex-presidente Dilma Rousseff e pelas eleições gerais de 2018, um cenário de extrema polarização ideológica tem pautado o debate político e a mídia nacional ([Recuero et al., 2020](#)). Esse mesmo fenômeno também tem se refletido nas redes sociais, que se tornou palco de intensos debates políticos entre a população, políticos e partidos. Diante dessa divergência ideológica é importante entender a dinâmica de como o conteúdo é gerado e disseminado nas redes sociais, principalmente devido à crescente influência dessas plataformas na formação da opinião das pessoas.

Portanto, entender o posicionamento de parlamentares nessas redes sociais é essencial para compreender fatos recentes na sociedade e como esses posicionamentos contribuem para a formação de outras práticas na política. Nesse contexto, neste trabalho, nós propomos o uso de modelos de tópicos para identificar e analisar os principais assuntos políticos abordados pelos parlamentares brasileiros ao longo do tempo.

Para esta tarefa, nós utilizamos mais de 3 milhões de publicações no Twitter feitas por 662 deputados brasileiros entre os anos de 2013 a 2019 ([Oliveira et al., 2021](#)). Para identificar e analisar os tópicos abordados pelos políticos ao longo desses anos, nós propomos uma adaptação do modelo de tópicos curtos BTM (Biterm Topic Model) ([Yan et al., 2013](#)). Nessa adaptação, nós utilizamos uma abordagem de grafos para juntar tópicos similares e identificar como ocorre a disseminação desses tópicos longo do tempo.

Portanto, este trabalho apresenta duas contribuições principais:

- Primeiro, em NLP, a contribuição é a extensão do modelo de grafos, proposto por [Moreira \(2019\)](#), para análise temporal. No modelo original, o agrupamento é feito em períodos distintos, gerando apenas um único resultado com os tópicos ao longo dos anos, perdendo assim a evolução temporal. Em nossa abordagem, fizemos um agrupamento por período e depois criamos um grafo de similaridade temporal para identificar tópicos semelhantes em períodos distintos, possibilitando não só a visualização por período, mas a evolução desses tópicos ao longo do tempo.
- Nas ciências sociais e análise computacional do discurso

político, nossas principais contribuições foram a análise de tópicos em um grande volume de dados, de maneira longitudinal e uma quantidade significativa de políticos. Trabalhos anteriores focaram em apenas um desses aspectos em suas análises (ver [Seção 2](#)).

Este estudo está organizado da seguinte forma. A [Seção 2](#) contém alguns dos trabalhos presentes na literatura relacionados à nossa pesquisa. A [Seção 3](#) fornece uma descrição mais detalhada da metodologia e da base de dados utilizada nesse estudo. A [Seção 4](#) mostra os resultados da aplicação da metodologia no *dataset* e mostra a similaridade de tópicos políticos, descrevendo a análise da evolução dos tópicos temporais no twitter. Finalmente, a [Seção 5](#) conclui este trabalho e descreve direções futuras.

2 Trabalhos Relacionados

Com a consolidação da web como um dos principais veículos de comunicação a partir dos anos 2000, pessoas comuns, segundo [Magalhães et al. \(2013\)](#), foram transformadas em produtores de opinião e divulgação online de conteúdos da Internet, sendo a maioria desse conteúdo distribuído em diversos fóruns profissionais, redes sociais e sites em geral. Devido ao grande volume de dados textuais, detectar tópicos nessas mensagens é importante para compreender os assuntos centrais que são discutidos no mundo virtual. Um dos modelos mais usados para identificar tópicos é o LDA (*Latent Dirichlet Allocation*) proposto por [Blei et al. \(2003\)](#), cujo objetivo é identificar tópicos em documentos com grande quantidade de palavras. No entanto, de acordo com [Pereira \(2019\)](#), o LDA, assim como outros modelos tradicionais de modelagem de tópicos, quando aplicados em documentos com texto curto (*tweets* por exemplo), tendem a não funcionar da forma prevista devido a sua incapacidade para lidar com dados esparsos ([Oliveira and Vaz Melo, 2017](#)).

Existem algoritmos que exploram a natureza dos textos curtos e lida com o problema de dispersão de dados que não é contemplado pelos modelos convencionais de tópicos. Por exemplo, [Yan et al. \(2013\)](#) sugeriram um novo modelo voltado para a modelagem de textos curtos intitulado Biterm Model Topic (BTM). De acordo com o autor, o BTM implementa a modelagem de tópicos evidenciando a co-ocorrência de palavras padrões (bitermos) em um contexto curto (por exemplo, uma pequena janela de tamanho fixo em uma sequência de termos dentro de um documento), desse modo, se aperfeiçoa o aprendizado do tópico e resolve o problema de palavras esparsas no documento. Assim, o processo de análise de documentos composto em essência por textos curtos se torna mais eficiente.

Como os dados textuais gerados pelo twitter normalmente refletem as manifestações de seus usuários acerca dos acontecimentos da sociedade, é esperado que os temas gerados a partir desses dados variem com o passar do tempo. Por isso é importante analisar as mudanças dos tópicos, para entender os fatores que levam a ascensão e o declínio de assuntos dado um período de tempo. Nesse cenário, alguns estudos utilizam variantes de modelos tradicionais de modelagem de tópicos que levam em consideração o tempo como fator determinante para afe-

rir essas mudanças. Essa tarefa de descobrir padrões em informações de texto coletadas ao longo do tempo foi definida por Mei and Zhai (2005) como mineração de texto temporal (TTM).

Por exemplo, Mei and Zhai (2005) utilizaram um modelo de probabilidade geral para encontrar padrões de tema e determinar por meio de um gráfico como os tópicos variam ao longo do tempo e como seus antecessores os influenciam. Em outro estudo similar, Blei (2012) apresenta o modelo de tópico correlacionado (CTM). Ele modela explicitamente a correlação entre os assuntos da coleção e possibilita a construção de gráficos e navegadores de documentos dinâmicos que possibilitam ao usuário navegar na coleção de maneira orientada por tópicos. Com isso é possível analisar a evolução dos mesmos com o passar do tempo. Nesse modelo, os documentos são divididos por fração de tempo (ano, mês, etc) e, como resultado, os tópicos finais são sequências de distribuições sobre palavras, ao invés de uma distribuição única. Ainda nessa perspectiva, Bertalan and Ruiz (2019) aplicou vários modelos diferentes, como Indexação Semântica Latente, LDA e Processo Hierárquico de Dirichlet para encontrar tópicos que representam o discurso dos dois espectros ideológicos (direita e esquerda). Para isso, foram coletados dados de três mídias alternativas de esquerda e três de direita no Brasil.

Finalmente, Moreira (2019) propôs um método que compara tópicos em intervalos de tempo consecutivos através da construção de um gráfico de evolução e, consequentemente, detecta tópicos estáveis e temporais. Este modelo de tópicos temporais se diferencia dos demais pelo fato de comparar tópicos de todos os intervalos de tempo e medir suas similaridades.

Neste trabalho analisamos *tweets* de políticos brasileiros entre 2013 e 2019 para encontrar e observar as variações dos tópicos durante o período estudado. Utilizamos o BTM (Blei et al., 2003) para a modelagem dos tópicos nas mensagens curtas do Twitter devido a sua especificidade em lidar com esse tipo de dados. Além disso, estendemos e ampliamos a estratégia de grafos utilizado por Moreira (2019) para identificar e agrupar os tópicos ao longo dos anos, pois os métodos na literatura não possibilitam observar a evolução temporal dos tópicos.

3 Metodologia e Base de dados

Neste trabalho, utilizamos uma base de dados coletada através da API (interface de programação de aplicativos) do Twitter contendo mais de 3 milhões de *tweets* públicos de 692 deputados brasileiros de outubro de 2013 a outubro de 2019 (Oliveira et al., 2021). Para o processamento dos dados foi utilizada a linguagem de programação Python 3 com as diversas bibliotecas para análise de dados. Como recursos computacionais, foram utilizados o Google Colab na versão gratuita (Intel Xeon CPU @2.20 GHz, 13 GB RAM) e um MacBook Pro (2,8 GHz Intel Core i7 Quad-Core, 16 GB RAM). Para gerar as imagens foram utilizadas ferramentas do Microsoft Office.

Nossa metodologia está dividida em algumas etapas como mostrado na Fig. 1.

Na etapa de aquisição e limpeza de dados, os nomes dos parlamentares atuantes nesse período foram recuperados e validados por um pesquisador em 2015 e 2019 por meio

do perfil dos parlamentares no portal dados abertos da câmara federal¹. A lista das contas do Twitter associadas aos congressistas foi coletada nas páginas de perfil pessoal de cada congressista. Após esse processo, cada conta foi validada manualmente. Preparamos o texto dos *tweets* para processamento removendo *tweets* duplicados, pontuação, palavras com menos de 3 caracteres, *stopwords* em português, *hashtags*, URLs e menções. Após esse passo, nós utilizamos o *stemmer* e reduzimos cada palavra desse conjunto ao seu radical gramatical por meio da função *stemmer* presente na biblioteca *nlTK* do python e obtemos o seguinte resultado, ver Fig. 2:

Vale destacar que nem sempre esse processo resulta em palavras sintaticamente existentes na língua portuguesa, contudo isso não é um problema em se tratando de modelagem de tópicos, visto que buscamos compreender apenas o contexto.

Após esse processo de limpeza dos dados, na etapa de modelagem de tópicos, nós utilizamos o classificador automático de mensagens políticas e não-políticas desenvolvido por Oliveira et al. (2018) para classificar cada *tweet* do nosso *dataset*. Como um dos objetivos deste estudo é identificar semelhanças de tópicos políticos ao longo do tempo, utilizamos nas análises somente os *tweets* do grupo político que engloba em torno de 2,5 milhões do total coletado.

Em seguida nós agrupamos os dados ano a ano e geramos os tópicos de cada agrupamento utilizando o modelo de tópico BTM. Por meio de testes empíricos e com base nas métricas de *perplexity* e *coherence* foi definida a quantidade ideal de 20 (vinte) tópicos para cada período e mostrado as top 10 (dez) palavras mais importantes de cada tópico.

Na etapa de agregação de tópicos, ao analisar as top 10 palavras geradas, observamos que haviam tópicos semelhantes nos mesmos anos e então decidimos que seria necessário unir esses tópicos. Para isso, utilizamos o teste de Jaccard (Hamers et al., 1989), uma medida estatística usada para medir a similaridade e diversidade de conjuntos de tamanho finito, definido pela razão entre o tamanho da interseção entre dois conjuntos e pelo do tamanho da união desses mesmos conjuntos. Quanto mais próximo de 1, mais os dois conjuntos são similares; e quanto mais próximo de 0, mais diferentes eles são. Por exemplo: tomemos os dois conjuntos A e B, com os seguintes elementos: A = [câmara, plenário, votação, sessão, projeto] e B = [câmara, político, votação, Brasília, projeto]. A partir desses dois conjuntos calculamos o índice de Jaccard pela interseção entre esses dois conjuntos dividido pela união deles. Dado pela fórmula: $Jaccard(IJ) = |A \cap B| / |A \cup B|$.

No exemplo apresentado teríamos o seguinte índice de Jaccard:

$$IJ = \frac{|[camara, votacao, projeto]|}{|[camara, plenario, votacao, politico, hoje, sessao, projeto]|}$$

$$IJ = 3/7$$

$$IJ = 0,43$$

Com base no índice de Jaccard, para cada ano de 2013 a 2019, nós comparamos o índice de similaridade entre os tópicos daquele ano. Em seguida, criamos um grafo não-direcionado com todos os tópicos e conectamos os semelhantes. Cada tópico é representado por um vértice do grafo, e existe uma aresta que conecta esses tópicos

¹www.dadosabertos.com.br



Figura 1: Metodologia para criação de tópicos

	Original	nlTK_stemmer
0	amigos	amig
1	amigas	amig
2	amizade	amizad
3	carreira	carr
4	carreiras	carr

Figura 2: Palavras após aplicação do stemming

caso eles sejam semelhantes. Na Fig. 3 é mostrado o grafo de semelhança para o ano de 2013. Nele podemos observar que o tópico 1 é semelhante aos tópicos 13 e 17, do mesmo modo que o tópico 8 é similar aos tópicos 5, 14 e 15.

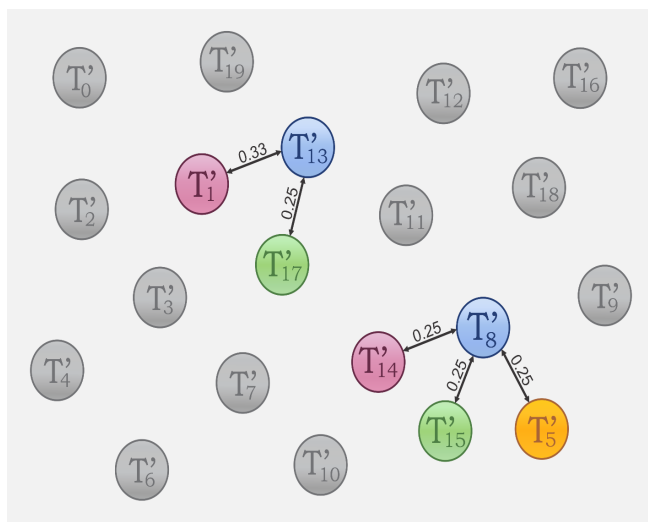


Figura 3: Grafo de semelhança de tópicos para o ano de 2013.

Neste trabalho, consideramos um tópico semelhante ao outro caso o índice de Jaccard seja maior ou igual a 0.25 (que representa $\frac{1}{4}$ da quantidade de palavras de cada tópico). Outros valores foram testados, mas 0.25 teve o resultado satisfatório para esse contexto.

Com base nesse grafo de similaridade, nós unimos os tópicos semelhantes em um único tópico. Por exemplo, o tópico 1 do ano de 2013 (T_{1_2013}) gerado pelo BTM retornou o seguinte conjunto de palavras, $T_{1_2013} = [\text{contra, governo, política, democracia, corrupção, ditadura, sempre,}$

respeito, oposição, lulopetista]. Identificamos que havia semelhança com os tópicos 13 e 17 do mesmo ano, representados por: $T_{13_2013} = [\text{direito, política, contra, partido, democracia, governo, brasil, sociedad, debat, sempre}]$ e $T_{17_2013} = [\text{brasil, política, melhor, campanha, pessoa, contra, governo, grand, brasileiro, precisa}]$. Portanto há semelhança entre esses tópicos e eles serão unidos em um único tópico resultante.

Uma importante etapa na junção de tópicos semelhantes é o recálculo da distribuição de probabilidade sobre as palavras do novo tópico. Para isso, nós utilizamos o mesmo método de [Moreira \(2019\)](#), dado pela fórmula $\sum P(w|T_i) * P(T_i)$.

Para cada tópico a ser juntado, nós somamos a multiplicação da probabilidade de cada palavra do vocabulário dado o tópico $P(w|T_i)$, pela probabilidade do tópico $P(T_i)$. A Fig. 4 ilustra os tópicos resultantes do processamento para o ano de 2013, que resultou em uma redução de 20 para 15 tópicos. De maneira semelhante, na Fig. 5, é mostrado que os tópicos de 2014 sofreram uma redução ainda maior de 20 para 12 tópicos. Então, repetimos esses processos para todos os anos desse estudo de 2013 a 2019.

Uma vez unidos os tópicos semelhantes em cada ano, nós utilizamos uma metodologia similar para identificar tópicos análogos entre os diferentes anos. Para isso, criamos um grafo direcionado com todos os tópicos entre os anos de 2013 a 2019. Conectamos os semelhantes entre os anos caso a similaridade de Jaccard seja igual ou superior a 0.25 (valores testados empiricamente). Por exemplo, o tópico 0 de 2013 (T_{0_2013}) tem uma similaridade de Jaccard igual a 0.41 com o tópico 6 de 2014 (T_{6_2014}) (ver Fig. 5). De maneira análoga aplicamos o teste de similaridade entre todos os tópicos de 2013 e 2014 e conectamos os tópicos semelhantes.

A partir dessas conexões, produzimos a Fig. 6 que demonstra as similaridades dos tópicos ao longo do tempo. Nela cada tópico é representado por um retângulo, a quantidade de tópicos referente a cada ano é igual a quantidade de retângulos pertencentes a esse ano. Cada conexão (semelhança) entre os tópicos é caracterizada por uma aresta cuja cor varia conforme o seu tópico de origem partindo do ano de 2013. A partir desse grafo de similaridade entre os anos é possível realizar uma análise longitudinal dos temas que persistem ao longo do tempo e dos temas que são específicos de um único período. Possibilitando atingir um dos objetivos desse trabalho que é a compreensão da comunicação dos políticos brasileiros no Twitter ao longo dos anos.

4 Resultados

Com base na análise dos dados coletados, identificamos que os tópicos gerados com base na metodologia apresen-

	topic0	topic1	topic2	topic3	topic4	topic5	topic6	topic7	topic8	topic9	topic10	topic11	topic12	topic13	topic14
0	domingo	brasil	freir	junto	ciudad	deputado	flávio	trabalhador	deputado	ministro	facebook	contra	amigo	contra	programa
1	alimento	política	brasil	momento	município	câmara	governo	junto	prefeito	georg	publiquei	deputado	obrigado	violência	escola
2	sábado	contra	roberto	espaço	governo	projeto	maranhão	mineiro	reunião	hilton	amigo	justiça	grand	joven	educação
3	amanhã	governo	governo	descontração	estado	comissão	sarney	obrigado	vereador	brasil	reunião	president	parabén	mulher	brasil
4	semana	direito	econômico	oportunidad	prefeito	governo	campanha	companheiro	president	esport	visita	feder	abraço	direito	professor
5	amigo	melhor	artigo	compartilhar	região	pública	pesquisa	rurai	liderança	entrevista	wagner	petrobra	semp	brasil	aluno
6	ciudad	democracia	política	estarmo	milhõ	reunião	candidato	agricultor	estadu	atleta	mont	tucano	senhor	peessoa	cultura
7	campanha	peessoa	portal	esqueça	recurso	audiência	edinho	trabalhadora	grand	brasileiro	univers	pedido	familia	familia	inscriçõ
8	agenda	semp	nacion	ligado	prefeitura	nacion	governador	grand	encontro	programa	denúncia	processo	carinho	redução	mulher
9	encontro	partido	petrobra	mídia	população	debat	oligarquia	brasil	companheiro	futebol	brasília	câmara	semana	polícia	música

Figura 4: Tópicos processados para o ano de 2013.

	topic0	topic1	topic2	topic3	topic4	topic5	topic6	topic7	topic8	topic9	topic10	topic11	topic12
0	contra	política	trabalhador	brasil	amigo	parabén	alimento	ciudad	deputado	governo	prefeito	flávio	junto
1	direito	debat	companheiro	governo	obrigado	grand	domingo	amanhã	câmara	estado	facebook	maranhão	momento
2	mulher	brasil	familiar	contra	família	brasil	polícia	grand	comissão	público	publiquei	campanha	espaço
3	violência	reforma	mineiro	freir	grand	peessoa	amigo	debat	reunião	educação	vereador	sarney	oportunidad
4	brasil	pública	rurai	impeach	senhor	trabalho	amanhã	centro	projeto	projeto	reunião	governo	descontração
5	democracia	governo	agricultor	roberto	parabén	abraço	senhor	programa	president	recurso	deputado	contra	compartilhar
6	defesa	educação	agricultura	política	igreja	obrigado	sábado	sábado	audiência	município	liderança	pesquisa	mention
7	movimento	programa	junto	corrupção	semana	semp	semana	região	ministro	milhõ	president	candidato	estarmo
8	humano	campanha	sindicato	artigo	carinho	melhor	brasil	bairro	nacion	escola	governador	mentira	twitter
9	joven	important	reunião	president	ciudad	mulher	inscriçõ	município	feder	programa	grand	edinho	precioso

Figura 5: Tópicos processados para o ano de 2014.

tada entre os anos de 2013 a 2019 se relacionam entre si e os tópicos variam ano a ano. Na Fig. 7 é mostrada a nuvem de palavras sobre os tópicos de 2013. Esses tópicos se referem, em sua maioria, a temas como: política, reunião, Brasil, governo. No entanto, há alguns outros temas menos representativos como: família, educação, mulher, parabenização, etc.

Ao analisar a propagação ao longo do tempo, observamos que alguns tópicos quando se propagam ao longo do tempo têm estruturas de bifurcação e confluência entre tópicos distintos que merecem ser melhor investigados.

Para melhor visualizar esses comportamentos, na Fig. 8, mostramos algumas estruturas selecionadas de propagações ao longo do tempo.

Nesta figura podemos observar a bifurcação do tópico 9 (2015) que possui as seguintes top 10 palavras [ciudad, campanha, amanhã, semana, amigo, sábado, deputado, encontro, domingo, grand]. Esse tópico se assemelha aos tópicos 9 e 16 de 2016 compostos respectivamente pelos conjuntos [deputado, prefeito, reunião, vereador, ciudad, president, liderança, grand, estadu, encontro], [programa, entrevista, amigo, semana, amanhã, jornal, campanha, deputado, confira, domingo].

Apesar de serem estruturalmente semelhantes eles retratam contextos diferentes, isso ocorre porque nosso algoritmo analisa apenas a semelhança entre o conjunto de

palavras e não a estrutura semântica do tópico. Contudo, para poder definir se um tópico se propaga ou não para os outros anos, tivemos que analisar semanticamente o contexto dos tópicos.

Por esse mesmo motivo, o tópico 9 de 2016, que aborda um encontro político, se propaga para o tópico 0 de 2017, que é caracterizado pelo conjunto de palavras [deputado, câmara, reunião, comissão, president, feder, projeto, prefeito, audiência, pública] e que trata da tramitação de projetos na câmara. Esses tópicos estão relacionados a um contexto maior, agenda governamental. No entanto, o tópico 16 de 2016 não se propaga para tópico 0 de 2017, porque se refere ao contexto de entrevistas, o que difere um pouco do contexto dos tópicos 9 (2016) e 0 (2017).

Como exemplo de confluência temos os exemplos dos tópicos 1 e 2 (2013) que se propaga para o mesmo tópico 3 (2014). Analisando a propagação do tópico 1 (2013), vemos que ele é composto pelo conjunto de palavras [brasil, política, contra, governo, direito, melhor, democracia, peessoa, semp, partido] e se relaciona com o tópico 3 (2014) que é composto pelo conjunto [brasil, governo, contra, freir, impeach, roberto, política, corrupção, artigo, president] que, por sua vez, se relaciona com o tópico 6 (2015) composto por [contra, política, governo, brasil, impeach, democracia, reforma, partido, corrupção, campanha] que, finalmente, se relaciona com os tópicos 6 (2016), 9 (2017),

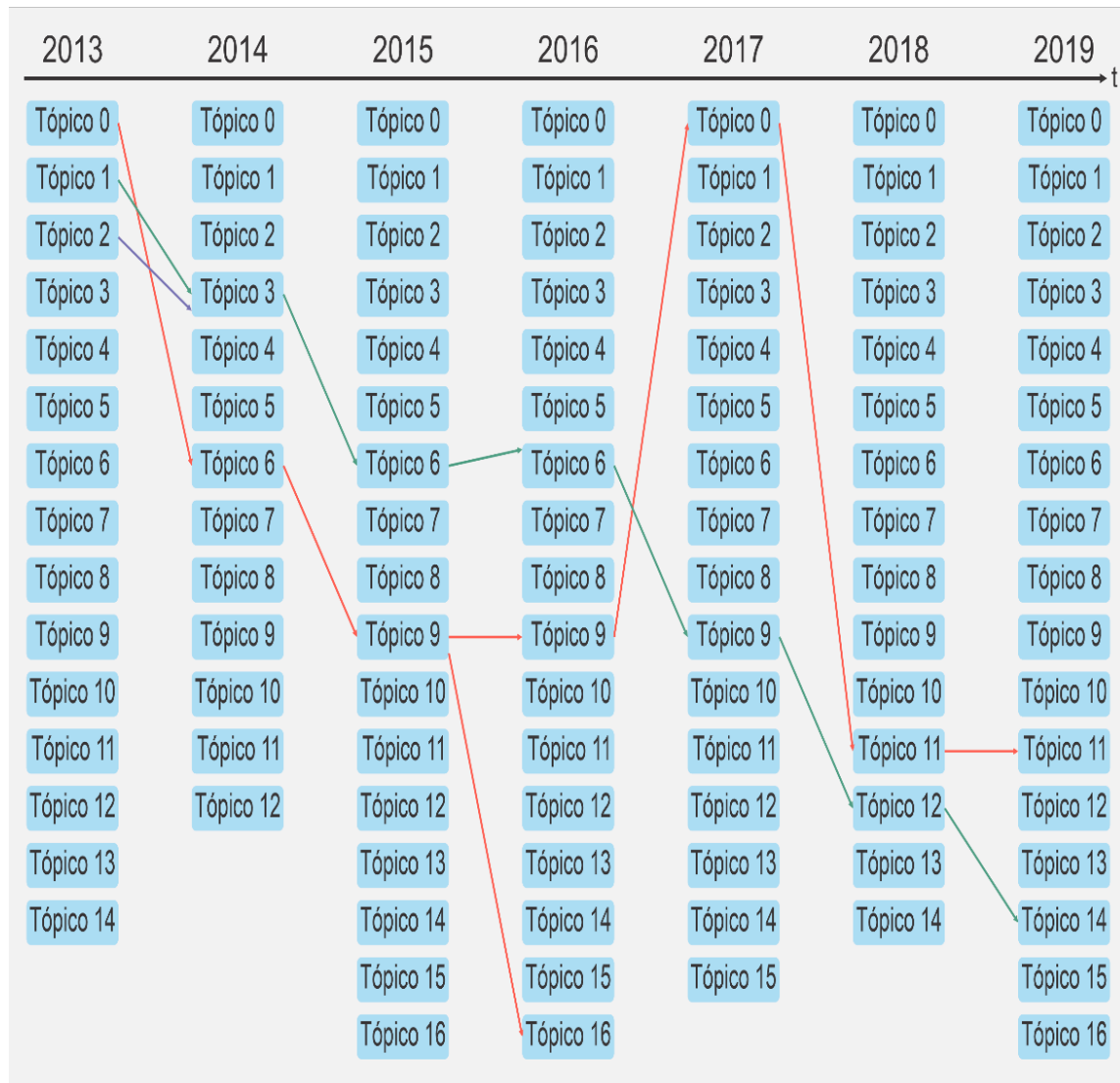


Figura 8: Tópicos ao longo do tempo (selecionados)

“Lauro Jardim: medo do impeachment é causa do repúdio a Eduardo Cunha...”

Por sua vez, o tópico 2 de 2013 composto pelas palavras [freir, brasil, roberto, governo, econômico, artigo, política, portal, nacion, petrobra] não se propaga. Isso ocorre porque os tópicos de anos posteriores não possuem uma relação semântica com esse tópico, apesar de uma similaridade estrutural pelo índice de Jaccard. A seguir temos alguns exemplos de *tweets* que remetem aos assuntos pontuais desse tópico:

“Roberto Freire no Brasil Econômico: Quando o cinismo e a inépcia atrasam o país URL.”

“...disse em função do momento nacional delicado onde a sociedade olha a política com desconfiança, não fulanizei.”

“‘Quebro a Petrobrás, mas me reelejo’, artigo de Hélio Duque URL”

Portanto, a abordagem de grafos se mostra adequada para análise temporal de tópicos. É possível observar que há tópicos recorrentes no contexto político que se propagam por vários anos, como, por exemplo, os tópicos relacionados a governo, campanha, corrupção, etc. Entretanto, há tópicos que são específicos de um determinado ano, capturando assuntos e fatos políticos ocorridos no período analisado. Por exemplo, o tópico 9 (2013) que contém as palavras [ministro, george, hilton, brasil, esporte, entrevista, atleta, brasileiro, programa, futebol] é relacionado a esporte e não se propagou para o ano de 2014. De modo semelhante os tópicos 8, 10 e 14 de 2013 não se propagaram e são temas específicos daquele ano.

5 Conclusão

A exposição de escândalos de corrupção, principalmente por meio da operação Lava-Jato e o processo de impeach-

ment da ex-presidente Dilma Rousseff em 2016, bem como a bipolarização política no Brasil desde as eleições de 2018, instalaram uma acentuada e profunda crise política e social no país. Esses conflitos também estiveram presentes nas redes sociais, que configuraram o palco para discussões acaloradas entre políticos e o público em geral. Nesse contexto, nosso trabalho teve como objetivo desenvolver métodos computacionais para modelar tópicos de mensagens políticas postadas por políticos brasileiros entre os anos de 2013 a 2019.

Nosso intuito é entender o discurso dos políticos nas redes sociais (Twitter). Para isso, apresentamos um método para investigar a evolução de tópicos, analisando os principais assuntos debatidos ao longo do tempo. Nossa abordagem divide o período estudado em segmentos anuais e compara a semelhança dos tópicos a partir de um grafo de similaridade. Esse grafo possibilita a identificação de tópicos correlatos em um mesmo ano e em anos diferentes. Em relação a investigação da evolução de temas políticos temporais, nossos resultados mostraram que temas relativos à crise política e as atividades relacionadas à câmara dos deputados foram os mais relevantes e discutidos pelos representantes durante o período estudado.

Como trabalho futuro, pretendemos aprimorar a metodologia e estender o estudo com uma investigação mais abrangente da evolução de tópicos políticos, acrescentando mais dados e diminuindo a granularidade temporal para 3 ou 6 meses. Esperamos com isso obter uma visão mais ampla do contexto e identificar elementos chave do comportamento dos parlamentares brasileiros nas redes sociais, para assim obtermos uma compreensão mais completa da dinâmica por trás do processo de propagação desses tópicos ao longo do tempo.

Referências

- Adamic, L. A. and Glance, N. (2005). The political blogosphere and the 2004 us election: divided they blog, *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43. <https://doi.org/10.1145/1134271.1134277>.
- Bertalan, V. G. and Ruiz, E. E. S. (2019). Using topic modeling to find main discussion topics in brazilian political websites, *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web, WebMedia '19*, Association for Computing Machinery, New York, NY, USA, p. 245–248. <https://doi.org/10.1145/3323503.3360644>.
- Blei, D. M. (2012). Probabilistic topic models, *Communications of the ACM* 55(4): 77–84. <https://doi.org/10.1145/2133806.2133826>.
- Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation, *Journal of machine Learning research* 3(Jan): 993–1022. Disponível em <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>.
- Farrell, H. (2012). The consequences of the internet for politics, *Annual review of political science* 15: 35–52. <https://doi.org/10.1146/annurev-polisci-030810-110815>.
- Hamers, L. et al. (1989). Similarity measures in scientometric research: The jaccard index versus salton's cosine formula, *Information Processing and Management* 25(3): 315–18. [https://www.doi.org/10.1016/0306-4573\(89\)90048-4](https://www.doi.org/10.1016/0306-4573(89)90048-4).
- Magalhães, C. M., Araújo, C. G., de Oliveira, E. T. and de Souza, H. L. (2013). O boom da internet e o impacto na relação entre empresas e consumidores, *Bauru: XVIII Congresso de Ciências da Comunicação na Região Sudeste*. Disponível em <https://portalintercom.org.br/anais/sudeste2013/resumos/R38-0315-1.pdf>.
- Mei, Q. and Zhai, C. (2005). Discovering evolutionary theme patterns from text: an exploration of temporal text mining, *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 198–207. <https://doi.org/10.1145/1081870.1081895>.
- Moreira, R. C. N. (2019). *Polarização política e o impeachment de 2016: uma análise de dados reais e de mídias sociais*, Dissertação de mestrado em ciência da computação, Universidade Federal de Minas Gerais, Belo Horizonte. Disponível em <https://repositorio.ufmg.br/bitstream/1843/30402/1/RobertaCoeli.pdf>.
- Oliveira, L. S. d., Amaral, M. S. and Vaz-de Melo, P. O. (2021). Long-term characterization of political communications on social media, *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 95–102. <https://doi.org/10.1145/3486622.3493934>.
- Oliveira, L. S. D. and Vaz Melo, P. O. (2017). How to Find the Relevant Words Politicians Use in Twitter?, *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web - WebMedia '17*, ACM Press, New York, New York, USA, pp. 465–468. <https://doi.org/10.1145/3126858.3131590>.
- Oliveira, L. S., de Melo, P. O. V., Amaral, M. S. and Pinho, J. A. G. (2018). When politicians talk about politics: Identifying political tweets of brazilian congressmen, *Twelfth International AAAI Conference on Web and Social Media*. <https://doi.org/10.48550/arXiv.1805.01589>.
- Pereira, M. (2019). *Análise exploratória de tweets utilizando modelagem de tópicos para textos curtos: caso olimpíadas rio 2016*, Tcc de graduação em ciência da computação, Universidade Federal da Fronteira Sul. Disponível em <https://rd.ufes.edu.br/bitstream/prefix/3371/1/PEREIRA.pdf>.
- Recuero, R., Soares, F. B. and Gruz, A. (2020). Hyper-partisanship, Disinformation and Political Conversations on Twitter: The Brazilian Presidential Election of 2018, *Ted Rogers School of Management, 3 Social Media Lab* 1(2): 569–578. Disponível em <https://ojs.aaai.org/index.php/ICWSM/article/view/7324/7178>.
- Reis, J. C. and Benevenuto, F. (2021). Supervised learning for misinformation detection in whatsapp, *Proceedings of the Brazilian Symposium on Multimedia and the Web*, pp. 245–252. <https://doi.org/10.1145/3470482.3479641>.

- Silva, M., Santos de Oliveira, L., Andreou, A., Vaz de Melo, P. O., Goga, O. and Benevenuto, F. (2020). Facebook ads monitor: An independent auditing system for political ads on facebook, *Proceedings of The Web Conference 2020*, pp. 224–234. <https://doi.org/10.1145/3366423.3380109>.
- Yan, X., Guo, J., Lan, Y. and Cheng, X. (2013). A biterm topic model for short texts, *Proceedings of the 22nd international conference on World Wide Web*, pp. 1445–1456. <https://doi.org/10.1145/2488388.2488514>.