

Ativismo da extrema direita brasileira no WhatsApp: O que mudou das eleições de 2018 para 2022?

Thiago Gadelha¹, José Maria Monteiro¹, Javam Machado¹,
Ivandro Claudino¹, Rangel Santos¹, Luciano Galick¹, Carlos Santos¹

¹Universidade Federal do Ceará (UFC)

Av. Humberto Monte, s/n, Pici - CEP 60440-593 – Fortaleza – CE – Brasil

{thiago.gadelha, jose.monteiro, javam.machado}@lsbd.ufc.br

{desapb, rangel.jungles, lgalicki, carlospsantos27}@gmail.com

Abstract. *The purpose of this article is to analyze the dynamics of public WhatsApp groups used by the Brazilian far-right in two distinct moments: the electoral campaigns of 2018 and 2022. To do this, we built two datasets for these periods, containing the messages that circulated in the observed groups. Next, we used the LDA algorithm to uncover the latent topics in each of the analyzed datasets. Subsequently, we utilized the LIWC tool to analyze the emotional, cognitive, and structural components of the texts that compose each dataset. The results point to significant changes in the dynamics of the groups used in 2018 compared to those used in 2022. Out of the 64 LIWC categories, 57 presented significant differences between 2018 and 2022. LDA showed that in 2022 there was a significant increase in questioning about electoral fraud.*

Resumo. *O presente artigo tem por finalidade analisar a dinâmica dos grupos públicos de WhatsApp utilizados pela extrema direita brasileira em dois momentos distintos: as campanhas eleitorais de 2018 e 2022. Para isso, construímos dois conjuntos de dados, referentes aos períodos estudados, contendo as mensagens que circularam nos grupos observados. Em seguida, aplicamos o algoritmo LDA para descobrir os tópicos latentes em cada um dos conjuntos de dados analisados. Posteriormente, utilizamos a ferramenta LIWC para analisar os componentes emocionais, cognitivos e estruturais dos textos que compõem cada conjunto de dados. Os resultados apontam mudanças significativas na dinâmica dos grupos estudados. Das 64 categorias do LIWC, 57 apresentaram diferenças significativas entre 2018 e 2022. O LDA mostrou que em 2022 ocorreu um aumento relevante nos questionamentos sobre fraude eleitoral.*

1. Introdução

Nos últimos anos, temos observado um espetacular crescimento político e eleitoral de forças de extrema direita no Brasil. A extrema direita é o posicionamento político cujas principais ideais são nacionalismo, ultraconservadorismo e extremismo, as quais, em geral, levam às percepções de superioridade, preconceito e xenofobia [Löwy 2015]. Neste contexto, a campanha de 2018 pode ser considerada um marco. Particularmente, pela vitória de Jair Bolsonaro, cuja campanha estruturou-se por meio de uma rede de comunicação digital, que teve o WhatsApp como um dos principais meios de disseminação de conteúdo.

Neste trabalho, abordamos o desafio de compreender a dinâmica dos grupos públicos de WhatsApp utilizados pela extrema direita brasileira, especificamente, de apoio ao presidente Jair Bolsonaro, em dois momentos distintos: as campanhas eleitorais de 2018 e 2022. Procuramos responder à seguinte questão de pesquisa: Quais as principais mudanças nas dinâmicas dos grupos públicos de WhatsApp, empregados pela extrema direita brasileira, das eleições de 2018 para 2022?

Para responder esta questão de pesquisa, construímos dois conjuntos de dados, referentes aos períodos investigados, contendo todas as mensagens que circularam nos grupos observados. Esses conjuntos de dados foram denominados ERB2018 e ERB2022, respectivamente. Em seguida, aplicamos o algoritmo *Latent Dirichlet Allocation* (LDA) [Blei et al. 2003] para descobrir os tópicos latentes em cada um dos conjuntos de dados analisados. Posteriormente, utilizamos a ferramenta LIWC (*Linguistic Inquiry and Word Count*) [Tausczik and Pennebaker 2010] para analisar os componentes emocionais, cognitivos e estruturais dos textos que compõem cada conjunto de dados. Os resultados apontam mudanças significativas na dinâmica dos grupos utilizados em 2018 em relação àqueles utilizados em 2022. Por exemplo, das 64 categorias do LIWC, 57 apresentaram diferenças significativas entre 2018 e 2022. O LDA mostrou que em 2018 houve uma ênfase nas questões eleitorais, enquanto que 2022 ocorreu um aumento significativo nos questionamentos sobre fraude eleitoral e pedidos de intervenção militar.

O restante deste artigo está organizado da seguinte forma. Na Seção 2 são discutidos os principais trabalhos relacionados. A Seção 3 descreve os conjuntos de dados ERB2018 e ERB2022. Na Seção 4 apresentamos as análises realizadas a fim de responder a questão de pesquisa. Por fim, na Seção 5 são apresentadas as conclusões obtidas.

2. Trabalhos Relacionados

Em [Ebeling et al. 2020], os autores apresentam um *framework* para analisar como a polarização política afeta o comportamento de grupos com posicionamentos opostos no cenário brasileiro da COVID. Os posicionamentos pró/contra o isolamento social são representados por dois grupos no Twitter denominados *Cloroquiners* e *Quarenteners*, e seus comportamentos foram comparados a um grupo sem motivações políticas. Já em [de Sousa and Becker 2021], os autores realizaram uma análise temporal com a finalidade de compreender os argumentos utilizados por brasileiros no Twitter para defender posicionamentos pró e contra a vacinação da COVID-19. Um estudo com o objetivo de caracterizar o tipo de conteúdo que foi disseminado durante a pandemia de COVID-19, a partir dos *links* do YouTube que foram compartilhadas em grupos públicos do WhatsApp e do Telegram no Brasil, focados em discussões políticas, foi apresentado em [Pereira et al. 2022].

Uma metodologia chamada LOCPOC, que visa caracterizar os padrões de comunicação dos políticos brasileiros ao longo dos anos com base na quantidade de mensagens políticas e não políticas que eles postam, foi proposta em [Oliveira and Melo 2022]. Em [de Oliveira et al. 2022], os autores apresentam uma análise multiplataforma da comunicação dos políticos brasileiros no Facebook, Instagram e Twitter. Eles quantificaram as postagens desses políticos, o quanto eles replicam conteúdo e o nível de engajamento. Como resultado, observou-se que a maioria dos políticos posta nas três redes sociais, há diversidade na escolha estratégica das principais mídias e que existem muitos deles que apenas replicam conteúdo nas três redes.

3. Os Conjuntos de Dados ERB2018 e ERB2022

Com o objetivo de analisar a dinâmica dos grupos públicos de WhatsApp utilizados pela extrema direita brasileira nas eleições de 2018 e 2022, construímos dois conjuntos de dados, ERB2018 e ERB2022, respectivamente. Esses conjuntos de dados incorporam todas as mensagens que circularam nos grupos observados. O ERB2018 foi construído a partir do conjunto de dados disponibilizado por [Cabral et al. 2021]. Este conjunto de dados foi coletado no período de 02 de julho de 2018 a 29 de outubro de 2018, que compreende o primeiro e segundo turno das eleições gerais de 2018. O ERB2018 contém 282.601 mensagens obtidas de 5.369 usuários de 59 grupos. Já o ERB2022 foi construído utilizando a plataforma proposta em [de Sá et al. 2021] no período de 29 de setembro de 2022 a 14 de novembro de 2022, que compreende parte do 1º turno e todo o 2º turno das eleições gerais de 2022. O ERB2022 contém 598.971 mensagens obtidas de 15.286 usuários de 240 grupos.

Na Tabela 1, apresentamos estatísticas básicas computadas sobre os conjuntos de dados ERB2018 e ERB2022, incluindo quantidade de mensagens únicas, média e desvio padrão do número de *tokens*, número mínimo e máximo de *tokens*, além da média e desvio padrão da quantidade de compartilhamentos. A partir da Tabela 1, percebe-se que as características das mensagens parecem não mudar substancialmente. Porém, a média de compartilhamentos é bem maior em 2022. De fato, em 2018 os grupos apresentavam uma discussão maior em torno das postagens. Já em 2022, os grupos funcionaram mais com repositórios de conteúdos a serem disseminados.

Tabela 1. Estatísticas Básicas dos Conjuntos de Dados ERB2018 e ERB2022

Estatística	2018	2022
Quantidade de mensagens únicas	114.444	161.245
Média e desvio-padrão do número de tokens	28,55 ±98, 27	29,67 ±63.03
Número mínimo de tokens	1	1
Número máximo de tokens	7.219	8.749
Média e desvio-padrão dos compartilhamentos	2,36 ±6, 83	7,81 ±34, 7

4. Análise das Mensagens

Neste seção, iremos analisar as principais mudanças nas dinâmicas dos grupos públicos de WhatsApp, empregados pela extrema direita brasileira, das eleições de 2018 para 2022. Para isso iremos explorar dois métodos distintos: LDA e LIWC.

4.1. Análise de Tópicos

Para analisar as mensagens compartilhadas nas eleições de 2018 e 2022, utilizamos uma abordagem de sumarização de tópicos. Os tópicos são formados por agrupamentos de palavras, que frequentemente ocorrem nas postagem que compõem um determinado conjunto de dados. A fim de identificar o melhor número de tópicos (k), usamos a métrica de pontuação de coerência e variamos k de 2 a 30, com o passo igual 2, em cada um dos conjuntos de dados. A partir desse procedimento, encontramos $k = 10$ tanto para o ERB2018 quanto para o ERB2022. Em seguida, aplicamos o algoritmo LDA, usando $k = 10$, em cada um dos conjuntos de dados.

Por fim, realizamos uma análise qualitativa para inferir quais os temas abordados e as dinâmicas sociais latentes nos dois períodos investigados. Os tópicos mais importantes e seus termos mais representativos estão na Tabela 2. Ao compararmos os tópicos (*clusters*) encontrados para 2018 e 2022, notáveis diferenças em vários aspectos são observadas. Em 2018, houve uma maior ênfase em textos relacionados ao debate eleitoral, campanha e divulgação. Por outro lado, em 2022, ocorreu um aumento significativo nos questionamentos acerca de fraude eleitoral, discursos inflamados sobre os candidatos e uma manifesta convocação à participação em protestos, incluindo o apelo ao apoio dos militares. Além disso, algumas temáticas são semelhantes nos dois anos analisados, como a utilização de apelos religiosos, o envolvimento das forças armadas e a disseminação de notícias por meio de grupos de discussão política.

Tabela 2. Tópicos gerados pelo LDA para as Eleições de 2018 e 2022

Tópico	Termos mais representativos (2018)	Tópico	Termos mais representativos (2022)
1	brasil; governo; contra; país; família; esquerda; partido; direita; militar; bolsonaro; pt; brasileiro; crianças; mulheres; políticos; poder; comunista; brasileira; partidos; política; dinheiro; venezuela	1	lula; tse; stf; armadas; urnas; forças; fraude; bolsonaro; intervenção; relatório; brasil; presidente; poder; eleições; sistema; militar; ffaa; eleitoral censura; ministros
3	bolsonaro; todos; turno; brasil; voto; pessoal; campanha; candidato; povo; eleição; segundo; primeiro; votar; jair; esquerda; nordeste; apoio; ganhar; eleitores; vitória; poder	3	deus; bom; senhor; vida; notícia; vídeo; jesus; sempre; melhor; casa
5	whatsapp; chat; bolsonaro; haddad; pág; tocantins; lula; carreata; link; grupos; pública; facebook; fernando; direita; recebido; fazenda;	4	brasil; povo; todos; agora; ruas; esquerda; país; quartéis; manifestações; liberdade; brasileiro; bem; ordem; comunismo; mundo; frente; podemos; exército
7	facebook; votar; b17; milhões; bolsonaro; brasil; ps1; urnas; votos; atenção; eleitoral; instagram; bolsonaro17; candidatos; tse; vote; candidato; fraude	9	lula; instagram; fazendo; terra brasil notícias; twitter; covardia; pronuncie; ladrão; faz; l; petista; jornal da cidade online; bandido; poderaopovo
8	pt; haddad; bolsonaro; globo; votos; pesquisa; fake; pesquisas; maior; ibope; eleições; urgente; rede; resultado	10	janja; robô; pt; milhões; ministério; governo; lula; ex; preso; dinheiro; campanha; bandidos; dilma; bilhões; secretaria; fraudes; corrupção; escândalo; presidência; brasil; presidiário

4.2. Análise Linguística

A fim de analisar os componentes emocionais, cognitivos e estruturais dos textos que compõem cada conjunto de dados utilizamos a ferramenta LIWC, a qual é capaz de extrair características psicolinguísticas de textos com base em léxicos. O LIWC utiliza um dicionário de palavras desenvolvido para o português do Brasil (LIWC2015-BR) [Carvalho et al. 2019], o qual possui 127.149 palavras, em que cada uma pode ser assinalada a uma ou mais categorias. Essas categorias representam perspectivas linguísticas, psicológicas, dentre outras. O dicionário é composto por um total de 64 categorias.

Levando em consideração todas as 64 categorias do LIWC, 57 apresentam diferenças significativas entre 2018 e 2022. A Tabela 3 apresenta os resultados das principais categorias do LIWC relacionadas a aspectos psicológicos. Para analisar a significância da diferença de percentuais de uso entre os dois períodos estudados, foi executado

o teste estatístico não paramétrico qui-quadrado com nível de significância de 95%. Os resultados que tiveram uma diferença estatística significativa estão destacados em negrito. Investigamos cada período eleitoral de acordo com os seguintes aspectos psicológicos, atrelados a sua relação com as categorias linguísticas:

- **Coesão e união:** A utilização de palavras da categoria “concordância” denota consenso. Logo, os grupos utilizados em 2018 apresentaram maior coesão e união.
- **Estados Afetivos:** Note que raiva e tristeza foram emoções mais fortes em 2018, onde tivemos discursos de ódio mais intensos. Já em 2022 observamos uma predominância de emoções negativas, provavelmente devido à possibilidade de derrota.
- **Sofisticação cognitiva:** A categoria mecanismos cognitivos apresenta um valor maior em 2022, isso significa uma elaboração dos textos. Adicionalmente, a categoria certeza diminuiu de 2018 para 2022, o que pode estar relacionados com argumentos em defesa do governo vigente.
- **Preocupações pessoais:** Notamos um apelo maior à religião em 2022. Enquanto mensagens de teor sexual diminuíram de 2018 para 2022. As preocupações relacionadas à saúde e morte são maiores em 2018. Seria natural esperar que essas preocupações fossem maiores em 2022, devido à pandemia de COVID19. Porém, as mensagens de 2022 buscam defender as ações do governo durante a crise sanitária, o que contribui para diminuir esses valores.

Tabela 3. Percentuais de categorias LIWC para as Eleições de 2018 e 2022

Dimensão	Categoria	2018	2022	Dimensão	Categoria	2018	2022
Senso de Grupo	nós	0,095	0,093	Emoções	raiva	0,478	0,469
	concordância	0,084	0,074		tristeza	0,325	0,300
Preocupações Pessoais	religião	0,289	0,334	ansiedade	0,171	0,006	
	dinheiro	0,536	0,512	em. negativa	1,042	1,055	
	lazer	0,342	0,491	em. positiva	1,412	1,419	
	sexo	0,229	0,212	Complexidade Cognitiva	exclusividade	1,541	1,461
saúde	0,189	0,185	conjunção		2,062	2,040	
morte	0,081	0,073	certeza		0,370	0,346	
			mec. cognitivos		8,815	8,892	

5. Conclusões e Trabalhos Futuros

Neste trabalho, buscamos responder à seguinte questão de pesquisa: Quais as principais mudanças nas dinâmicas dos grupos públicos de WhatsApp, empregados pela extrema direita brasileira, das eleições de 2018 para 2022? Os resultados apontam mudanças significativas na dinâmica dos grupos utilizados em 2018 em relação àqueles utilizados em 2022. Por exemplo, das 64 categorias do LIWC, 57 apresentaram diferenças significativas entre 2018 e 2022. As emoções ligadas a raiva e tristeza diminuíram de 2018 para 2022. Já o apelo à religião aumentou em 2022, assim como a sofisticação das mensagens. A concordância também diminuiu de 2018 para 2022, indicando uma redução na coesão e união. O algoritmo LDA mostrou que em 2018 houve uma ênfase nas questões eleitorais, enquanto que 2022 ocorreu um aumento significativo nos questionamentos sobre fraude eleitoral e pedidos de intervenção militar.

Referências

- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(null):993–1022.
- Cabral, L., Monteiro, J. M., da Silva, J. W. F., Mattos, C. L. C., and Mourão, P. J. C. (2021). Fakewhastapp.br: NLP and machine learning techniques for misinformation detection in brazilian portuguese whatsapp messages. In Filipe, J., Smialek, M., Brodsky, A., and Hammoudi, S., editors, *Proceedings of the 23rd International Conference on Enterprise Information Systems, ICEIS 2021, Online Streaming, April 26-28, 2021, Volume 1*, pages 63–74. SCITEPRESS.
- Carvalho, F., Rodrigues, R., Santos, G., Cruz, P., Ferrari, L., and Guedes, G. (2019). Avaliação da versão em português do liwc lexicon 2015 com análise de sentimentos em redes sociais. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 24–34, Porto Alegre, RS, Brasil. SBC.
- de Oliveira, L. S., Costa, W., de Melo, P. O. S. V., and Benevenuto, F. (2022). How politicians communicate in social media: A cross-platform study. In Silva, T. H., Dorini, L. B., Almeida, J. M., and Marques-Neto, H. T., editors, *WebMedia '22: Brazilian Symposium on Multimedia and Web, Curitiba, Brazil, November 7 - 11, 2022*, pages 75–83. ACM.
- de Sá, I. C., Monteiro, J. M., da Silva, J. W. F., Medeiros, L. M., Mourão, P. J. C., and da Cunha, L. C. C. (2021). Digital lighthouse: A platform for monitoring public groups in whatsapp. In Filipe, J., Smialek, M., Brodsky, A., and Hammoudi, S., editors, *Proceedings of the 23rd International Conference on Enterprise Information Systems, ICEIS 2021, Online Streaming, April 26-28, 2021, Volume 1*, pages 297–304. SCITEPRESS.
- de Sousa, A. and Becker, K. (2021). Pro/anti-vaxxers in brazil: a temporal analysis of covid vaccination stance in twitter. In *Anais do IX Symposium on Knowledge Discovery, Mining and Learning*, pages 105–112, Porto Alegre, RS, Brasil. SBC.
- Ebeling, R., Sáenz, C., Nobre, J., and Becker, K. (2020). Quarenteners vs. cloroquiners: a framework to analyze the effect of political polarization on social distance stances. In *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*, pages 89–96, Porto Alegre, RS, Brasil. SBC.
- Löwy, M. (2015). Conservadorismo e extrema-direita na europa e no brasil. *Serviço Social & Sociedade*, pages 652–664.
- Oliveira, L. and Melo, P. (2022). Large-scale and long-term characterization of political communications on social media. In *Anais Estendidos do XXVIII Simpósio Brasileiro de Sistemas Multimídia e Web*, pages 31–34, Porto Alegre, RS, Brasil. SBC.
- Pereira, É. V., Melo, P. F., Júnior, M., Mafra, V. O., Reis, J. C. S., and Benevenuto, F. (2022). Analyzing youtube videos shared on whatsapp and telegram political public groups. In Silva, T. H., Dorini, L. B., Almeida, J. M., and Marques-Neto, H. T., editors, *WebMedia '22: Brazilian Symposium on Multimedia and Web, Curitiba, Brazil, November 7 - 11, 2022*, pages 28–37. ACM.
- Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.