

Central de Fatos: Um Repositório de Checagens de Fatos

João M. M. Couto¹, Breno Pimenta¹, Igor M. de Araújo¹, Samuel Assis¹,
Julio C. S. Reis², Ana Paula C. da Silva¹, Jussara M. Almeida¹, Fabrício Benevenuto¹

¹Universidade Federal de Minas Gerais (UFMG) – Brasil

²Universidade Federal de Viçosa (UFV) – Brasil

{joaocouto,brenopimenta,igorma,samuellassis}@dcc.ufmg.br

jreis@ufv.br, {ana.coutosilva,jussara,fabricio}@dcc.ufmg.br

Abstract. *In recent times, the research interest in dissecting the dissemination and prevention of misinformation in the online environment has spiked dramatically. Given that scenario, a recurring obstacle is the unavailability of public datasets containing fact-checked instances. In this work, we performed an extensive data collection of such instances from the most important Brazilian fact-checking agencies. Particularly, this paper offers to the research community a novel dataset containing fact-checks from various trustworthy sources regarding a wide range of topics. In total, the resulting collection encompasses 11 647 fact-check instances collected across 6 different agencies that can be used to drive future efforts towards identifying and fighting misinformation on digital platforms in Brazil.*

Resumo. *Recentemente, o interesse por frentes de pesquisa analisando os mecanismos, bem como maneiras de evitar a disseminação de desinformação aumentou significativamente. Neste cenário, um recorrente obstáculo é a indisponibilidade de checagens de fatos. Neste trabalho, compilamos uma extensa coleção de checagens oriundas de importantes agências de checagem de fatos brasileiras. Oferecemos à comunidade científica uma coleção inédita contendo checagens de diversas fontes confiáveis que abrangem um largo espectro de tópicos. Ao todo, a coleção resultante engloba 11 647 instâncias de checagem de fatos coletadas em 6 agências diferentes que podem ser utilizadas em diversos estudos nos contextos de identificação e combate à desinformação em plataformas digitais no Brasil.*

1. Introdução

Nos últimos anos, plataformas digitais tornaram-se poderosos meios de compartilhamento de informações em larga escala no mundo e, em particular, no Brasil [Newman et al. 2019]. Apesar de promover meios eficientes para divulgação de informações de interesse público (por exemplo, divulgação científica, comunicação governamental), observamos um cenário preocupante no qual entidades com agendas bem definidas abusam dessa liberdade, promovendo a propagação em massa de conteúdos de desinformação [Resende et al. 2019]. Esses conteúdos muitas vezes visam influenciar a opinião pública no desenrolar de processos democráticos [Bessi and Ferrara 2016, Gomes Jr and Frizzon 2019, Machado et al. 2019], na promoção de movimentos sociais [Arun 2019] ou na adoção de comportamentos nocivos à saúde [Ferrara 2020,

Martins et al. 2021], possivelmente gerando novos caminhos de radicalização e intensificando conflitos sociais.

Em face a esse desafio, existem diferentes iniciativas que visam o combate à disseminação de desinformação [Myslinski 2012, Vlachos and Riedel 2014, Wu et al. 2014, Reis et al. 2019]. Entre elas, destacam-se as agências de checagem de fatos, cuja missão é prover uma fonte confiável e contínua de publicações descredenciando a maior quantidade possível de conteúdos de desinformação [Myslinski 2012, Vlachos and Riedel 2014, Ciampaglia et al. 2015]. Exemplos incluem “Snopes.com”¹, “PolitiFact”², “FactCheck.org”³, no âmbito internacional, e “Aos fatos”⁴, “Lupa”⁵, “Boatos.org”⁶ e “Projeto Comprova”⁷, no Brasil. Esses conteúdos são compartilhados em um desafiador e largo espectro de formatos: notícias, vídeos, ou até mesmo memes. Aqui, temos um cenário temeroso: a frequência com a qual publicações de baixa credibilidade são geradas supera, com uma extensa margem, a frequência com que verificações das agências de checagem são produzidas. A consequência disso é clara: existe um volume muito maior de conteúdos falsos do que verificações que os desmintam [Ciampaglia et al. 2015]. Na dura realidade deste fato, faz-se de suma importância a exploração do problema a fim de buscar novos entendimentos e soluções para a disseminação de desinformação.

Assim, as checagens de fatos realizadas por agências com alta credibilidade permitem o melhor entendimento das diferentes facetas que impulsionam a difusão de conteúdos de desinformação. Compreender esta dinâmica é de fundamental importância para mitigar os danos causados por este conteúdo. Um exemplo atual da importância de mecanismos que identifiquem conteúdos de desinformação é a pandemia da COVID-19. Neste contexto, a propagação de desinformação, seja com intuito de promover remédios sem eficácia comprovada, ridicularizar comportamentos (uso de máscaras e distanciamento social) ou desacreditar a eficácia de vacinas, pode ocasionar em perdas de vidas [Galhardi et al. 2020]. No entanto, as checagens realizadas por estas agências não estão centralizadas em um único repositório. Além disso, as poucas estratégias de unificação desses dados são focadas em assuntos específicos (i.e, saúde⁸), com pouca representatividade do contexto geral brasileiro. A existência de um repositório unificado de dados de alta qualidade que tenham sido rotulados por anotadores especialistas (i.e., jornalistas) com experiência de domínio [Reis et al. 2020] e que abranjam diferentes domínios (i.e., saúde, política, etc) e assuntos (i.e., COVID-19, eleições municipais, etc) é primordial para o desenvolvimento de estudos que visam compreender e combater a difusão em larga escala de conteúdos de desinformação no contexto brasileiro.

Neste trabalho, apresentamos uma metodologia de coleta e unificação de conteúdos provenientes de 6 agências brasileiras de checagem de fatos com reconhecimento e abrangência nacional, e algumas delas, inclusive, signatárias do código de

¹www.snopes.com

²www.politifact.com/

³www.factcheck.org/

⁴aosfatos.org

⁵piaui.folha.uol.com.br/lupa/

⁶www.boatos.org

⁷projeto.comprova.com.br/

⁸<https://www.poynter.org/coronavirusfactsalliance/>

conduta da IFCN⁹. O repositório, *Central de Fatos*, gerado a partir da unificação dos dados destas diferentes fontes possui 11 647 checagens de fatos, cobrindo assuntos diversos (e.g., saúde, política), realizadas no período entre 2013 e maio de 2021. Este repositório está disponibilizado publicamente em: <https://doi.org/10.5281/zenodo.5191798>. Acreditamos que estes dados poderão auxiliar em futuras pesquisas relacionadas à caracterização e detecção automática de desinformação no cenário nacional, visando combater, de forma eficiente, a difusão deste tipo de informação nociva à sociedade.

O restante do trabalho está organizado da seguinte forma. Na Seção 2 apresentamos a metodologia para construção do repositório e descrevemos algumas limitações dos dados coletados. Em seguida, a Seção 3 apresenta a descrição geral do conjunto de dados coletado. A Seção 4 discute algumas potenciais aplicações dos dados compartilhados no repositório *Central de Fatos*. Por fim, a Seção 5 conclui o trabalho.

2. Construção do Repositório *Central de Fatos*

Para combater de forma eficiente a difusão em larga escala de conteúdos de desinformação, é primordial obter um conjunto de dados de alta qualidade que tenha sido rotulado por anotadores especialistas (i.e., jornalistas) com experiência de domínio [Reis et al. 2020]. Além disso, é importante que esses dados abranjam diferentes domínios (i.e., saúde, política, etc) e assuntos (i.e., COVID-19, eleições municipais, etc). Diante deste cenário, dados oriundos de agências de checagem de fatos se tornam um recurso valioso para estudos no contexto da desinformação. Primeiro, as agências de checagem reúnem equipes altamente especializadas para realização da verificação dos fatos. Depois, todos os vereditos fornecidos são efetuados com base em uma análise bastante cuidadosa com base em dados que os suporte. Por fim, comumente as agências de checagem de fatos não são limitadas a domínios e assuntos específicos, o que proporciona uma grande variedade em relação aos fatos verificados.

Neste trabalho apresentamos um novo conjunto de dados composto de checagens de fatos coletadas a partir de 6 diferentes agências brasileiras. Esses dados podem ser utilizados tanto no entendimento quanto na proposição de estratégias para contenção da desinformação no contexto brasileiro (detalhes são apresentados na Seção 4). A Figura 1 apresenta uma visão geral da metodologia proposta para construção do repositório de dados disponibilizado neste trabalho.

De forma geral, a primeira etapa (1) consiste na *definição das agências de checagem de fatos* a serem coletadas. Para este trabalho, foram selecionadas 6 agências de checagem de fatos com reconhecimento e abrangência nacional, e algumas delas, inclusive, signatárias do código de conduta da IFCN⁹, que estabelece em nível mundial princípios que devem ser observados por essas entidades. São elas: Aos Fatos¹⁰, Estadão Verifica¹¹, Lupa¹², Boatos¹³, Comprova¹⁴ e Fato ou Fake¹⁵.

⁹<https://www.ifcncodeofprinciples.poynter.org/>

¹⁰<https://www.aosfatos.org/>

¹¹<https://politica.estadao.com.br/blogs/estadao-verifica/>

¹²<https://piaui.folha.uol.com.br/lupa/>

¹³<https://www.boatos.org/>

¹⁴<https://projeto comprova.com.br/>

¹⁵<https://g1.globo.com/fato-ou-fake/>

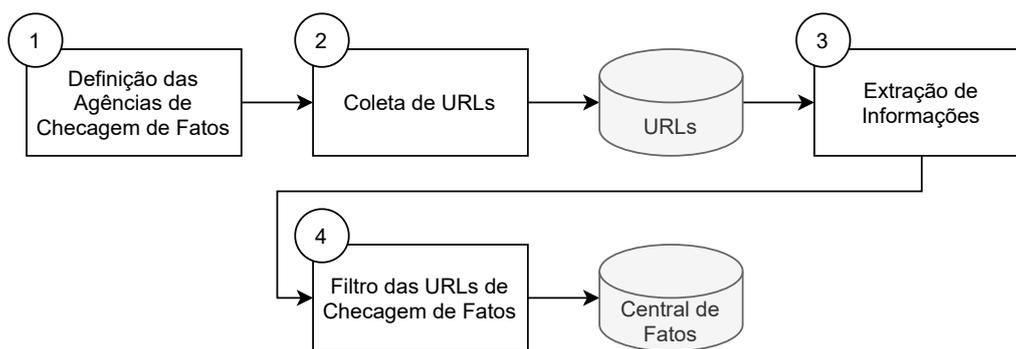


Figura 1. Visão geral do processo de construção do repositório de dados.

A segunda etapa (2) compreende o processo de *coleta das URLs* disponíveis na página (ou *websites*) de cada uma das agências de checagem de fatos selecionadas. Neste contexto, é válido ressaltar que as páginas das agências coletadas não são constituídas apenas de URLs relativas à checagens de fatos. Em geral, outros tipos de conteúdo também são publicados (e.g., notícias, etc). No entanto, como parte do processo de coleta de dados, todo o conteúdo de interesse (i.e., checagem de fatos) é filtrado na etapa posterior.

Em seguida, a partir do repositório de dados com as URLs coletadas, a próxima tarefa consiste na *extração das informações* de interesse (etapa (3)). Esta é uma tarefa complexa, uma vez que as páginas das agências de checagem de fatos não possuem uma padronização clara de estruturação do conteúdo publicado. Assim, foi necessário implementar coletores de dados específicos para cada uma das agências de checagem de fatos exploradas neste trabalho. Todos os coletores foram desenvolvidos com o uso da linguagem de programação Python versão 3. De forma geral, a partir de um parser do HTML de cada uma das URLs coletadas, foram extraídas as seguintes informações: título da checagem, subtítulo, data de publicação, texto, link da imagem principal associada (se houver), link do vídeo associado (se houver), autor(es) da checagem (se estiver(em) disponível(eis)), categoria (se estiver disponível), tags (se estiver disponível), e veredito (ou rótulo) da checagem. Por fim, a informação de veredito (coletada durante a etapa anterior) foi considerada na etapa (4) *filtro das URLs de checagens de fatos*. Neste processo, foram descartadas 3 266 URLs que não correspondiam a checagem de fatos.

Especificamente sobre os vereditos coletados, existem algumas considerações. Primeiro, é válido ressaltar que algumas agências de checagem de dados fornecem, para um único fato checado, vários rótulos. Por exemplo, em uma análise de um discurso político é comum que existam várias informações acerca de tópicos distintos. Neste caso, cada informação, ainda que dentro de um mesmo discurso, pode receber um rótulo distinto. Assim, para o repositório de dados disponibilizado neste trabalho, os vereditos (ou rótulos) em cada uma das checagens foram salvos como uma lista onde os elementos podem variar de 1 até n . Adicionalmente, existe um caso específico na agência Estadão Verifica onde os vereditos das checagens de fatos publicadas são fornecidos em formatos não textuais, principalmente imagens. Para esses casos, a informação textual desses campos foi extraída das imagens usando o reconhecimento óptico de caracteres (OCR, do inglês *Optical Character Recognition*) disponibilizado pela Python-tesseract¹⁶, uma

¹⁶<https://pypi.org/project/pytesseract>

biblioteca Python específica para este fim. Além disso, esse mesmo processo foi aplicado para coleta de tags para a agência Boatos.

Em suma, o repositório de dados construído neste trabalho é composto de 11 647 checagens de fatos realizadas durante os anos de 2013 e 2021 no contexto brasileiro. Os dados, salvos em formato CSV com a utilização do delimitador ';', estão disponíveis no seguinte endereço: <https://doi.org/10.5281/zenodo.5191798>. Especificamente, cada linha do arquivo possui os seguintes dados referentes à uma checagem:

- **url**: Link para a checagem;
- **source_name**: Agência responsável pela realização da checagem;
- **title**: Título da checagem;
- **subtitle**: Subtítulo da checagem (quando disponível);
- **publication_date**: Data da publicação da checagem, em formato YYYY-MM-DD, onde YYYY, MM e DD representam ano, mês e dia, respectivamente;
- **text_news**: Texto associado a checagem;
- **image_link**: Link para a imagem principal associada a checagem (se houver);
- **video_link**: Link para o vídeo associada a checagem (se houver);
- **authors**: Autor(es) responsável(is) pela realização da checagem;
- **categories**: Categorias relacionadas à checagem (se estiver disponível);
- **tags**: Palavras-chave relacionadas à checagem (quando disponibilizadas pela agência de checagem de fatos);
- **obtained_at**: Data da coleta, também em formato YYYY-MM-DD.
- **rating**: Vereditos das checagens (lista);

É importante ressaltar que é recorrente as agências analisarem a mesma notícia, o que pode acarretar em checagens similares. Para a construção dessa base optou-se por manter essas checagens, tendo em vista que elas podem conter dados e vereditos distintos ainda que tratem sobre a mesma informação. Além disso, é válido ressaltar que as agências de checagem de fatos comumente não disponibilizam a fonte original do fato (ou desinformação). No entanto, conforme discutido na Seção 4 deste trabalho, os dados disponibilizados podem ser utilizados como base para verificação de correspondência entre checagem e fonte da desinformação. Ademais, outra limitação dos dados está relacionada às checagens fornecidas pelas agências que não contêm rótulos explícitos dos vereditos. Esses casos foram removidos durante a etapa de (4) filtro das informações de interesse. Ainda assim, o repositório aqui disponibilizado é inédito e consiste em recurso valioso para diversos estudos nos contextos de identificação e combate à desinformação no cenário brasileiro.

3. Descrição do Repositório *Central de Fatos*

Nesta seção, apresentamos alguns resultados que descrevem, em linhas gerais, os dados disponibilizados no repositório *Central de Fatos*. A Tabela 1 apresenta o volume de dados de cada agência de checagem, a média de publicações por mês, bem como a média de palavras das checagens disponibilizadas por estas agências. A agência *Boatos* possui o maior número de checagens, sendo também a mais antiga entre as agências selecionadas. A média de palavras por checagem é similar entre quase todas as agências, com exceção da *Comprova*. Este fato pode ser explicado pela natureza das checagens compartilhadas por esta agência: cada checagem é elaborada de forma colaborativamente entre jornalistas

representantes dos veículos participantes da iniciativa e é apenas publicada quando pelo menos três das redações participantes aprovam as conclusões obtidas. Este processo, conhecido como checagem cruzada¹⁷, ou, CrossCheck¹⁸, pode resultar em uma média de palavras por checagem mais elevada. Interessante notar que, apesar de terem diferentes períodos de atuação, todas as agências possuem um elevado número médio de publicação por mês.

Agência	Data Inicial	Data Final	Total	Média de palavras	#publicações/#mês
Comprova	16-07-2019	19-05-2021	361	1.813	16,63
Estadão Verifica	09-08-2018	19-05-2021	593	806	18,13
Aos Fatos	09-10-2015	13-05-2021	1.679	756	25,49
Boatos	01-07-2013	14-05-2021	5.523	639	59,59
Fato-ou-Fake	30-07-2018	21-04-2021	917	553	28,54
Lupa	25-11-2015	19-04-2021	2.574	620	40,46

Tabela 1. Algumas estatísticas referentes às checagens publicadas por agência.

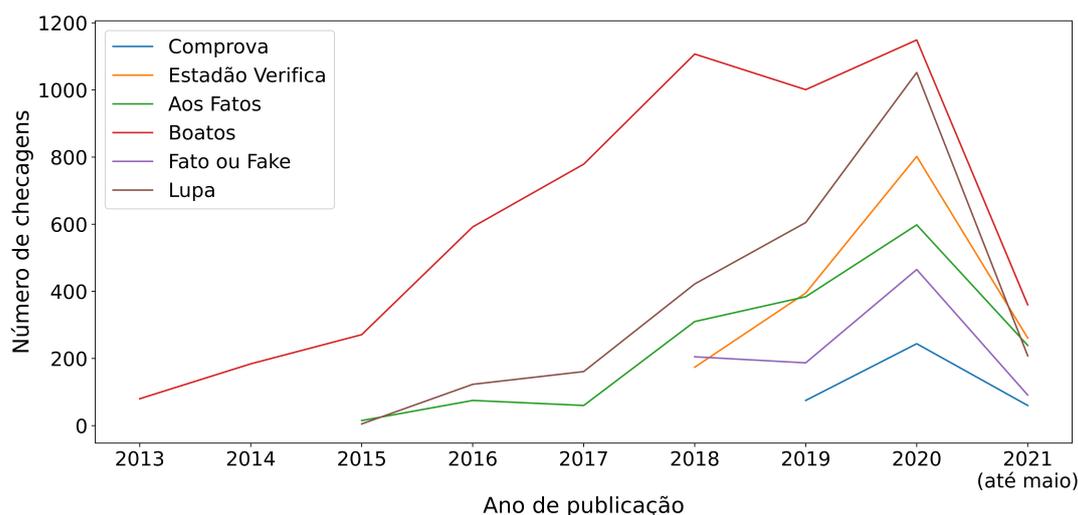


Figura 2. Total de publicações por agência e ano (de 2013 até maio de 2021).

A Figura 2 apresenta a série temporal do total de checagens compartilhadas por cada agência. Exceto pelo ano de 2021 (considerando os dados coletados até o mês de maio), notamos uma clara tendência no aumento do número de publicações das agências de checagem, em grande parte decorrente do aumento da disseminação de desinformação em diferentes plataformas digitais que traz como consequência uma alta demanda por checagens. Em particular, destacamos o vertiginoso crescimento no número de checagens em 2020, provavelmente devido à disseminação de conteúdos de desinformação em larga escala associada à pandemia da COVID-19 [Ferrara 2020].

A seguir, analisamos algumas características das checagens presentes no nosso repositório. A Tabela 2 sumariza a quantidade de checagens associadas à cada rótulo atribuído aos conteúdos. Vale ressaltar que cada agência possui um conjunto próprio de

¹⁷<https://projeto comprova.com.br/about/faqs/>

¹⁸<https://firstdraftnews.org/about/crosscheck-newsroom/>

Agência	Rótulos
Comprova	enganoso (183), falso (159), comprovado (9), evidência comprovada (6), contexto errado (4)
Estadão Verifica	falso (299), enganoso (160), fora de contexto (134)
Aos Fatos	falso (2522), verdadeiro (637), impreciso (394), exagerado (239), distorcido (125), insustentável (120), contraditório (65)
Boatos	boato (5523)
Fato-ou-Fake	fake (1098), fato (394), não é bem assim (346)
Lupa	falso (3209), verdadeiro (1469), exagerado (866), verdadeiro mas (723), de olho (222), contraditório (189), subestimado (108), ainda é cedo para dizer (108), insustentável (97)

Tabela 2. Rótulos associados à checagens publicadas por cada agência.

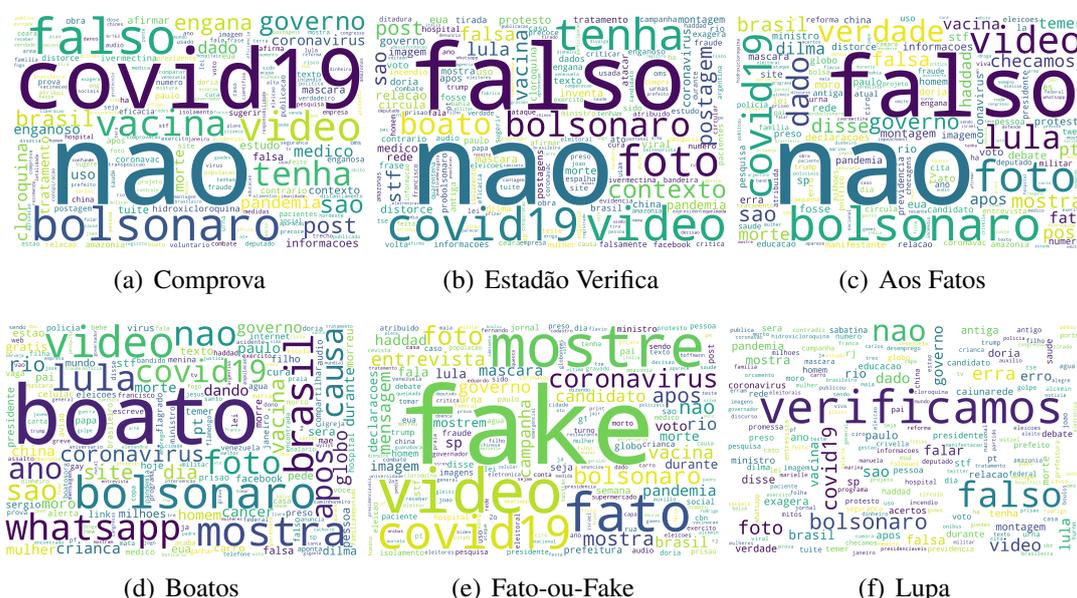


Figura 3. Nuvem de palavras extraída dos títulos das checagens coletadas.

rótulos a serem atribuídos. Podemos notar que vereditos como “enganoso”, “falso”, “boato”, e “fake”, são predominantes. Além disso, a agência Lupa possui a maior variedade de rótulos disponíveis seguida da agência Aos Fatos.

Por fim, a Figura 3 mostra a nuvem de palavras extraída dos títulos das checagens coletadas. De forma geral, é possível notarmos a expressividade de palavras como “falso”, “boato”, e “fake” indicando que as diferentes agências incluem o veredito no título da checagem. Além disso, a presença de palavras como “vacina”, “covid19”, “bolsonaro”, “lula” e “governo” destaca a diversidade de assuntos (i.e., saúde, política, etc) cobertos pelo repositório disponibilizado neste trabalho.

4. Potencial de Utilização dos Dados do *Central de Fatos*

A seguir, apresentamos algumas direções de pesquisa que podem se beneficiar dos dados disponibilizados no repositório *Central de Fatos*.

Caracterização das agências brasileiras de checagem de fatos. No combate à

desinformação, a checagem de fatos obteve um grande destaque como uma ferramenta efetiva. Neste contexto, as checagens e as agências que as publicam tornam-se também objetos de interesse acadêmico [Pereira and Marques-Neto 2021]. Os dados disponibilizados através do nosso repositório podem ser utilizados para a elaboração de diversas análises que visam caracterizar o cenário de verificação de fatos no Brasil. Entre estas análises, podemos citar a comparação entre os resultados de checagem provenientes de diferentes agências ou ainda a metrificacão da interseção entre as checagens que são publicadas por elas.

Caracterização da desinformação no cenário nacional. É importante salientar a magnitude do efeito danoso causado pela disseminação descontrolada de instâncias de desinformação. Em particular, esses efeitos foram exacerbados pela pandemia do coronavírus, na qual a disseminação deste tipo de conteúdo foi amplamente potencializada pelo sentido de vulnerabilidade generalizado e a consequente histeria coletiva, criando um cenário no qual a desinformação torna-se verdadeiramente um problema de saúde pública [Galhardi et al. 2020]. Neste cenário, buscar novas percepções sobre conteúdos de desinformação tal como identificar os padrões explorados pelas agências que viabilizam a etiquetagem ou ainda como as temáticas destes conteúdos se relacionam com acontecimentos concorrentes é de suma importância para a elaboração de novas soluções. Além disso, nosso repositório permite, através do casamento entre checagens de fatos e conteúdos suspeitos, a identificação de instâncias de desinformação em propagação no Brasil, a partir das quais é possível extrair padrões de disseminação e características de sua concepção.

Detecção automática de desinformação. Um desafio frequente no combate à desinformação é a falta de dados classificados por entidades reconhecidas. Nosso repositório pode ser utilizado para identificar instâncias de desinformação em diversas fontes suspeitas. Essas instâncias podem então ser utilizadas no treinamento de algoritmos de aprendizado de máquina na tentativa de identificar características e distribuições associadas à este tipo de conteúdo. Esses modelos de aprendizado de máquina podem ser posteriormente utilizados para a detecção automatizada de conteúdos de desinformação, inclusive em uma variedade ampla de tipos de fontes, tais como imagens, textos e vídeos.

5. Conclusões

Neste trabalho, apresentamos o repositório de dados *Central de Fatos* composto por 11 647 checagens de fatos realizadas no contexto brasileiro entre 2013 e 2021, provenientes de 6 importantes agências de checagens de fatos brasileiras: Aos Fatos, Estadão Verifica, Lupa, Boatos, Comprova e Fato ou Fake.

As checagens incluídas no nosso repositório cobrem uma ampla e diversificada amostra das instâncias de desinformação em circulação no Brasil, valendo-se também ressaltar que ao abranger diversas fontes de checagens, nosso repositório contempla instâncias de desinformação em um largo espectro de mídias: de postagens em redes sociais a vídeos, memes e declarações de figuras públicas. Por fim, acreditamos que os dados da *Central dos Fatos* podem auxiliar futuras pesquisas relacionadas à caracterização e detecção automática de desinformação no cenário nacional, visando combater, de forma eficiente, a difusão deste tipo de informação nociva à sociedade.

Agradecimentos

Este trabalho foi realizado com apoio financeiro do Ministério Público de Minas Gerais (MPMG), projeto Capacidades Analíticas, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Referências

- Arun, C. (2019). On whatsapp, rumours, and lynchings. *Economic & Political Weekly*, 54(6):30–35.
- Bessi, A. and Ferrara, E. (2016). Social bots distort the 2016 us presidential election online discussion. *First Monday*, 21(11).
- Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F., and Flammini, A. (2015). Computational fact checking from knowledge networks. *PLOS ONE*, 10(6):e0128193.
- Ferrara, E. (2020). What types of covid-19 conspiracies are populated by twitter bots? *First Monday*.
- Galhardi, C. P., Freire, N. P., de Souza Minayo, M. C., and Fagunde, M. C. M. (2020). Fato ou fake? uma análise da desinformação frente à pandemia da covid-19 no brasil. *Ciência & Saúde Coletiva* [<https://doi.org/10.1590/1413-812320202510.2.28922020>], 25:4201–4210.
- Gomes Jr, L. and Frizzon, G. (2019). Fake news and brazilian politics—temporal investigation based on semantic annotations and graph analysis. In *Anais do Simpósio Brasileiro de Banco de Dados (SBBDD)*, pages 169–174.
- Machado, C., Kira, B., Narayanan, V., Kollanyi, B., and Howard, P. (2019). A study of misinformation in whatsapp groups with a focus on the brazilian presidential elections. In *Companion Proceedings of the World Wide Web conference (WWW)*, pages 1013–1019.
- Martins, A. D. F., Cabral, L., Mourão, P. J. C., Monteiro, J. M., and Machado, J. (2021). Detection of misinformation about covid-19 in brazilian portuguese whatsapp messages. In *Proc. of the Int’l Conference on Applications of Natural Language to Information Systems*, pages 199–206.
- Myslinski, L. J. (2012). Fact checking method and system. Google Patents. US Patent 8,185,448.
- Newman, N., Fletcher, R., Kalogeropoulos, A., and Nielsen, R. K. (2019). Reuters Institute Digital News Report 2019. Reuters Institute for the Study of Journalism.
- Pereira, C. G. and Marques-Neto, H. T. (2021). Caracterização da reação de agências de fact-checking às publicações sobre a pandemia da covid-19 em redes sociais. In *Anais do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, pages 113–126.
- Reis, J. C., Melo, P., Garimella, K., Almeida, J. M., Eckles, D., and Benevenuto, F. (2020). A dataset of fact-checked images shared on whatsapp during the brazilian and

- indian elections. In *Proc. of the Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 903–908.
- Reis, J. C. S., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, 34(2).
- Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., and Benevenuto, F. (2019). (mis)information dissemination in whatsapp: Gathering, analyzing and countermeasures. In *Proc. of the ACM Web Conference (WWW)*, pages 818–828.
- Vlachos, A. and Riedel, S. (2014). Fact checking: Task definition and dataset construction. In *Proc. of the ACL Workshop on Language Technologies and Computational Social Science*, pages 18–22.
- Wu, Y., Agarwal, P. K., Li, C., Yang, J., and Yu, C. (2014). Toward computational fact-checking. *Proce. of the VLDB Endowment*, 7(7):589–600.