

Trending Fake⁽¹⁾: Análise estatística via algoritmo NER (Named Entity Recognition) das correlações entre as fake news brasileiras e os tópicos abordados no Twitter – Fase 2



Pietro Andrade Quinzani e João Pedro Sassi Sandre
Prof. Dr. Francisco Tupy Gomes Correa (orientador)

COLÉGIO VISCONDE DE PORTO SEGURO
Rua Floriano Peixoto Santos, 55 – São Paulo/SP



Fig. 1. ODS 16/4
Fonte: Organização das Nações Unidas

INTRODUÇÃO

Na fase anterior de pesquisa, identificou-se que as *fake news* aparentam acompanhar as tendências temáticas da sociedade. Como podemos então verificar essa hipótese?



PROBLEMA DE PESQUISA

Qual é a correlação existente entre os tópicos abordados nas *fake news* e os tópicos abordados no mesmo período no Twitter e na grande mídia?

OBJETIVO GERAL

Comparação estatística entre os tópicos identificados pelo algoritmo de reconhecimento automático de temas (NER) nas notícias falsas brasileiras, nos trending topics do Twitter e na amostra coletada de notícias de grande circulação no segundo semestre de 2022.

OBJETIVOS ESPECÍFICOS

Automação do algoritmo de NER para a criação de uma ferramenta on-line de exposição dos principais tópicos das *fake news*.
Adicionalmente, testagem do uso de análise multivariada para a predição de temas futuros de *fake news*.

MÉTODOS E TÉCNICAS

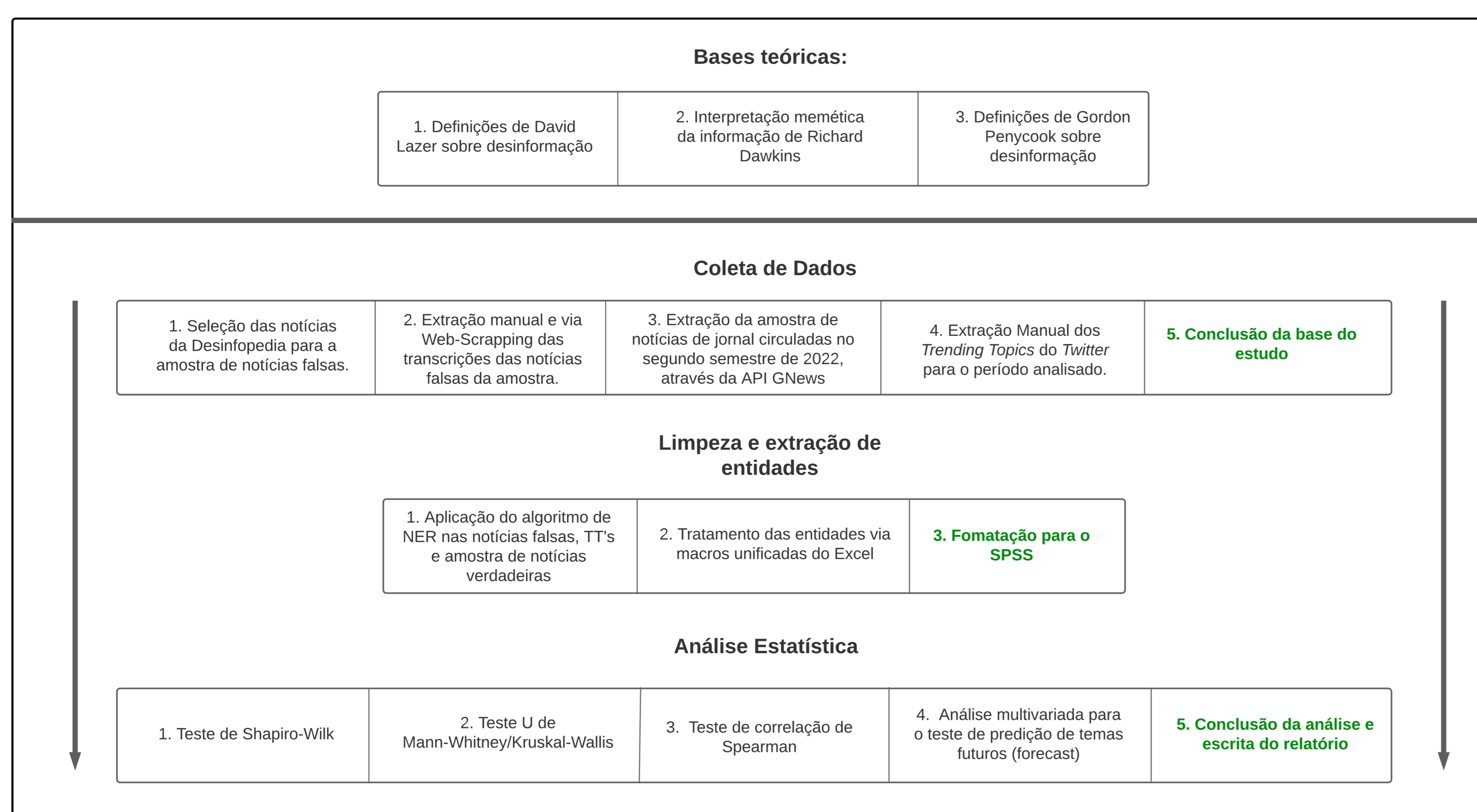


Fig. 2. Desenvolvimento da pesquisa
Fonte: Autores

RESULTADOS

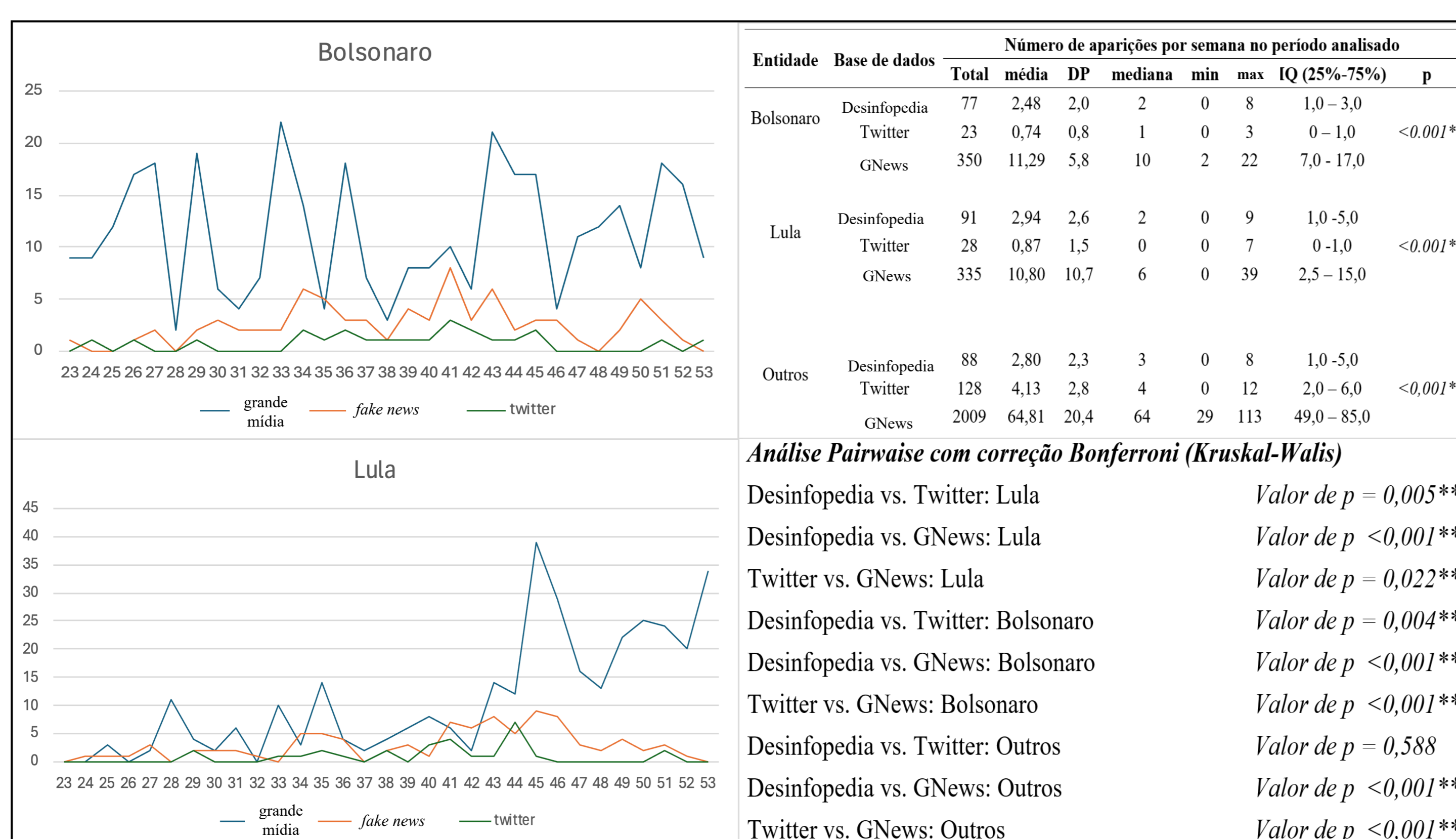


Fig. 3. Conjunto das análises estatísticas comparativas das distribuições
Fonte: Autores

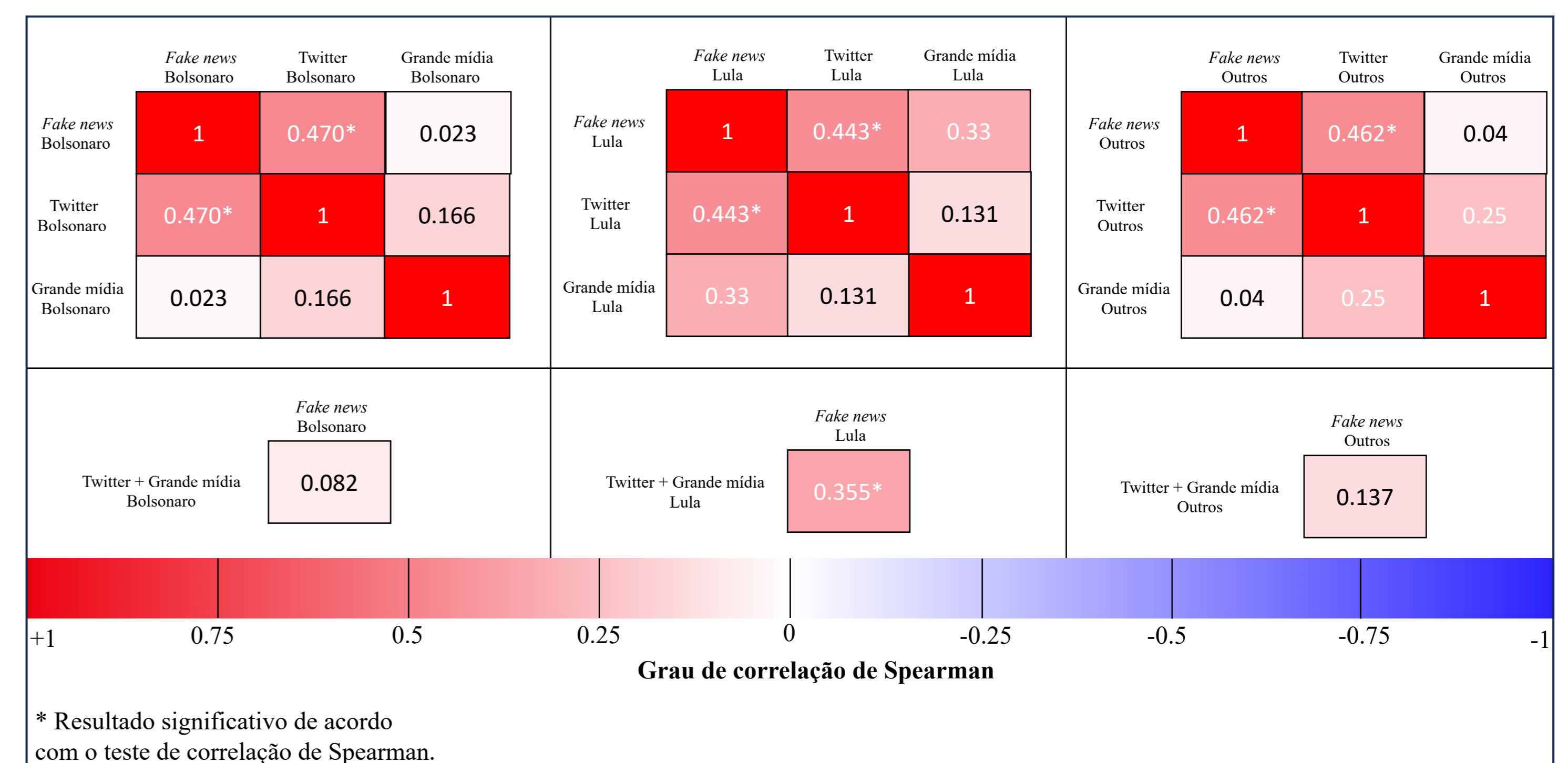


Fig. 4. Matrizes de correspondência Spearman das métricas comparadas
Fonte: Autores

Modelo de predição para a base de dados fake news					
	Coefficiente	Erro padrão de coeficiente	t value	p value	
Bolsonaro	Constante	1,535077	0,702071	2,186	0,03730*
	Preditor: GNews	-0,004866	0,053945	-0,090	0,92877
	Preditor: Twitter	1,352857	0,381375	3,547	0,00139*
Erro padrão residual: 1,685 com 28 graus de liberdade					
R ² : 0,3131, R ² ajustado: 0,2641					
F: 6,382 com 2 e 28 graus de liberdade, p-value: 0,005203					
Lula	Constante	1,46701	0,65040	2,256	0,0321*
	Preditor: GNews	0,08815	0,03971	2,220	0,0347*
	Preditor: Twitter	0,57116	0,27610	2,069	0,0479*
Erro padrão residual: 2,321 com 28 graus de liberdade					
R ² : 0,245, R ² ajustado: 0,1911					
F: 4,543 com 2 e 28 graus de liberdade, p-value: 0,01955					

Legenda: * resultado significativo de acordo com o modelo de regressão linear.

Fig. 5. Resultado da predição do comportamento futuro de fake news
Fonte: Autores

DISCUSSÃO E CONCLUSÃO

- Após a coleta e análise via NER de 265 fake news, 2 130 trending topics do Twitter e 2 140 notícias verdadeiras veiculadas na grande mídia, identificamos os temas abordados por estes.
- Obtivemos evidências que sustentam a afirmação inicial de que as notícias falsas são dinâmicas e acompanham as tendências da sociedade (PENNYCOOK, 2020).
- Na comparação entre os temas das fake news e os temas em alta no Twitter, observamos uma correlação moderada. Isso indica que as notícias realmente acompanham em certo grau o que é mencionado nessa rede social.
- Os resultados obtidos na análise dos temas veiculados na grande mídia contestam a teoria de "Agenda Setting", indicando uma possível necessidade de revisão do conceito sob as novas dinâmicas digitais de comunicação. Tal defasagem da teoria é respaldada por outros pesquisadores do campo.
- Completamos a automação do algoritmo de NER criado para o trabalho a fim de utilizá-lo na montagem de uma ferramenta on-line de acompanhamento em tempo real das tendências temáticas das fake news.
- Com base na correlação identificada no estudo, obtivemos sucesso na predição do comportamento futuro das fake news usando análise multivariada.

LAKATOS, E. M.; MARCONI, M. de A. Fundamentos de metodologia científica. 3. ed. São Paulo: Atlas, 1991.

LAZER, D. M. J.; BAUM, M. A.; BENKLER, Y.; BERINSKY, A. J.; GREENHILL, K. M.; MENCZER, F.; METZGTER, M. J.; NYHAN, B.; PENNYCOOK, G.; ROTHSCHILD, D.; SCHUDON, M.; SLOMAN, S. A.; SUNSTEIN, C. R.; THORSON, E. A.; WATTS, D. J.; ZITTRAIN, J. L. The Science of Fake News. *Science*, 359:1094-1096, 2018. Disponível em: <<https://www.science.org/doi/10.1126/science.aao2998>>.

PENNYCOOK, G.; RAND, D. G. The Psychology of Fake News. *Trends in Cognitive Sciences*, 25:388-402, 2021. Disponível em: <<https://doi.org/10.1016/j.tics.2021.02.007>>.

(1) Os nomes Trending Fake e Desinfopedia são marcas que se encontram sob processo de patente junto ao INPI.