# Classificação em tempo real de Fake News considerando o conteúdo textual e conteúdo social.

Proposta de Pesquisa por Thiago Amado Costa

## 1. Resumo

O fenômeno das Fake News se tornou um desafio significativo na era da mídia social, representando sérias ameaças à credibilidade das informações compartilhadas on-line. Além disso, indivíduos e grupos específicos podem se tornar alvos dessas ações maliciosas, criando comunidades que acreditam e compartilham notícias das mesmas fontes maliciosas. Atualmente, pesquisas demonstram a importância da detecção de notícias falsas, principalmente em seus estágios iniciais de propagação, utilizando diferentes abordagens. Este estudo aborda esse desafio, ao utilizar um método que integra duas abordagens distintas que consideram tanto o conteúdo textual da notícia como o conteúdo social, adaptando-as para a detecção precoce de notícias falsas.

## 2. Introdução

O fenômeno das Fake News é, atualmente, um dos maiores desafios para a sociedade moderna, ameaçando a credibilidade das informações compartilhadas online. Segundo [Lazer et al. (2018)], Fake News é definida como informação incorreta ou enganosa fabricada para imitar a estrutura das mídias de notícias, mas sem o processo de garantir a precisão e credibilidade. Esse fenômeno representa sérias ameaças às mídias sociais atuais, pois se espalha mais rapidamente e de forma mais ampla do que as informações confiáveis, conforme [Vosoughi et al. (2018)]. Além disso, segundo [Zhou e Zafarani (2020)], quanto mais uma notícia falsa se espalha, maior a probabilidade de pessoas confiarem e acreditarem nelas, o que se torna um sério problema considerando também que notícias falsas são criadas e publicadas de forma mais fácil e rápida em mídias sociais do que em mídias tradicionais.

A existência do efeito das chamadas "echo chambers" (câmaras de eco) nas redes sociais também é um agravante ao problema. Esse conceito pode ser definido, segundo [Cinelli et al. 2021], como grupos de indivíduos que se envolvem em torno de opiniões, crenças e inclinações políticas parecidas, possuindo tendências e atitudes similares. Ainda, a existência dessas câmeras de eco podem reforçar e levar esses indivíduos a desenvolverem ideais cada vez mais extremos. Considerando esse efeito, informações tendenciosas são reforçadas e ampliadas, causando efeitos diferentes de acordo com o grupo atingido ([Zhou e Zafarani (2020)]).

Isso reforça a necessidade de mais estudos e desenvolvimento de metodologias de detecção de Fake News, principalmente em seus estágios iniciais. Nesse contexto, o User Preference-aware FakeNews Detection Framework (Detecção de Notícias Falsas Sensível às Preferências do Usuário - UPFD), proposto por [Dou et al. 2021], além de

utilizar informações textuais das notícias para a detecção de fake news, utiliza informações sociais dos usuários que as propagaram, o que segundo os autores desempenha um papel fundamental nessa detecção, e o modelo proposto atinge performances consideráveis. Uma explicação mais detalhada do framework pode ser lida na Metodologia.

Dessa forma, o objetivo deste estudo é adaptar o framework UPFD no contexto de detecção precoce de notícias falsas, de modo a confirmar a importância dos contextos sociais nessa detecção.

## 3. Conteúdo

Atualmente, diversas abordagens para a detecção automática de notícias falsas empregam técnicas como Aprendizado de Máquina, Processamento de Linguagem Natural e Teoria dos Grafos. Estas incluem perspectivas baseadas no conteúdo textual das notícias e outras que exploram o contexto de propagação, como câmaras de eco, além de abordagens combinadas que mesclam análise textual e contextual para identificação das notícias falsas.

Avanços recentes na detecção de notícias falsas têm aproveitado Redes Neurais em Grafos (GNNs) para identificar padrões na propagação de informações, independentemente do conteúdo textual. [Han et al. (2020)] propõe um método de detecção baseado na propagação usando GNNs que explora os comportamentos distintos de disseminação de notícias falsas versus reais online. Esse estudo mostra que a exploração do grafo de propagação das notícias pode ser utilizada na identificação de notícias falsas, ao mostrar que o mesmo método em grafos menores também obtém performances consideráveis.

Em outro estudo que foca na detecção precoce de notícias falsas, [Zhou et al. 2020] apresenta um modelo orientado por teoria, com foco único na análise do conteúdo textual das notícias em vez de sua propagação nas plataformas de mídia social. Aproveitando as técnicas de aprendizado de máquina, o modelo demonstra um desempenho superior em comparação com os métodos existentes, demonstrando sua capacidade de obter detecção  precoce mesmo quando confrontado com informações limitadas sobre o conteúdo.

Dessa forma, considerando os estudos anteriores, o objetivo deste estudo é adaptar o framework UPFD (Dou, 2021) no contexto de detecção precoce de notícias falsas, de modo a confirmar a importância dos contextos sociais nessa detecção. Esse framework combina um método de detecção baseado em grafos com um método baseado em análise do conteúdo textual, e será explicado na próxima seção.

## 4. Metodologia

Este estudo utiliza uma combinação de conteúdo textual de notícias e contexto social para detectar notícias falsas em seus estágios iniciais, utilizando o repositório de dados FakeNewsNet e uma versão do framework User Preference-aware Fake News Detection.

O conjunto de dados FakeNewsNet é um recurso abrangente projetado para auxiliar na detecção e estudo de notícias falsas. Ele oferece diversas características essenciais para análise, incluindo o conteúdo textual das notícias, contexto social e informações espaço-temporais. Ele inclui conjuntos de dados de duas fontes diferentes, Politifact e Gossipcop, e possui dados que consideram o tanto o conteúdo textual das notícias como contexto social e informações espaço-temporais, o que é instrumental para estudar a detecção precoce de notícias falsas, sua evolução e possíveis estratégias para minimizar seus danos.

O FakeNewsNet pode ser usado para muitas aplicações de pesquisa diferentes além da atual, como a identificação de atores maliciosos que desempenham um papel poderoso na disseminação de notícias falsas. Ele pode também ser expandido, adicionando suporte em diferentes idiomas e diferentes plataformas de mídia social, ao reproduzir sua metodologia, proposta por [Shu, 2019].

O framework User Preference-aware FakeNews Detection (Detecção de Notícias Falsas Sensível às Preferências do Usuário - UPFD) extrai o conteúdo textual das notícias e o contexto social dos conjuntos de dados para construir dois Codificadores. Primeiro, o Codificador Endógeno (ou Codificador Textual das Notícias) utiliza técnicas de representação de texto pré-treinadas para gerar o Embedding da informação textual das notícias e dos posts históricos dos usuários que interagiram com as notícias. O Codificador Exógeno (ou Codificador de Engajamento do Usuário) constrói um grafo de propagação de notícias onde o nó raiz é a notícia e os outros nós representam os usuários que repostaram a notícia. Em seguida, usa uma GNN (Graph Neural Network - Redes Neurais em Grafos) para fundir as características dos usuários com este caminho de propagação da notícia, adicionando o embedding textual das notícias e o embedding das preferências dos usuários como características dos nós, para aprender os embeddings dos nós. Aplicando uma função de leitura que realiza a média dos embeddings de todos os nós, o embedding do Grafo de Propagação de Notícias é obtido. Finalmente, é feita uma concatenação entre o Embedding Textual da Notícia e o Embedding de Engajamento do Usuário para criar o Embedding Final da Notícia, que é alimentado em um Perceptron Multicamadas (MLP) com duas saídas, representando as probabilidades de notícias reais e falsas. O treinamento do modelo utiliza Binary Cross-entropy como função de perda e é atualizado com a função de otimização Stochastic Gradient Descent (SGD).

O framework original UPFD é usado considerando todo o grafo de propagação das notícias. Para usá-lo na detecção precoce, o grafo de propagação é cortado para considerar apenas os primeiros 100 reposts de uma determinada notícia, e apenas 20 posts históricos de um determinado usuário que repostou, em vez dos 200 originais. Comentários dos usuários sobre o post original da notícia também são considerados. Um limite de tempo também é definido, para considerar apenas a primeira hora de disseminação. O objetivo é determinar se as preferências dos usuários e o contexto social podem desempenhar um papel significativo na identificação de notícias falsas em seus estágios iniciais, quando o grafo de propagação é menor e menos usuários interagem com a notícia.

5. Referências

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. Science, 359(6380):1094–1096.

Dou, Y., Shu, K., Xia, C., Yu, P. S., and Sun, L. (2021). User preference-aware fake news detection.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. Science, 359(6380):1146–1151.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H. (2019). Fakenewsnet: A data repository with news content, social context and spatialtemporal information for fake news detection

Zhou, X. and Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection, methods, and opportunities. ACM Computing Surveys, 53(5):1–40.

Ruchansky, N., Seo, S., and Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM ’17. ACM.