# User Preference-aware Fake News Detection

SIGIR'21, Online, July 11-15

**Yingtong Dou, Congying Xia, Philip Yu (University of Illinois at Chicago)**

**Kai Shu (Illinois Institute of Technology)**

**Lichao Sun (Lehigh University)**

Paper: https://arxiv.org/pdf/2104.12259.pdf
Code: https://github.com/safe-graph/GNN-FakeNews
Benchmark: https://paperswithcode.com/dataset/upfd
PyG Example: https://tinyurl.com/a6s92t37
DGL Example: https://tinyurl.com/yjwvd93b
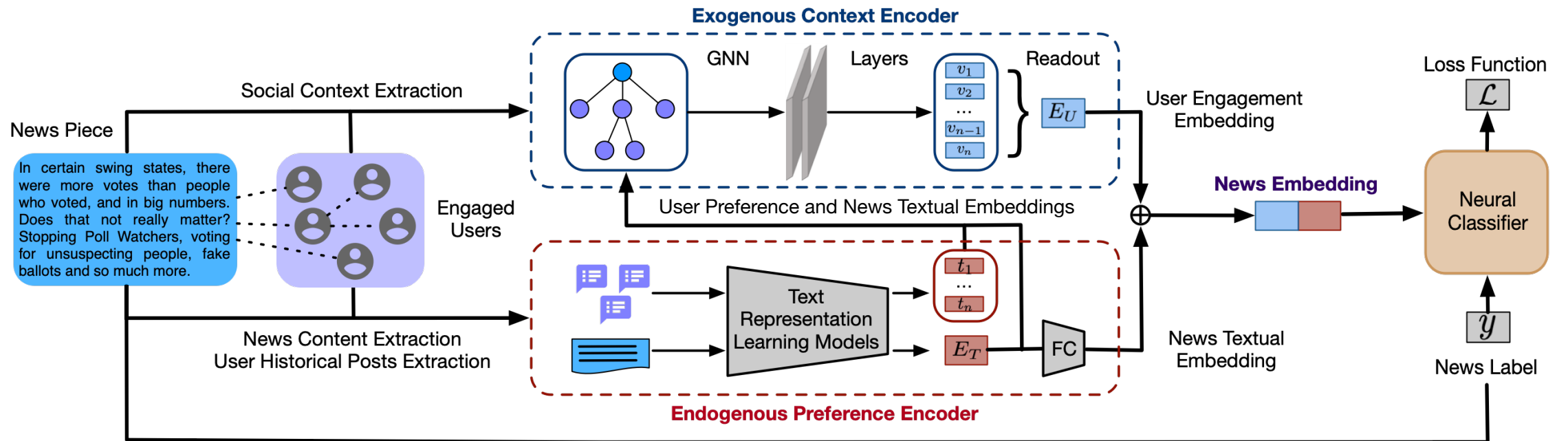
- Modeling news propagation paths on Twitter according to tweets and retweets information

- Crawling 20 million user historical tweets and encoding them as users' news consumption preference

- Proposing an end-to-end fake news classification model with text and graph encoders

- Opensource all baselines and datasets as a benchmark for fake news detection and graph classification

- People tend to consume news that confirm their existing views[1]

- An account engaged in a fake news post on social media may have shared similar news before

- The information in users' historical posts could help verify the truthfulness of the news they engaged

[1] Nickerson, Raymond S. "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of general psychology* 2, no. 2 (1998): 175-220.

**Exogenous Context Encoder**

GNN          Layers          Readout

$v_1$
$v_2$
...
$v_{n-1}$
$v_n$

$E_U$

User Engagement Embedding

Loss Function

$\mathcal{L}$

Social Context Extraction

News Piece

In certain swing states, there were more votes than people who voted, and in big numbers. Does that not really matter? Stopping Poll Watchers, voting for unsuspecting people, fake ballots and so much more.

Engaged Users

User Preference and News Textual Embeddings

$\oplus$

**News Embedding**

Neural Classifier

$t_1$
...
$t_n$

Text Representation Learning Models

$E_T$

FC

News Textual Embedding

News Content Extraction
User Historical Posts Extraction

**Endogenous Preference Encoder**

$y$

News Label

- The proposed UPFD model outperforms various fake news detection baselines

- The user preference information plays a significant role in fake news classification

- Different graph encoders and text encoders are in favor of different news datasets

- The graph data built in this paper has been released as a graph classification/anomaly detection benchmark

- Two sets of tree-structured graph, four types of node features, binary classification task

- Hosted on PaperWithCode, integrated with PyG and DGL

| Data | #Graphs | #Fake News | #Total Nodes | #Total Edges | #Avg. Nodes per Graph |
|------|---------|------------|--------------|--------------|-----------------------|
| Politifact | 314 | 157 | 41,054 | 40,740 | 131 |
| Gossipcop | 5464 | 2732 | 314,262 | 308,798 | 58 |

# User Preference-aware Fake News Detection

SIGIR'21, Online, July 11-15

**Yingtong Dou, Congying Xia, Philip Yu (University of Illinois at Chicago)**

**Kai Shu (Illinois Institute of Technology)**

**Lichao Sun (Lehigh University)**

Slides PDF: http://ytongdou.com/files/sigir21slides.pdf

Project Homepage: https://github.com/safe-graph

Twitter: @dozee_sim

Email: ydou5@uic.edu