Following the Trail of Fake News Spreaders in Social Media
A Deep Learning Model

Antonela Tommasel, Juan Manuel Rodriguez, Filippo Menczer

ISISTAN — Instituto Superior de Ingeniería de Software Tandil

CONICET

INDIANA UNIVERSITY BLOOMINGTON

Fake or unreliable content can pose significant threats to **democracies**, **public health** and **economy**.

- **Can affect how people perceive content.**
  - Alter the likelihood of accepting fake content as truth.
  - The line between what is fake or not becomes more uncertain.

The trustworthiness of the entire news ecosystem might be at risk.

Users play a fundamental role as **creators and disseminators** of fake content.

Detecting spreaders will provide **valuable information** for the design of **mitigation or intervention strategies** to rapidly contain the spreading.

We presented **a model for identifying fake news spreaders in social media** by combining content and user features, the induced propagation trees, and features learned from user interactions.
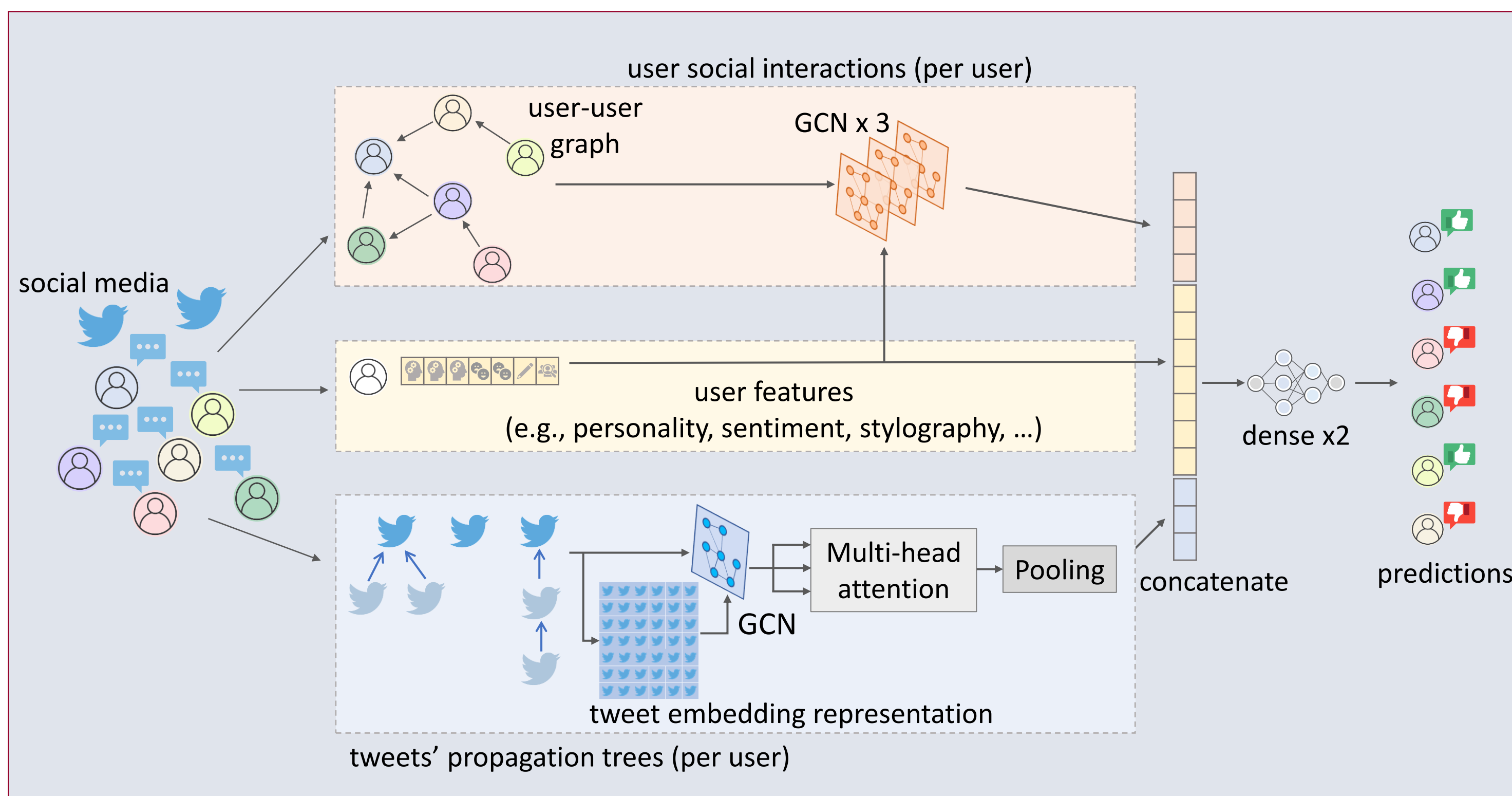
A preliminary evaluation showed the **models' potential for accurately detecting fake news spreaders** and the i**mportance of combining the different aspects of user representation** for effectively characterize spreaders.

	Traditional	State-of-the-art
Avg. precision Improvements	43%	54%
Avg. recall improvements	61%	184%
Avg. AUC-ROC improvements	51%	42%

- Best baselines results were obtained with simple user/tweet features.

- **High precision, but low recall.**

- **Network topology** and **hand-crafted features** seemed to be **more useful** than **content**.

- Our model achieved the **highest results**.

- **Better balance between precision and recall** than the evaluated baselines.

- Some baselines achieved similar precision, but lower recall.

---

social media

user social interactions (per user)
user-user graph
GCN x 3

user features
(e.g., personality, sentiment, stylography, …)

Multi-head attention — Pooling
GCN
tweet embedding representation
tweets' propagation trees (per user)

concatenate — dense x2 — predictions

- User representation is divided into **three components**.

- Features. Vector concatenating **personality traits**, **readability scores**, **sentiment** and **emotions**, …

- Social interactions. **Three concatenated GCNs** allow including interactions from up to **3-hop neighbours** (user community).

- Tweets. Each tweet is represented by a **propagation tree derived from the triggered replies** and the **pooled BERT embeddings** of the involved tweets.

There is still work to do!

- Evaluate with other data collections varying scale and domain.

- Explore user relation representations.

- Explore the temporal relation of tweets.

- Perform an ablation study.