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

ANTONELA TOMMASEL

JUAN MANUEL RODRIGUEZ

FILIPPO MENCZER

Identifying fake news spreaders

- **Fake or unreliable content can severely affect society**, posing significant threats to democracies and economy.
    - With the COVID-19 pandemic, health misinformation arose as a threat to public health.

- **Can affect how people perceive content.**
    - Repeated exposure can alter the likelihood of accepting fake content as truth, especially when the fake content aligns with internal beliefs.
    - The line between what is fake or not becomes more uncertain hindering the differentiation between fake and authentic content.
    - <u>The trustworthiness of the entire news ecosystem might be at risk.</u>

Identifying fake news spreaders

- **Fake or unreliable content can severely affect society**, posing significant threats to democracies and economy.
  - With the COVID-19 pandemic, health misinformation arose as a threat to public health.

- **Can affect how people perceive content.**
  - Repeated exposure can alter the likelihood of accepting fake content as truth, especially when the fake content aligns with internal beliefs.
  - The line between what is fake or not becomes more uncertain hindering the differentiation between fake and authentic content.
  - <u>The trustworthiness of the entire news ecosystem might be at risk.</u>

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

It is **essential to detect both fake content and the users spreading it**, as the latter will provide **valuable information** for the design of **mitigation or intervention strategies** to rapidly contain the spreading.

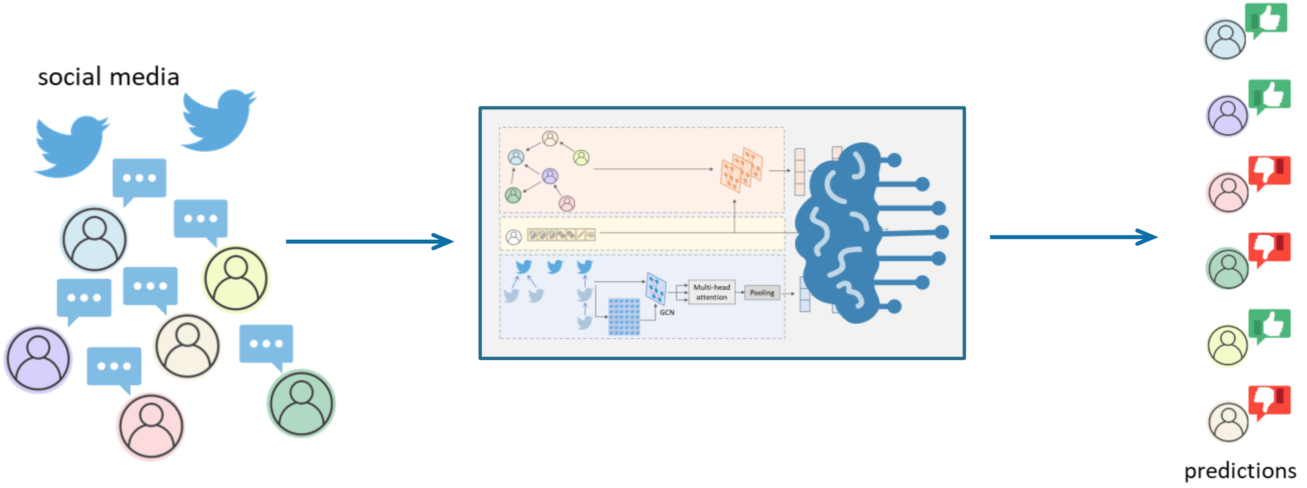
Identifying fake news spreaders

How can we effectively detect fake news spreaders in social media?

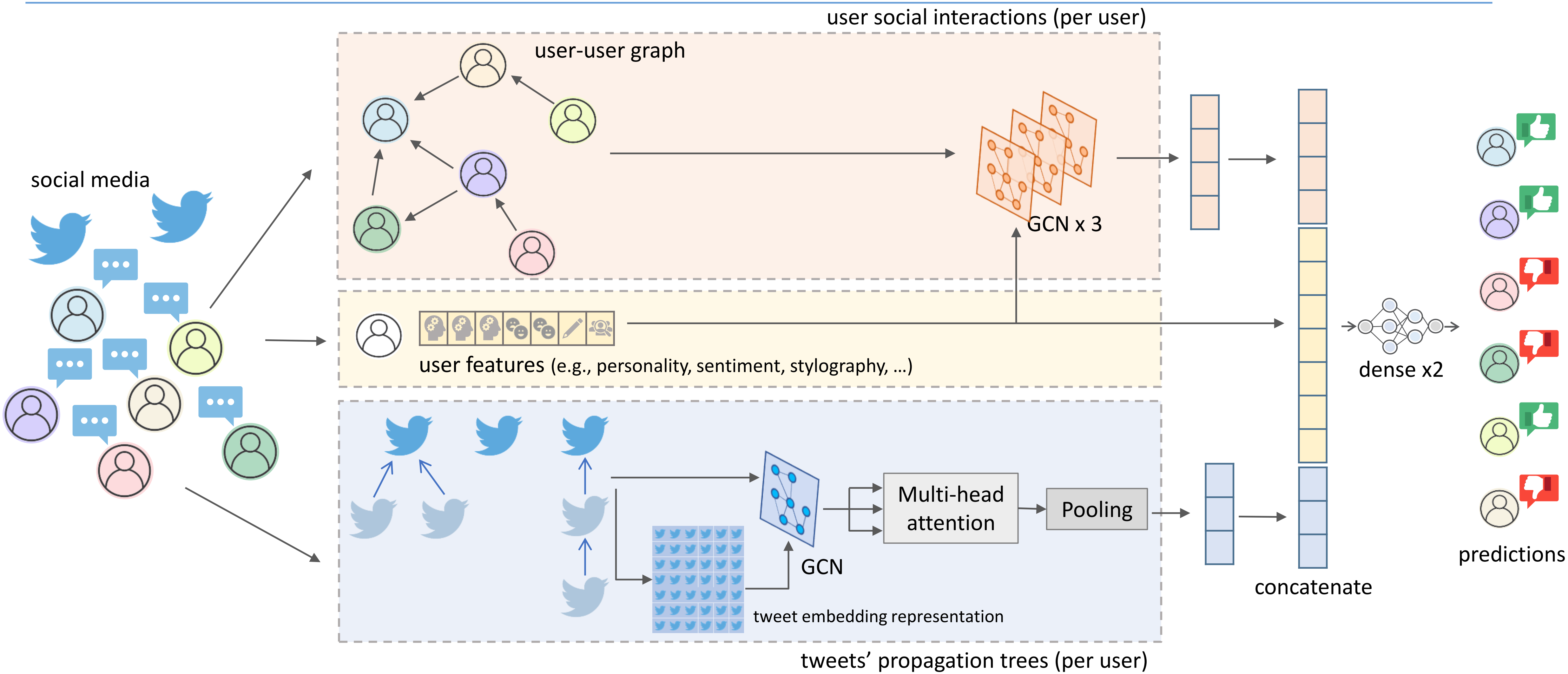
Identifying fake news spreaders

How can we effectively detect fake news spreaders in social media?

For a given user u_i and their **social interactions**, the **shared content** and the **content propagation trees**, the goal is to learn a function $F \rightarrow \{1, -1\}$, such that 1 indicates that u_i is a fake news spreaders, and -1 otherwise.



Model overview



Experimental evaluation
Data

- We used the **FibVid** data collection.
  - Tweets related to the COVID-19 pandemic.
  - The collection is based on news claims appearing in **Politifact** and **Snopes**.
- Tweets were retrieved using the **Faking it!** tool.
- The collection comprised **772** COVID-related news claims and **112k** relevant tweets belonging to **24k** users, which were shared during 2020.
- Tweets have a authentic/fake label based on Politifact and Snopes.
  - 26% authentic content, 74% fake content.
  - Labels were used to determine whether users were fake news spreaders.
  - Users were deemed as spreaders if the **proportion of shared fake content was higher than a certain threshold** (0.5).

Experimental evaluation

Baselines

Traditional

- Based on hand-crafted feature sets.
  - Tweet/user stats (popularity, screenname length, account age, …).
  - LIWC.
  - Personality traits.
  - Readability.
  - Content-based embeddings.
  - Topology-based embeddings.

State-of-the-art

- All based on deep-learning models.
- Mostly based on content-based information.

Experimental evaluation

Evaluation

Evaluation Metrics

- Binary/weighted precision and recall.
  - More importance to recall.
- AUC-ROC.

Data split

- All evaluations were performed over the **same data partitions** and evaluated using the same set of metrics.

- Temporal data split.
- Training set: first 70% users sorted according to the date of their first interaction.
- Test set: remaining users.

Experimental evaluation

Results - Highlights

	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 relatively low recall.
- Hand-crafted content features achieved similar results than considering content embeddings.
- Network topology 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 to our model, but lower recall.

Summary & conclusions

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** to achieve a more effective characterization of spreaders.

Summary & conclusions

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** to achieve a more effective characterization of spreaders.

- Data and code are publicly available.

- Evaluate with other data collections varying scale and domain.
- Explore the representation of user relations.
- Explore the temporal relation of tweets.
- Perform an ablation study.

Thanks!

Questions?

antonela.tommasel@isistan.unicen.edu.ar

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

ANTONELA TOMMASEL
JUAN MANUEL RODRIGUEZ
FILIPPO MENCZER