CERIAS

The Center for Education and Research in Information Assurance and Security

An Interactive Framework for Profiling News Sources

Nikhil Mehta and Dan Goldwasser

(1) Motivation

- Social media is very influential in today's society
- But…harmful information can spread quickly
- Fake news, Politically biased content
- Important to detect it!

- News is being spread on new topics everyday, these are **emerging news events**
- Ex: When COVID-19 happened, a pandemic like that was never seen!
- Makes task harder!

- News often spreads in communities on social media
- Tightly connected groups of users
- Some communities are more likely to spread fake news, others not
- Identifying communities can be useful!

- But Identifying communities is hard!
- Even the best models struggle!
- Humans can help, but can't identify everything
- We can use LLMs + Humans + Graphs to do better!

- **Key Contribution**: ***Propose an interactive approach using LLM, Graph, and Human insight to profile news sources better!***

(2) Task Defn.

- News Media Profiling: Detecting the factuality and political bias of news sources
- Fake News Detection: High, Low, or Mixed Factuality
- Political Bias : Left, Right, or Center

(2) Approach: Graph

- Model data in an information graph
- Nodes: News sources, users, and articles
- Edges: Relationships between nodes
- Train graph for the source detection tasks

(3) User Communities

- Users form communities on social media
- Similar users likely have similar interests and share similar content, which has similar factuality/bias
- Idea: If we can model user preferences and thus user similarity, we can better model the content they propagate

(4) Interactive Approach

- Question: How to identify user communities?
- Humans can look at a few users and determine if they are similar
- Graph architecture models user preferences, captures community structures, but likely incorrect
- LLMs can look at text and determine textual similarity, but can't model user preferences well
- How can we best use this?

- Approach:

1. Cluster users in graph model -> get candidate user communities
2. For each community, determine users the graph model is unsure about
3. Ask humans, are these users similar?
4. Ask LLMs to identify the same user similarity the humans identified, but on other sets of users. Text similarity detection is easier for LLMs!
5. Connect edges between users LLM identifies as similar in the graph
6. Now, profile sources better!

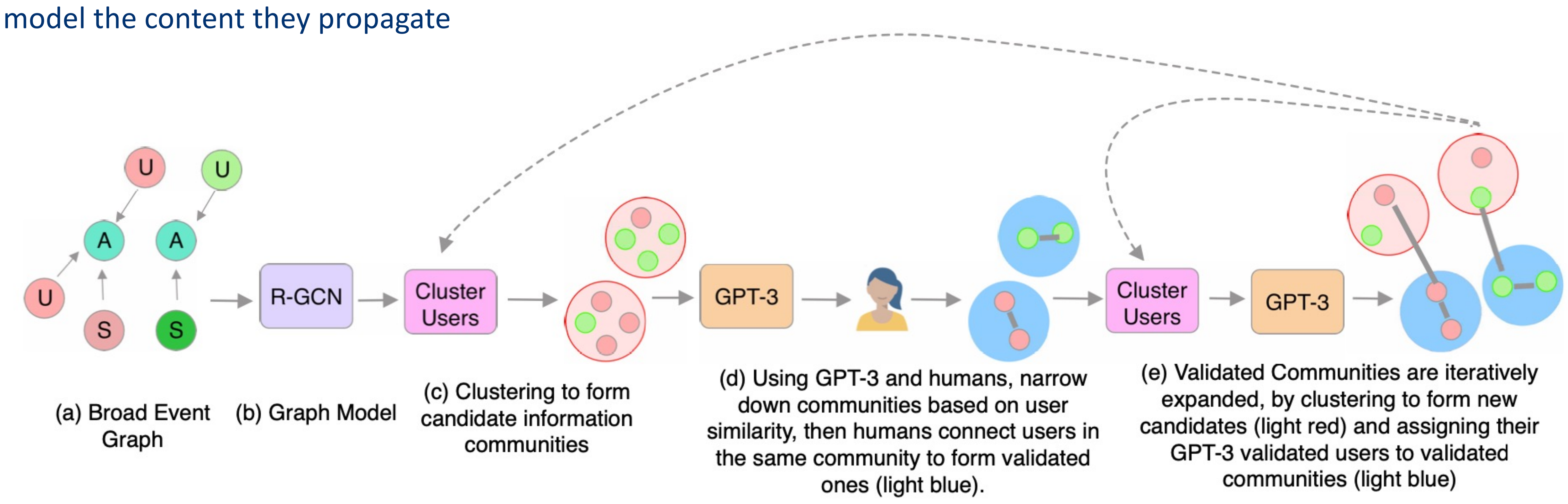

Fig 1: Framework Overview

(5) Results and Discussion

RQ1: Do Interactions Help?

- Evaluate on Black Lives Matter data
- Evaluation setting: No test data, users, or topics seen at test time, hard!
- Tab 1: Human interactions lead to significant improvements
- ~33% improvement on FN, ~40% on Bias

Which users have the same perspectives?
User 1: This user is discussing the Black Lives Matter protest and their perspectives is that these leaders are stealing money from the organization.
User 2: This user is discussing the death of a black man in Seattle who was shot by police. Their perspective appears to be critical of the Black Lives Matter movement, and suggest it is hypocritical for the movement to not be speaking out about this man's death.
User 3: This is discussing the case of Shaun King, a civil rights activist, and their perspective is that King is being unfairly attacked. They also express support for the Black Lives Matter movement.
Positive Users;;;;Negative Users:
User 1, User 2;;;;User 3

Which users have the same perspectives?
User 4: ….
User 5: …
User 6: …
Positive Users;;;;Negative Users:

Figure 2: LLM usage

RQ2: How are LLMs used?

- Fig 2: LLMs get few-shot examples based on human interaction
- LLMs tasked to either "accept" or "reject" new users into the human formed community

Future Work

- Building larger and better communities
- Application on other social media tasks
- General framework to use human insight to help LLMs

Model	FN Acc	FN F1	Bias Acc	Bias F1	# Users; # Sources	# Edges	# Inter-actions
Baseline: (Mehta et al., 2022)	41.89	28.48	46.79	27.43	-	-	-
Graph Only: High Purity 2 Communities (Comms.)	43.01	28.85	46.15	28.59	25; 25	1,200	-
Graph Only: High Purity 4 Communities (Comms.)	41.89	27.23	48.71	21.83	-	-	-
LLM Only: 2 Comms, 2 Expansion Rounds	42.70	28.05	45.01	27.84	38; 63	494	-
LLM Only: 4 Comms, 2 Expansion Rounds	42.70	28.62	39.50	33.22	69; 56	1,791	-
LLM Only: 6 Comms, 2 Expansion Rounds	40.54	26.88	37.03	29.22	73; 63	1,612	-
LLM + Humans: 2 Comms, 2 Expansion Rounds	**52.51**	**38.03**	44.23	33.40	25; 26	367	1
LLM + Humans: 2 Comms, 4 Expansion Rounds	46.36	35.03	**49.35**	**45.13**	72; 56	1,087	1
LLM + Humans: 4 Comms, 2 Expansion Rounds	43.01	32.36	47.43	32.00	55; 43	808	2
LLM + Humans: 6 Comms, 2 Expansion Rounds	41.34	32.36	48.07	33.91	82; 61	1,696	3

Table 1: Results

PURDUE UNIVERSITY

CERIAS