Towards Fine-Grained Reasoning for Fake News Detection

Yiqiao Jin[1], Xiting Wang[2]*, Ruichao Yang[3], Yizhou Sun[1], Wei Wang[1], Hao Liao[4], Xing Xie[2]

[1] University of California, Los Angeles
[2] Microsoft Research Asia
[3] Hong Kong Baptist University
[4] Shenzhen University

* Corresponding author

Fake News

Fake news is “intentionally and verifiably false news published by a news outlet”[1]

Fake news are produced because

- Fake news attracts more readers
- Anonymous users intentionally disseminate noisy and misleading information

The widespread of fake news can mislead the public, and “produce unjust political, economic, or psychological profit for some parties”[2]

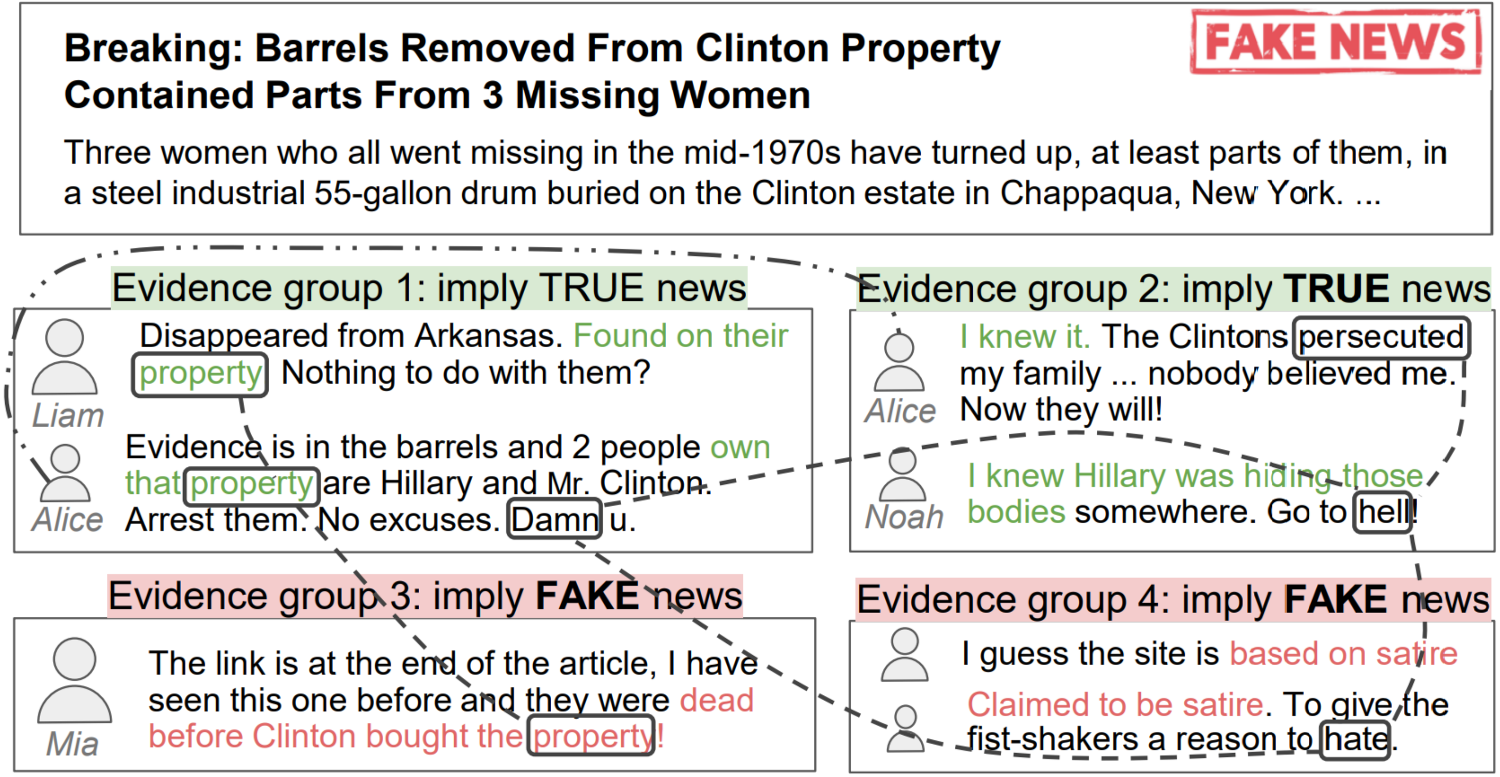
[1]: Peckham, Oliver. *AI Squares Against Fake News*, https://www.datanami.com/2019/09/02/ai-squares-off-against-fake-news/. September 2019
[2]: Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. *Truth of varying shades: Analyzing language in fake news and political fact-checking.* EMNLP 2017, pages 2931–2937.

Reasoning

Reasoning is the process of forming inferences from facts or premises. Detection of fake news often requires sophisticated reasoning skills, such as logically combining information by considering word-level subtle clues.

To reason about the authenticity of a news article, humans perform <u>fine-grained</u> analysis for identifying subtle (e.g., word-level) clues, and connect <u>different types of clues</u> (e.g., textual and social) to draw conclusions.

Motivating Example



Issues about Current Works

Works that approach the task from the perspective of reasoning are still lacking.

Explainability

Most existing works on fake news detection either do not provide explanations or enable explainability for a small part of the model.

Accuracy

Existing methods lack such fine-grained reasoning capability, either do not model the interactions between different types of evidence, or model them at a coarse-grained (e.g., sentence or post) level.

Issues about Current Works

Do not provide sound explanations for the detection results
Only enable explainability for a small part of the model, e.g., the attention layer
The major part of the model, e.g. the overall workflow, remains obscure to humans.
This prevents us to better understand & trust the model, or steer the model for performance refinement
Existing models on **Explainable Detection**: dEFEND GCAN

Research Question

RQ1. Can the model be designed by following the human's information processing model?

RQ2. Can human knowledge about which evidence (e.g. posts and users) is important be better incorporated?

RQ3. How does one achieve fine-grained modeling of different types of subtle clues?

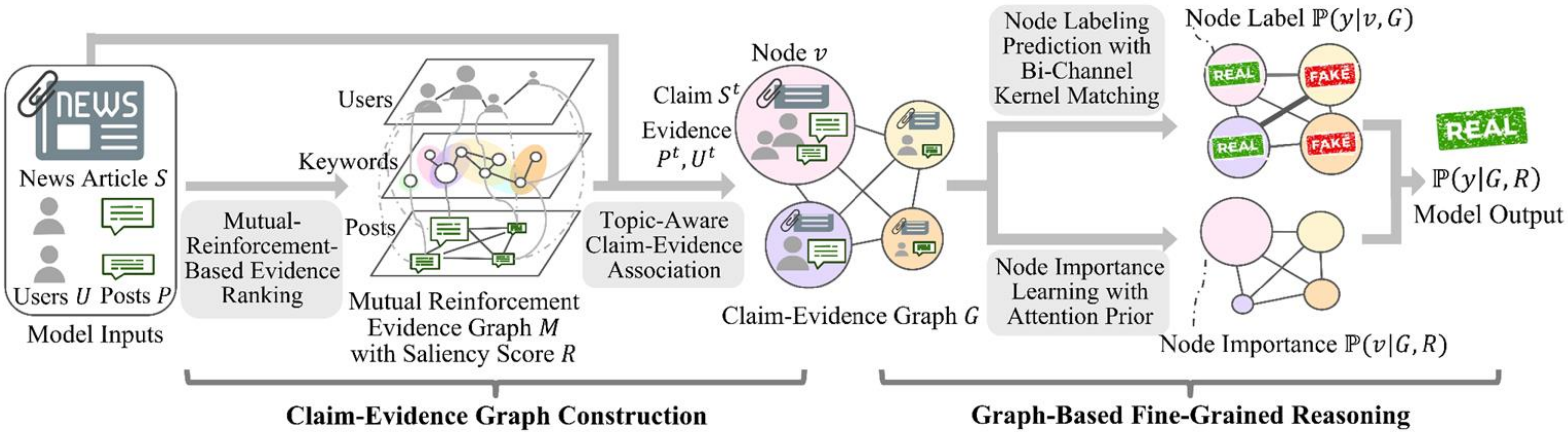
Our Contribution

C1. Design a **Fine**-grained **r**easoning framework for **Fa**ke news dete**ct**ion (**FinerFact**) by following the human's information-processing model.

C2. Propose a **mutual-reinforcement-based method** for evidence ranking, which enables us to better incorporate prior human knowledge about which types of evidence are the most important.

C3. Design a **prior-aware bi-channel kernel graph network** to achieve fine-grained reasoning by modeling different types of subtle clues.

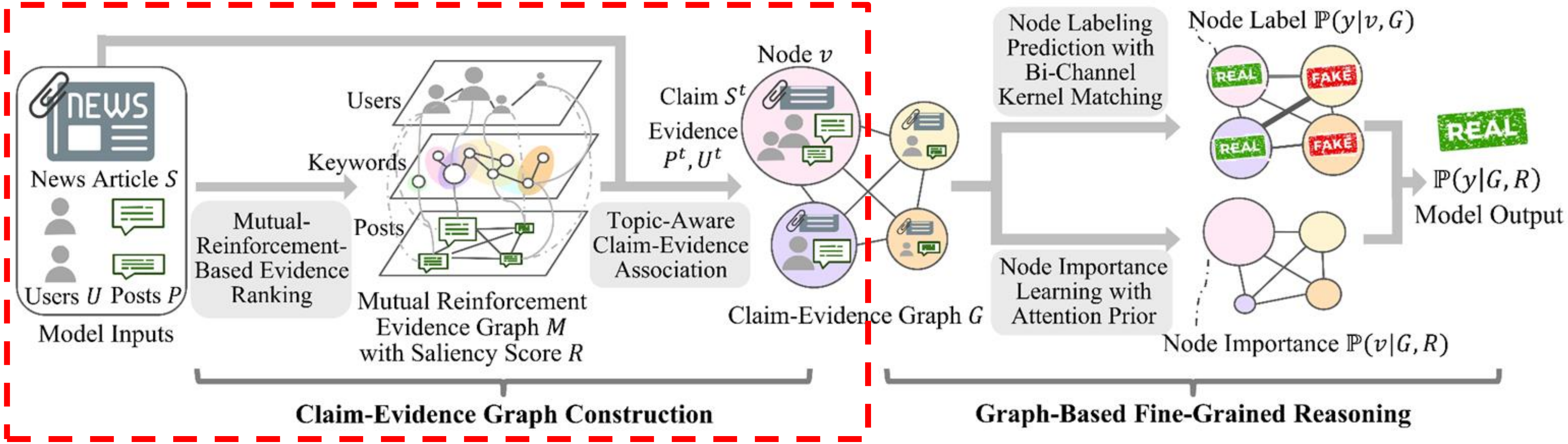
Proposed Method



RQ1. Can the model be designed by following the human's information processing model?

FinerFact detects fake news by better reflecting the logical process of human thinking, which enhances interpretability and provides the basis for incorporating human knowledge.

Proposed Method

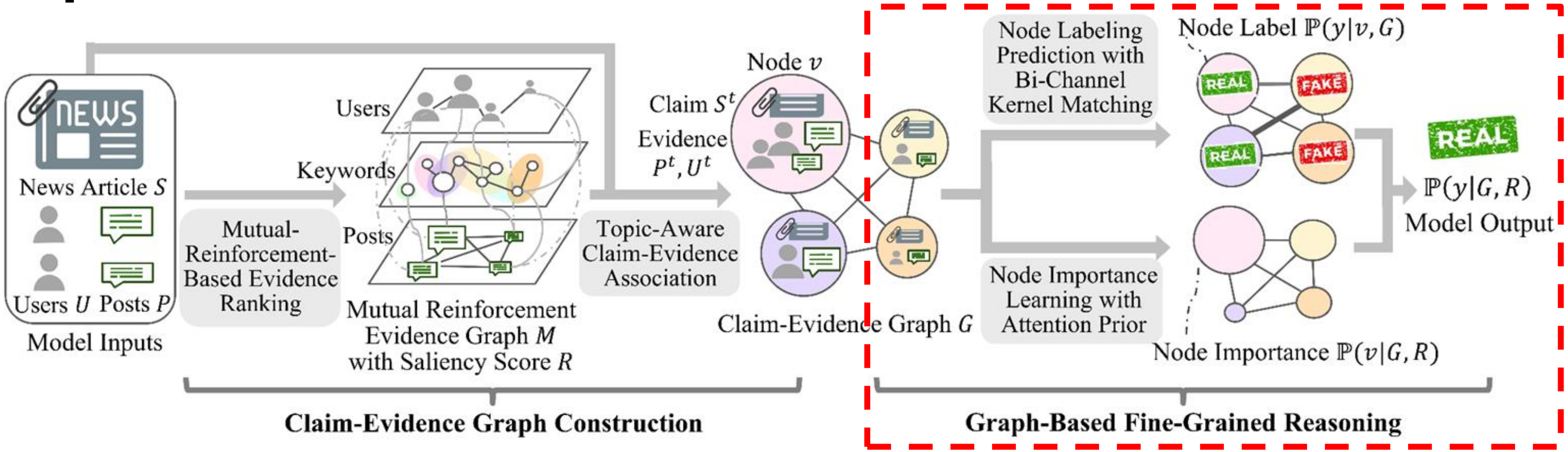


<u>RQ2</u>. Can human knowledge about which evidence (e.g. posts and users) is important be better incorporated?

Claim-Evidence Graph Construction

Corresponds to the <u>*storage*</u> sub-process of the human's information-processing model, in which people select the most important pieces of information and build their in-between associations to store them in the memory. Essential for filtering noise, organizing facts, and speeding up the fine-grained reasoning process at the later stage.

Proposed Method

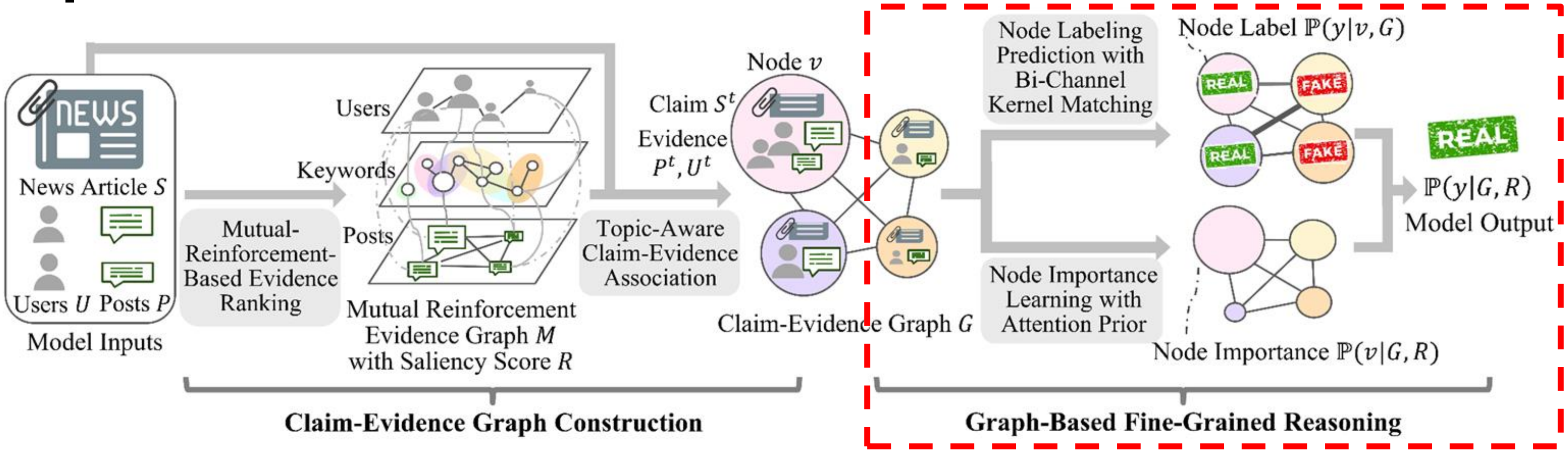


<u>**RQ3**</u>**.** How does one achieve fine-grained modeling of different types of subtle clues?

Graph-Based Fine-Grained Reasoning

Corresponds to the <u>*retrieval*</u> sub-process of the human's information-processing model, in which people reactivate specific pieces of information based on their associations for decision making. This enables fine-grained modeling of evidential relations by considering subtle clues.

Proposed Method



Node label prediction with bi-channel kernel matching
Node importance learning with attention priors

$$\mathbb{P}(y \mid G, R) = \sum_{v \in G} \underbrace{\mathbb{P}(y \mid v, G)}_{\text{Node label prediction}} \underbrace{\mathbb{P}(v \mid G, R)}_{\text{Node importance learning}}$$

Evaluation

Datasets

- PolitiFact and GossipCop.
- Contain news articles and the social context information about the news.
- News Labels are provided by journalists and domain experts.
- Labels can be fake (y = 1) or real (y = 0).

Baselines

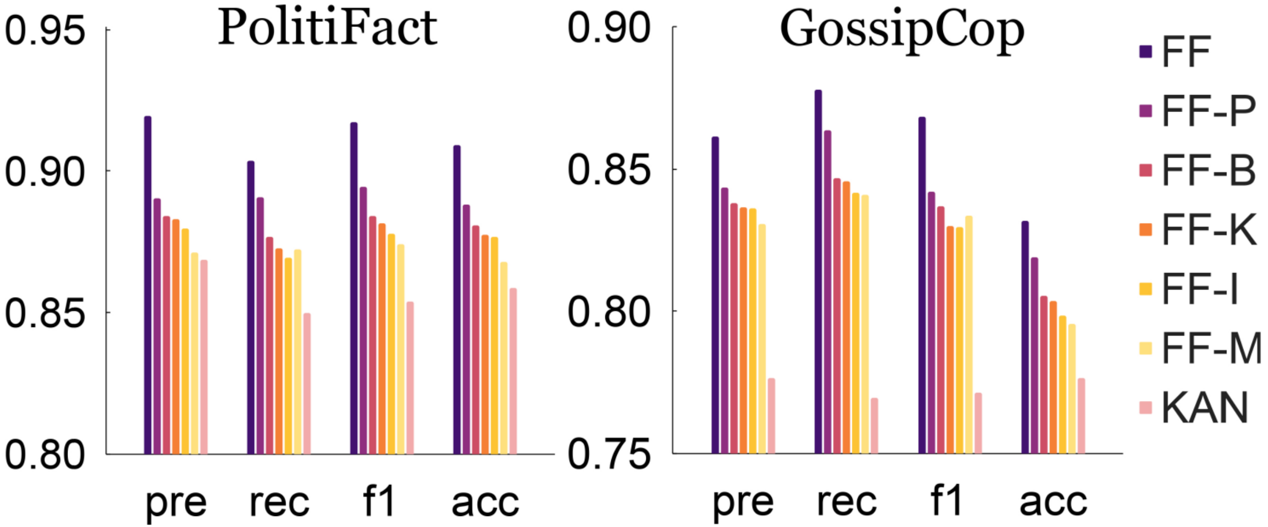
- *content-based methods* (**G1**): leverage the textual or visual content of the news.
- *knowledge-aware methods* (**G2**): leverages auxiliary knowledge such as knowledge graphs and social information about the online posts.

Performance

		PolitiFact					GossipCop				
		Pre	Rec	F1	Acc	AUC	Pre	Rec	F1	Acc	AUC
G1	SVM	0.7460	0.6826	0.6466	0.6694	0.6826	0.7493	0.6254	0.5955	0.6643	0.6253
	RFC	0.7470	0.7361	0.7362	0.7406	0.8074	0.7015	0.6707	0.6691	0.6918	0.7389
	DTC	0.7476	0.7454	0.7450	0.7486	0.7454	0.6921	0.6922	0.6919	0.6959	0.6929
	GRU-2	0.7083	0.7048	0.7041	0.7109	0.7896	0.7176	0.7079	0.7079	0.718	0.7516
G2	B-TransE	0.7739	0.7658	0.7641	0.7694	0.8340	0.7369	0.7330	0.7340	0.7394	0.7995
	KCNN	0.7852	0.7824	0.7804	0.7827	0.8488	0.7483	0.7422	0.7433	0.7491	0.8125
	GCAN	0.7945	0.8417	0.8345	0.8083	0.7992	0.7506	0.7574	0.7709	0.7439	0.8031
	KAN	0.8687	0.8499	0.8539	0.8586	0.9197	0.7764	0.7696	0.7713	0.7766	0.8435
Ours	FinerFact	**0.9196**	**0.9037**	**0.9172**	**0.9092**	**0.9384**	**0.8615**	**0.8779**	**0.8685**	**0.8320**	**0.8637**
	Impv.	+5.1%	+5.4%	+6.3%	+5.1%	+1.9%	+8.5%	+10.8%	+9.7%	+5.5%	+2.0%

Our kernel-attention-based approach can better model the interactions between news articles and evidence.
Methods that incorporate external knowledge (**G2**) generally perform better than content-based methods (**G1**), which illustrates the usefulness of external knowledge in fake news detection.

Ablation Study



FF-P removes the attention prior when learning node importance;

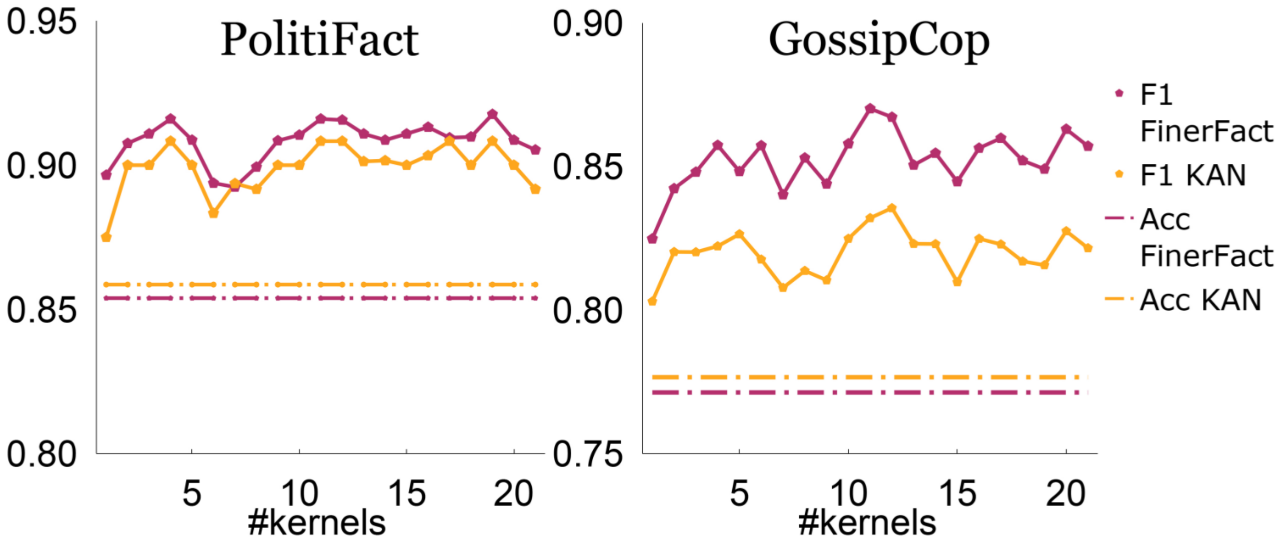
FF-B eliminates bi-channel reasoning by removing the user-based reasoning channel;

FF-K replaces kernel-based representation learning with GNN-based aggregation scheme;

FF-I excludes node importance learning and assigns an equal weight to every node;

FF-M eliminates mutual-reinforcement-based evidence ranking.

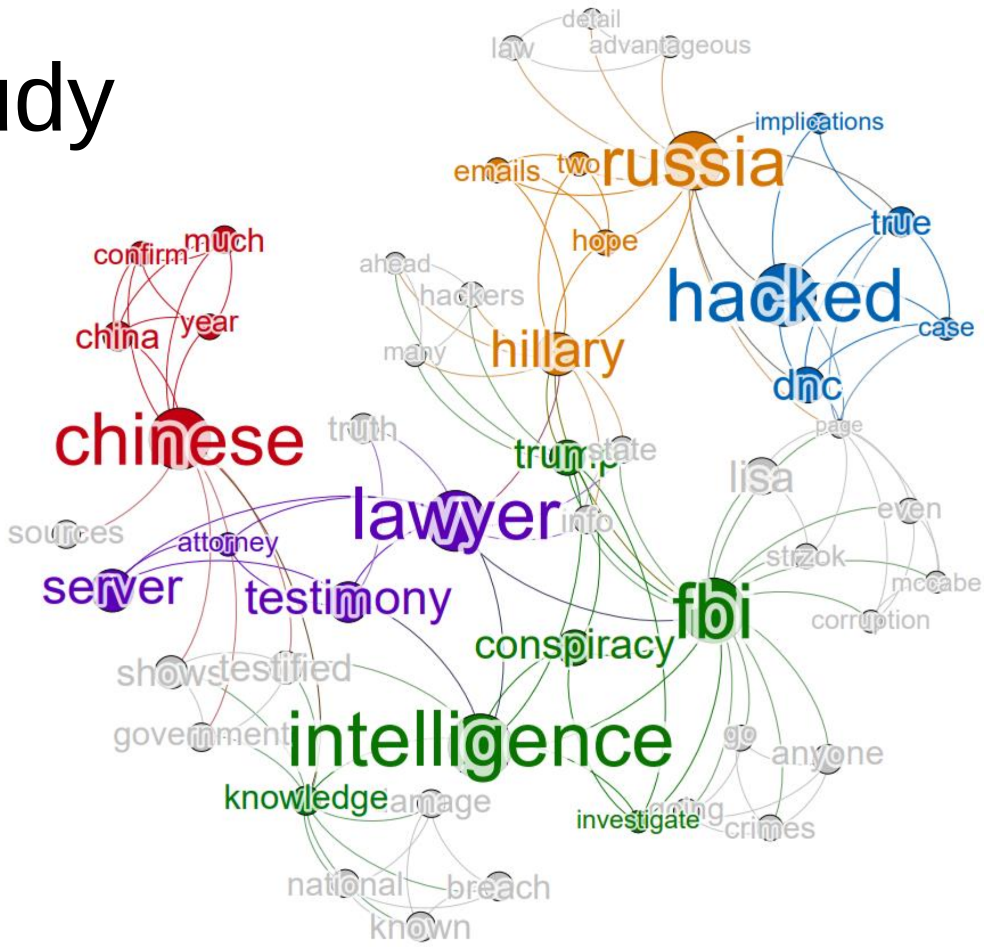
Study on Number of Kernels



FinerFact consistently outperforms the best baseline KAN with varying numbers of kernels, which demonstrates its robustness.

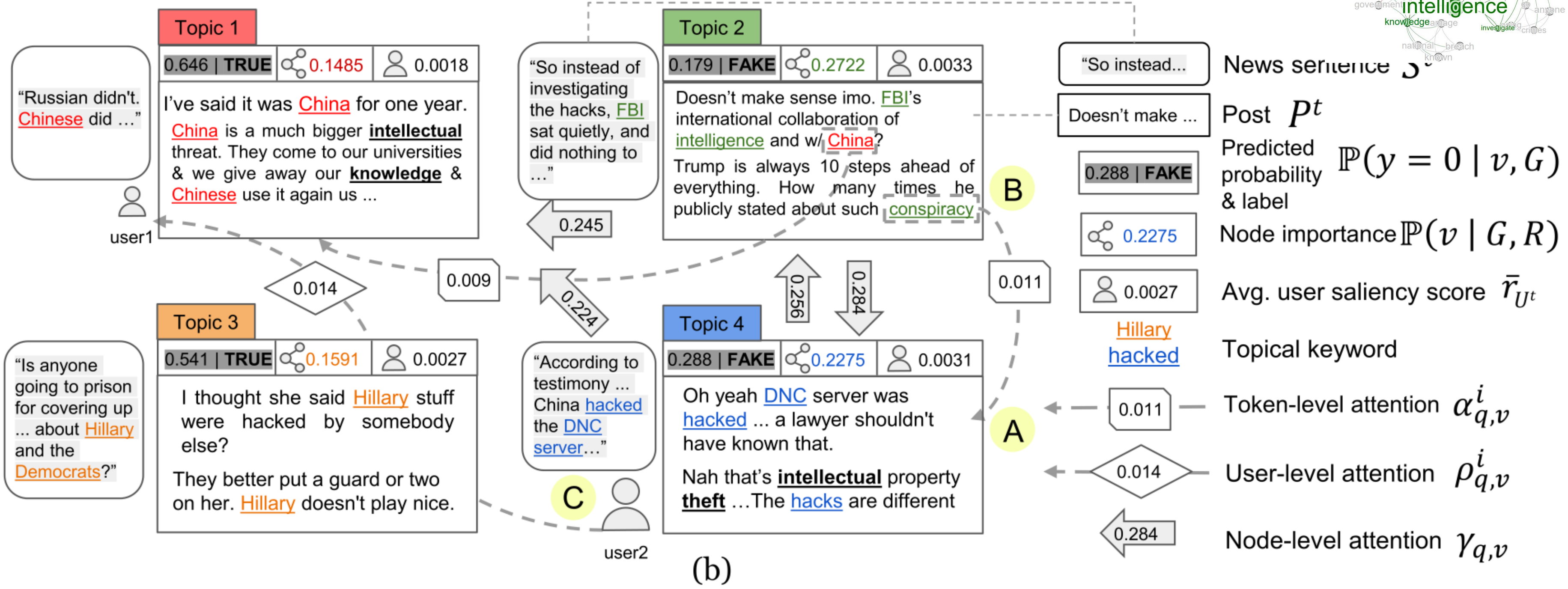
Best performance with around 11 kernels. More kernels does not necessarily lead to better performance due to overfitting.

Case Study

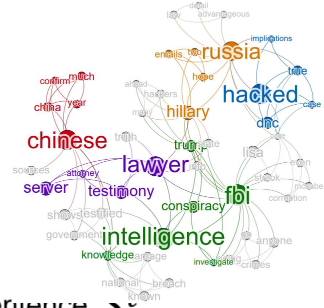


detail
law
advantageous
implications
emails
two
russia
true
much
confirm
hope
ahead
hacked
hackers
year
case
china
many
hillary
dnc
page
chinese
truth
trump
state
lisa
lawyer
even
info
sources
attorney
strzok
mccabe
server
testimony
fbi
corruption
conspiracy
shows
testified
government
intelligence
go
anyone
knowledge
damage
going
investigate
crimes
national
breach
known

Case Study



Topic 1
0.646 | TRUE
0.1485
0.0018
"Russian didn't. Chinese did …"
user1
I've said it was China for one year.
China is a much bigger intellectual threat. They come to our universities & we give away our knowledge & Chinese use it again us ...
Topic 2
0.179 | FAKE
0.2722
0.0033
"So instead of investigating the hacks, FBI sat quietly, and did nothing to …"
Doesn't make sense imo. FBI's international collaboration of intelligence and w/ China?
Trump is always 10 steps ahead of everything. How many times he publicly stated about such conspiracy
B
0.245
0.009
0.014
0.224
0.256
0.284
0.011
Topic 3
0.541 | TRUE
0.1591
0.0027
"Is anyone going to prison for covering up … about Hillary and the Democrats?"
I thought she said Hillary stuff were hacked by somebody else?
They better put a guard or two on her. Hillary doesn't play nice.
"According to testimony … China hacked the DNC server…"
C
user2
Topic 4
0.288 | FAKE
0.2275
0.0031
Oh yeah DNC server was hacked ... a lawyer shouldn't have known that.
Nah that's intellectual property theft …The hacks are different
A
(b)
"So instead...
News sentence S^t
Doesn't make ...
Post P^t
0.288 | FAKE
Predicted probability & label $\mathbb{P}(y = 0 \mid v, G)$
0.2275
Node importance $\mathbb{P}(v \mid G, R)$
0.0027
Avg. user saliency score $\bar{r}_{U^t}$
Hillary
hacked
Topical keyword
0.011
Token-level attention $\alpha^i_{q,v}$
0.014
User-level attention $\rho^i_{q,v}$
0.284
Node-level attention $\gamma_{q,v}$



russia
hacked
hillary
chinese
lawyer
fbi
server
testimony
conspiracy
intelligence
knowledge