Joint Estimation of User And Publisher Credibility for Fake News Detection

Presenter:
Sriram Srinivasan

Authors:
Sriram Srinivasan, Rajdipa Chowdhury*, and Lise Getoor*

* Equal contribution

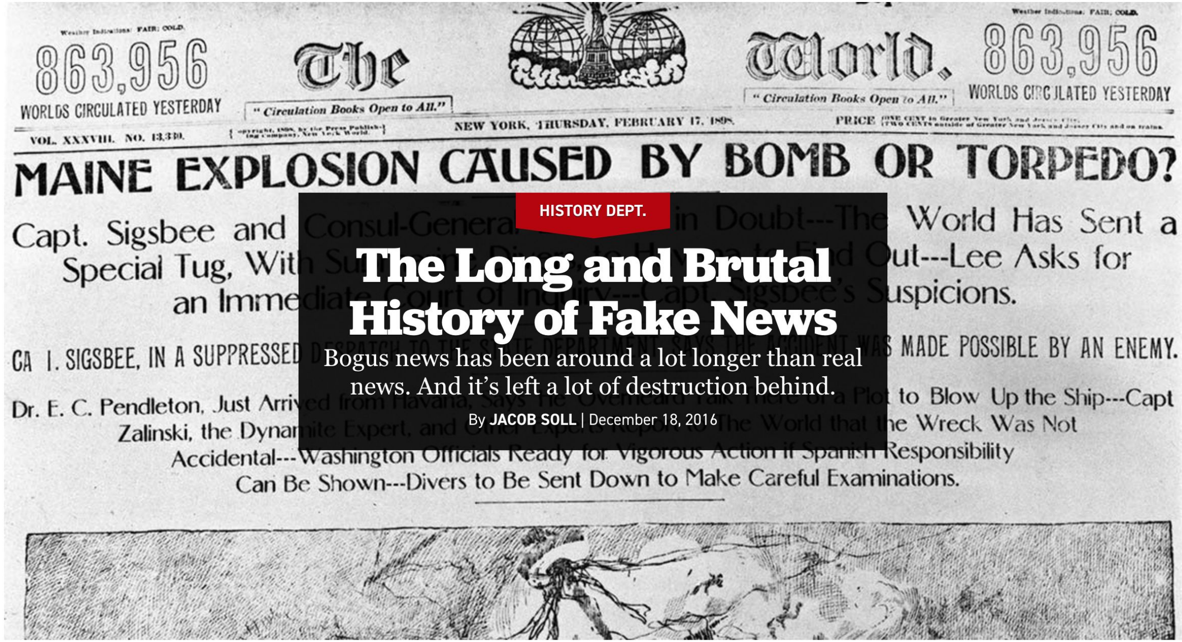
Fake News

News consisting of deliberate disinformation

• • •

Fake News

Not a new problem



https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535

Political or economic gain

Impact of Social Media

Easy access

Cheap

Everyone is a publisher

Easy to spread

High engagement

No moderator

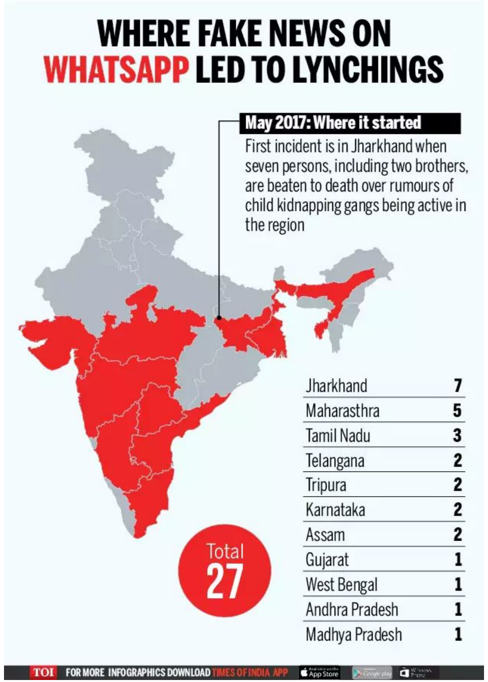
Everyone is a distributor

Impact of Social Media

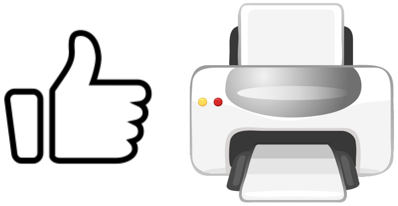
Inform/Misinform millions in a minutes

Serious impact in the society

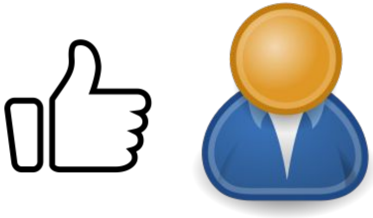
Example:



Contributions



Identify credibility of publisher publishing the news



Identify credibility of user sharing the news

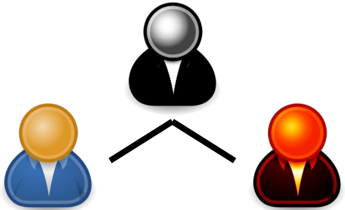
Identify if news is fake or not

Joint inference

Empirically show the effectiveness of the approach

Previous Works

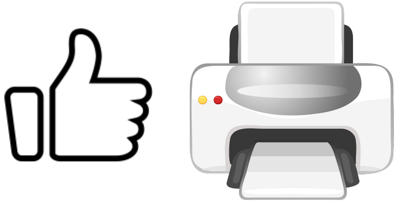
- Content-based approach
  - Wang et al. In *Association for Computational Linguistics* (2017)
  - Reed et al. In *Values and knowledge* (2018)



- Social context-based approach
  - Wu et al. In *Web Search and Data Mining* (2018)
  - Yang et al. In *National Conference on Artificial Intelligence* (2019)
- Hybrid approach
  - Volkova et al. In *The Web Conf.* (2018)
  - Shu et al. In *Web Search and Data Mining* (2019)

Publisher Credibility

Credible publisher $\Rightarrow$ Credible news



External source

Media Bias / Fact Check

https://mediabiasfactcheck.com/

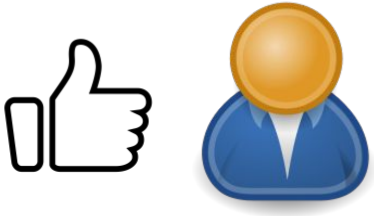
Publisher bias

Factual reporting score

Learn from fake news publishing behaviour

User Credibility

Credible user ⇒ Shares credible news



Different types of users

Fact check before share

Share their bias

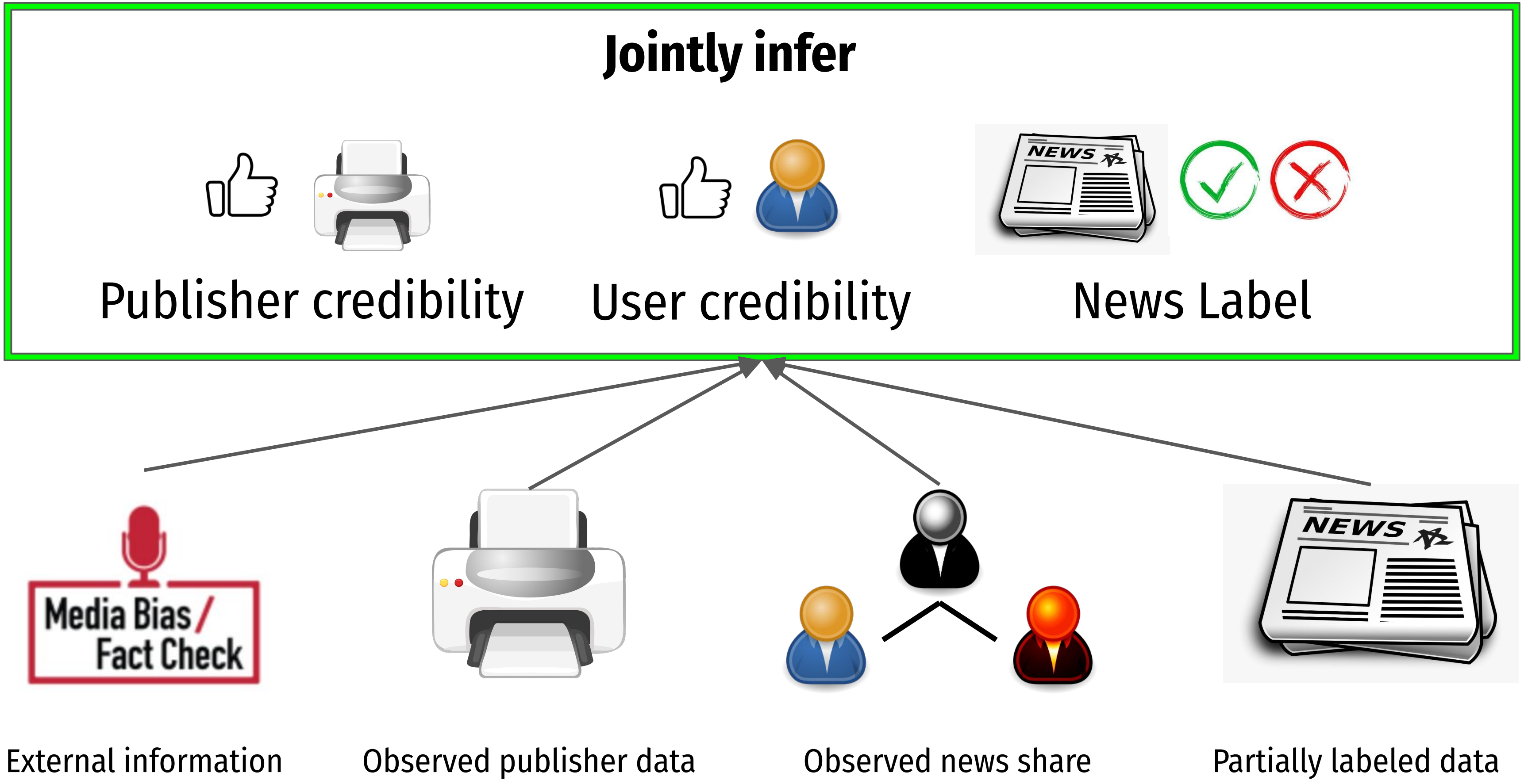
Random share

Malicious uses

Credibility score 1 → **Credibility score 0**

Learn from fake news sharing behaviour

Joint Estimation



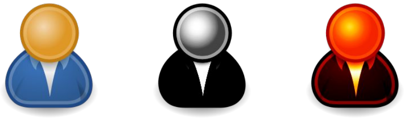
Probabilistic Soft Logic (PSL)

- A statistical relational learning framework
- Model defined via weighted first-order logical rules
- Generates a specific type of Markov random field (MRF): hinge-loss MRF (HL-MRF)*
- Efficient MAP inference

*Stephen H. Bach, Matthias Broecheler, Bert Huang, and Lise Getoor. Hinge-Loss Markov Random Fields and Probabilistic Soft Logic. JMLR 2017

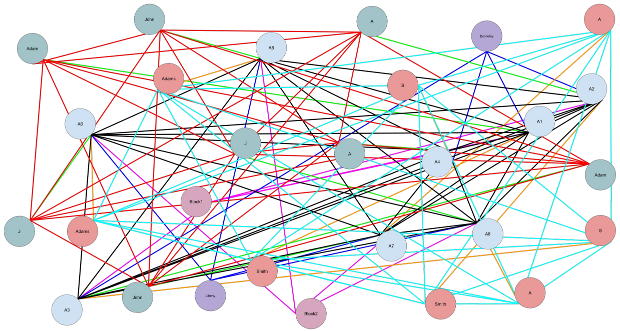
Probabilistic Soft Logic (PSL)

w: UserShare(U, N) $\land$ ¬FakeNews(N) $\Rightarrow$ UserCredible(U)

Users (U) = 

News (N) = 

Instantiate rules with data



$$= \quad P(Y|X) \propto exp(\sum_{i=1}^{|G|} w_i \phi_i)$$

Empirical Evaluation

Datasets

- Data obtained from the TriFN paper by Shu et al. (2019)
- Contains news content and social-context information
- Post information from Twitter

Platform	Politifact	Buzzfeed
Users	23,865	15,257
Publishers	88	27
Engagements	37,259	25,240
News (Fake:Real)	120:120	91:91

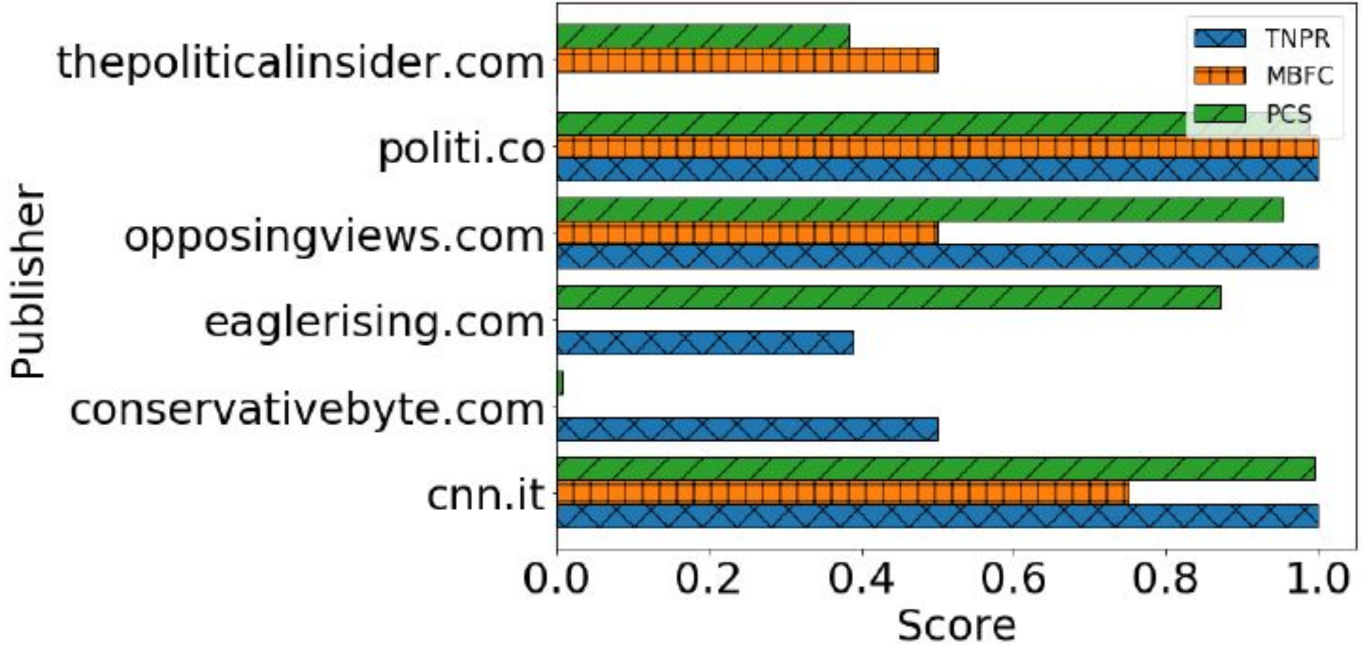
Performance

	Politifact		Buzzfeed	
Metrics	**TriFN**	**Our Approach**	**TriFN**	**Our Approach**
Accuracy	0.878	**0.913**	**0.864**	**0.858**
Precision	0.867	**0.879**	**0.849**	0.787
Recall	0.893	**0.961**	0.893	**0.979**
F1	0.880	**0.917**	**0.870**	**0.870**

Bold implies best or not significantly different from best

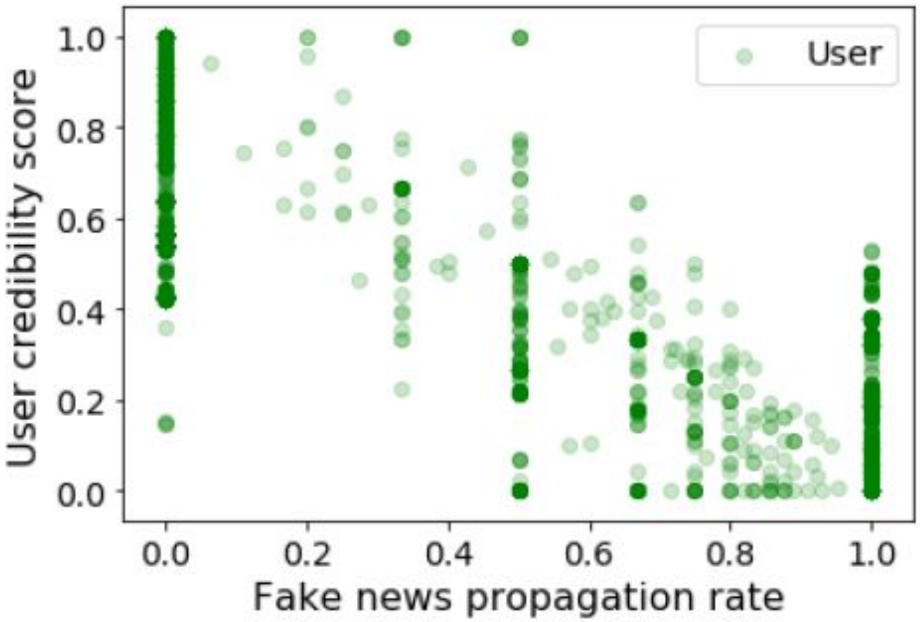
Publisher Credibility

- TNPR: True news propagation rate
- MBFC: Score from “media bias fact check” website
- PCS: Publisher credibility score

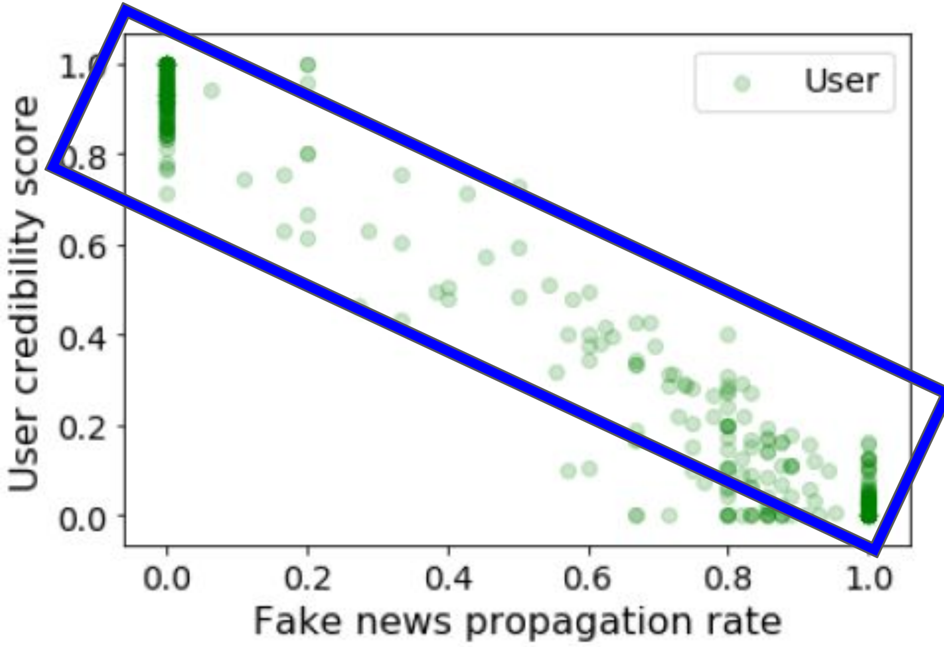


User Credibility

All users



Users with at least 5 shares



Conclusion and Future work

- Conclusion:
  - Fake news detection
  - Meaningful scores for publisher and user credibility
  - Show effectiveness of the approach on realworld datasets

- Future work
  - Augment content features and knowledge graphs to improve fake news detection
  - Analysis to determine minimum labeled data required
  - Early fake news detection to stop spread

Thank You

Questions?