

Learning with Weak Social Supervision for Disinformation Detection

Kai Shu

Department of Computer Science
Illinois Institute of Technology

<http://www.cs.iit.edu/~kshu>

kshu@iit.edu

Learning with Weak Social Supervision

Representation Learning

- Transfer Representation Learning (WSDM 18)
- Unsupervised Learning (AAAI 19, Preprint in CIKM 20)
- Adversarial Learning (ICDM 18, Preprints in AAAI 21)
- Reinforcement Learning (Preprints in WWW 21)
- Weak Supervision Learning (SIGIR 20, ECML-PKDD 20)

Social Network Analysis

- Multi-modal Networks (MICCAI 18)
- Temporal Networks (CIKM 14, TIST 19, ICWSM 20, CIKM 20)
- Location-based Social Networks (WWW 17)
- Heterogeneous Networks (WSDM 19)

Applications

- Disinformation Detection (SIGKDD 19, WSDM 19, ECML-PKDD 20 Preprints in AAAI 21)
- Social Recommendations (WSDM 18, WSDM 19, AAAI 18)
- Data Generation (ICDM 18, EMNLP 20)
- AI for social good

Develop novel machine learning algorithms to derive **supervision** from **noisy, heterogenous** and **linked** social media data

10 Wonderful Examples Of Using Artificial Intelligence (AI) For Good



Bernard Marr Contributor ⓘ

Enterprise Tech

Spot “Fake News”

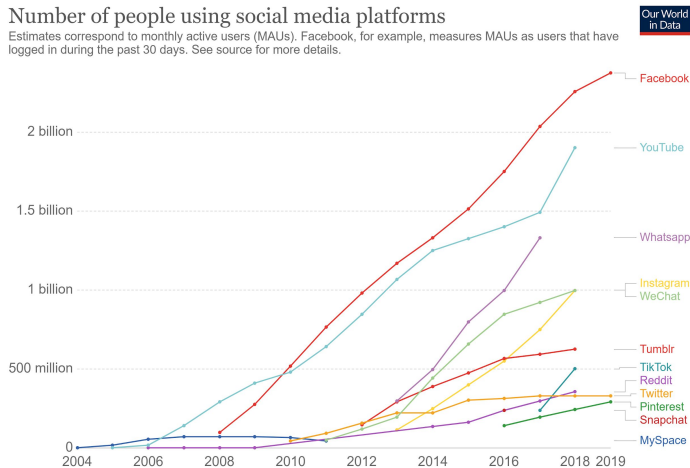
It's true: AI is the engine that pushes "fake news" out to the masses, but Google, Microsoft, and grassroots effort Fake News Challenge are using [AI \(machine learning and natural language processing\)](#) to assess the truth of articles [automatically](#). Due to the trillions of posts, Facebook must monitor and the impossibility of manually doing it, the [company also uses artificial intelligence](#) to find words and patterns that could indicate fake news. Other tools that rely on AI to analyze content include Spike, Snopes, Hoaxy, and more.

Social Media for Information Sharing

- People are increasingly using social media for information sharing, social networking, etc
- 68% of Americans get news on social media

Number of people using social media platforms

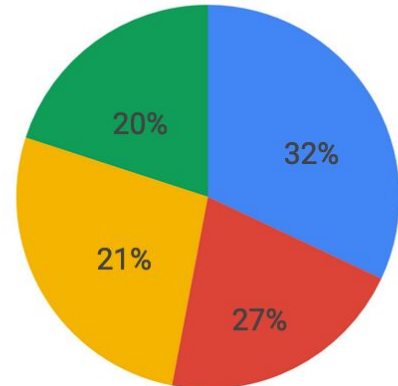
Estimates correspond to monthly active users (MAUs). Facebook, for example, measures MAUs as users that have logged in during the past 30 days. See source for more details.



Source: Statista and TNW (2019)

CC BY

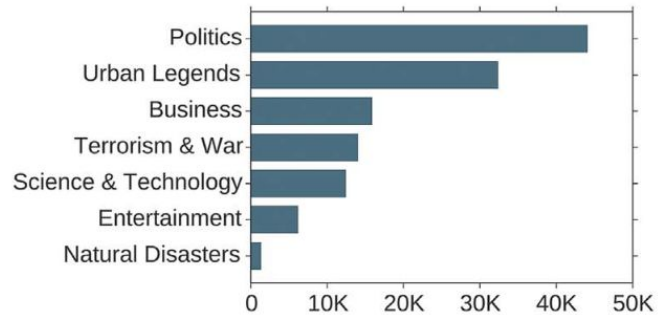
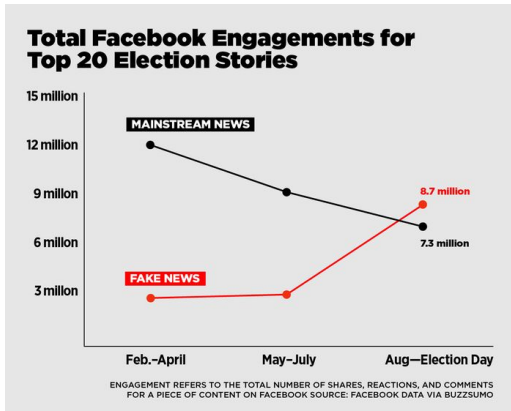
• Never • Hardly Ever • Sometimes • Often



<https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>

Disinformation is Rampant on Social Media

- **Disinformation** is false information with a bad intention aiming to mislead the public
- **Fake news** is news with intentionally false information

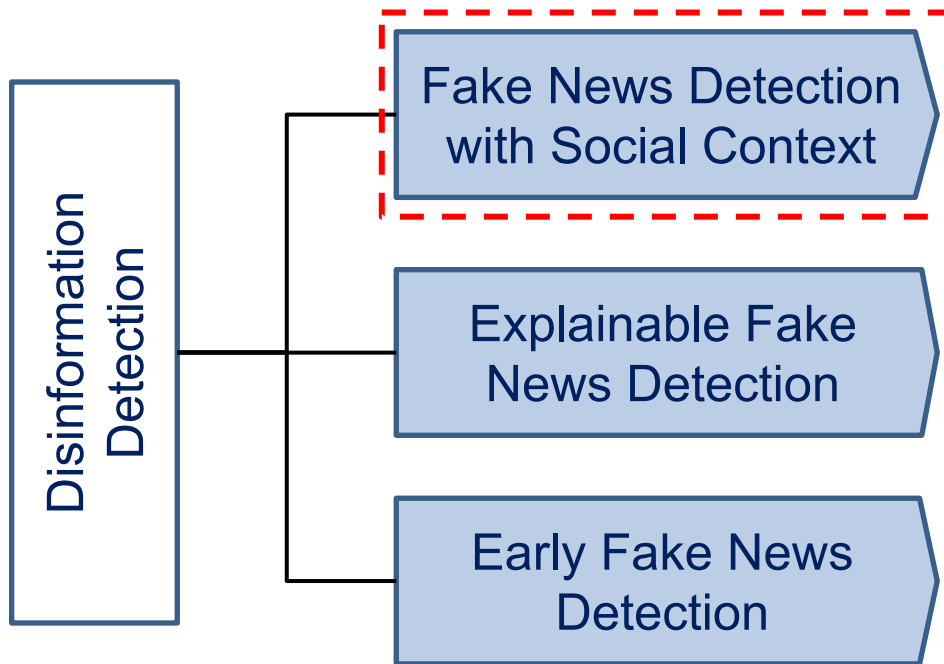


Studying Fake News

- Humans are **susceptible** to fake news
 - **Limited resources**: time, information, and expertise
 - **Confirmation Bias**: people tend to believe fake news when it confirms their pre-existing knowledge
- Fake news can have **detrimental societal effects**
 - **Misleading** people to false information
 - **Changing** the way people respond to true news
 - **Weakening** public trust in governments and journalism

Why it is so challenging

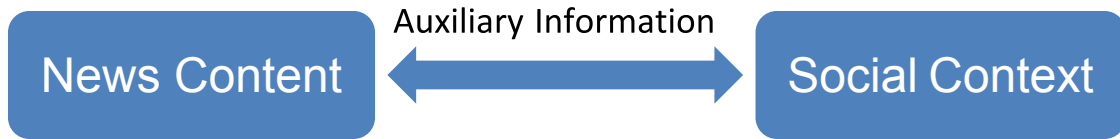
- *Fake news detection* is not just another competition
 - A competition gives a dataset with ground truth and shows who can fare best
- Fake news detection is complex in many dimensions
- We discuss some imperative challenges
 - Detection, Explainability, and Data



[Kai Shu](#), Suhang Wang, and Huan Liu. "Beyond News Content: The Role of Social Context for Fake News Detection". [WSDM 2019](#), February 11-15, 2019. Melbourne, Australia.

Motivation

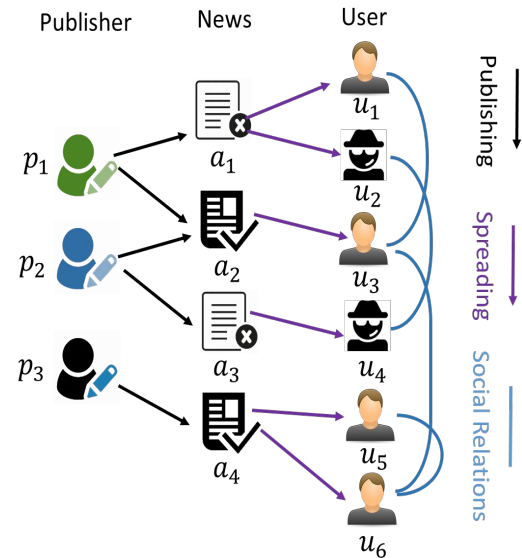
- It is **difficult** to differentiate fake news from real news using **only** news content
- **Social context** provides rich auxiliary information beyond news content



- What are the **actors** and **their relations** in social context?
- How to **model** social context to help detect fake news?

News, Actors, and Their Relations

- A typical news ecosystem with social context
 - **Entities:** publishers, news pieces, and social media users
 - **Relations:** publishing, spreading, and social relations



Modeling Social Context

- **Goal:** learn the news representations from the **heterogeneous network** for fake news prediction
- Jointly embedding news content and social context

$$\min_{\mathbf{D}, \mathbf{U}, \mathbf{V}, \mathbf{T} \geq 0, \mathbf{p}, \mathbf{q}} \left[\|\mathbf{X} - \mathbf{D}\mathbf{V}^T\|_F^2 \right]$$

Content Embedding

$$+ \alpha \|\mathbf{Y} \odot (\mathbf{A} - \mathbf{U}\mathbf{T}\mathbf{U}^T)\|_F^2$$

User-User Embedding

$$+ \beta \text{tr}(\mathbf{H}^T \mathbf{L} \mathbf{H})$$

User-News Embedding

$$+ \gamma \|\mathbf{e} \odot (\bar{\mathbf{B}}\mathbf{D}\mathbf{q} - \mathbf{o})\|_2^2$$

Publisher-New Embedding

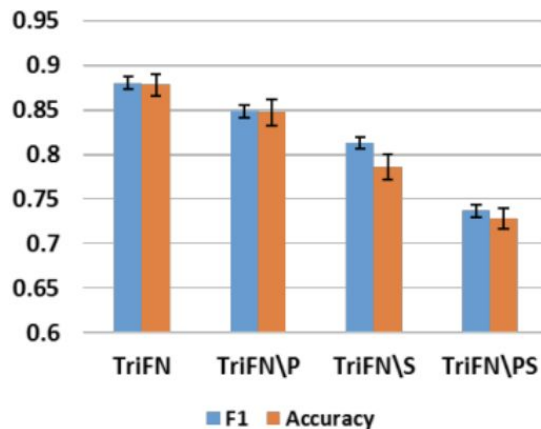
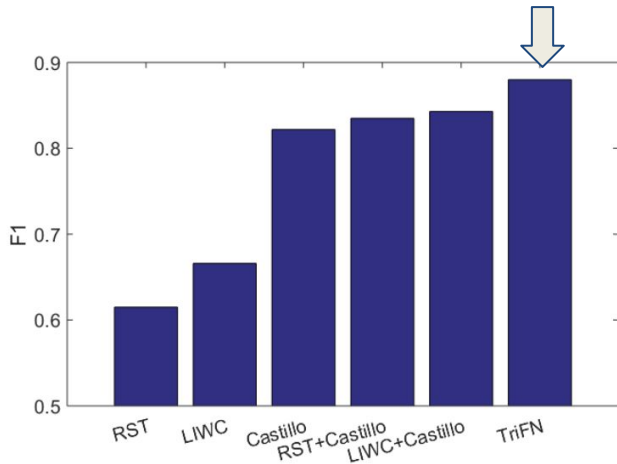
$$+ \eta \|\mathbf{D}_L \mathbf{p} - \mathbf{y}_L\|_2^2 + \lambda R$$

Fake News Classifier

Social Context

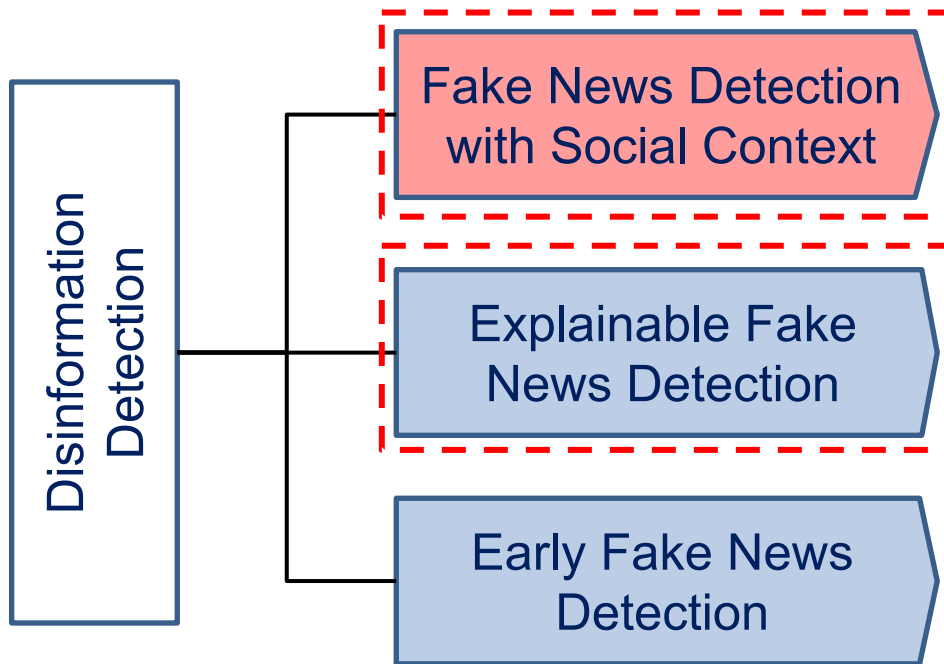
Experimental Evaluation

- The proposed model can achieve **best** performance in detecting fake news consistently
- It is necessary to model **both** news contents and social context because they contain complementary information



Summary

- ***Social context*** information brings ***additional*** signals to fake news detection
- It is important to capture the ***relations*** among publishers, news pieces, and users to detect fake news
- The proposed TriFN framework is ***effective*** to model ***tri-relationships*** through heterogeneous network embedding



[Kai Shu](#), Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. "dDEFEND: Explainable Fake News Detection", [KDD 2019](#), August 4-8, 2019. Anchorage, Alaska.

Motivation

- **Goal:** detecting fake news and explaining why it is detected as fake
 - Provide insights and knowledge to domain experts
 - Explainable features from noisy auxiliary information can further help detection performance
- Would ***comments*** be helpful to explain and detect fake news?
- How to ***model*** content-comment relations for explainable fake news detection?

Contents, Comments, and Their Relations

- News contents and user comments are **inherently related** and provided important cues for explanation and detection
 - News contents may contain false information
 - User comments have rich information from the crowd such as opinions, stances, and sentiment

Fake News

Iranian Official Drops Bombshell: Obama Secretly Gave Citizenship to 2500 Iranians as Part of Nuke Deal

By [redacted] - July 2, 2018

148 Comments

A senior Iranian cleric and member of parliament has just dropped a bombshell.

He is claiming that the Obama administration, as part of negotiating during the Iran Deal, granted U.S. citizenship to 2500 Iranians including family members of government officials.

...

There have been so many things hidden from the public about the Iran Deal if this was one more thing given up in bribe, it wouldn't be hard to believe.

Comments

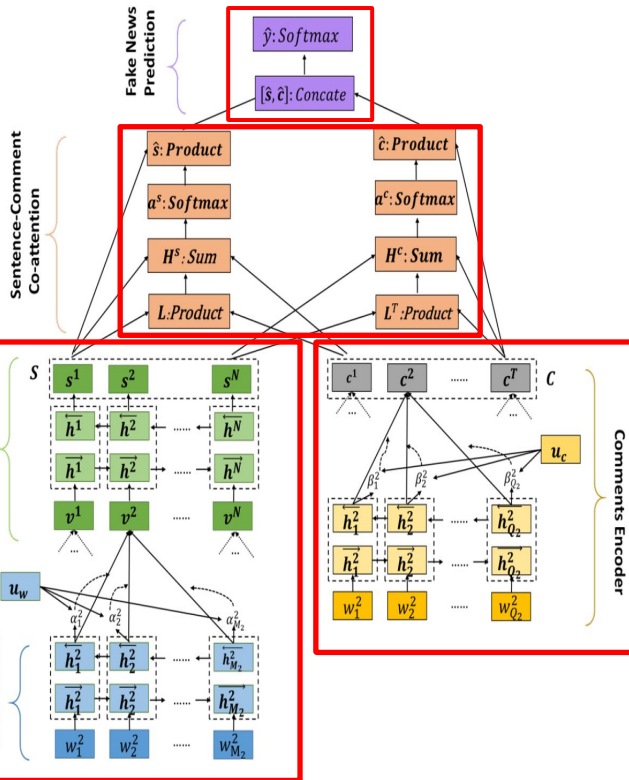
If you had done your research, you would know that the president does not have the power to give citizenship. This would have to be done as an act of congress... (0.0160)

Isn't graft and payoffs normally a offense even for a ex-president? (0.0086)

Wow! What's frightening is where will it end? We could be seeing some serious issues here. (0.0051)

Walkaway from their (0.0080)

dEFEND explains why it is detected as fake



- Learn news sentence representations through a hierarchical attention network
- Encode comment representations through a word-level attention network
- Select top explainable sentences and comments through a co-attention network
- Detect fake news with concatenated sentence and comment representations

Detection Performance

- User comments based methods are slightly more **effective** than news content based methods
- dEFEND performs the **best** among the methods using both news content and user comments

User comments

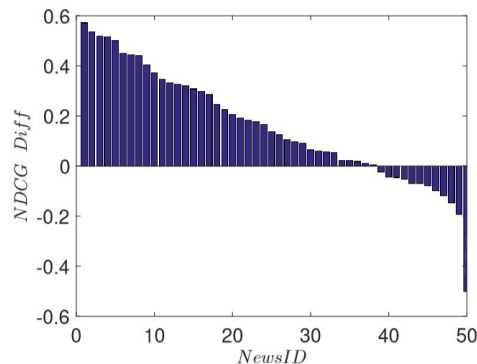
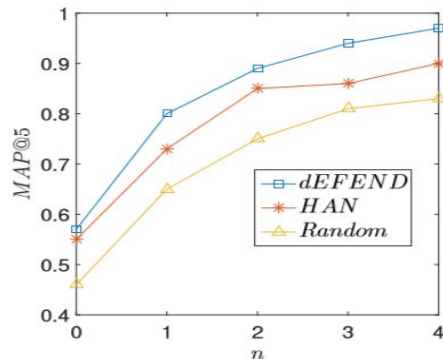
Datasets	Metric	RST	LIWC	text-CNN	HAN	TCNN-URG	HPA-BLSTM	CSI	dEFEND
PolitiFact	Accuracy	0.607	0.769	0.653	0.837	0.712	0.846	0.827	0.904
	Precision	0.625	0.843	0.678	0.824	0.711	0.894	0.847	0.902
	Recall	0.523	0.794	0.863	0.896	0.941	0.868	0.897	0.956
	F1	0.569	0.818	0.760	0.860	0.810	0.881	0.871	0.928
GossipCop	Accuracy	0.531	0.736	0.739	0.742	0.736	0.753	0.772	0.808
	Precision	0.534	0.756	0.707	0.655	0.715	0.684	0.732	0.729
	Recall	0.492	0.461	0.477	0.689	0.521	0.662	0.638	0.782
	F1	0.512	0.572	0.569	0.672	0.603	0.673	0.682	0.755

News Content

News Content + User comments

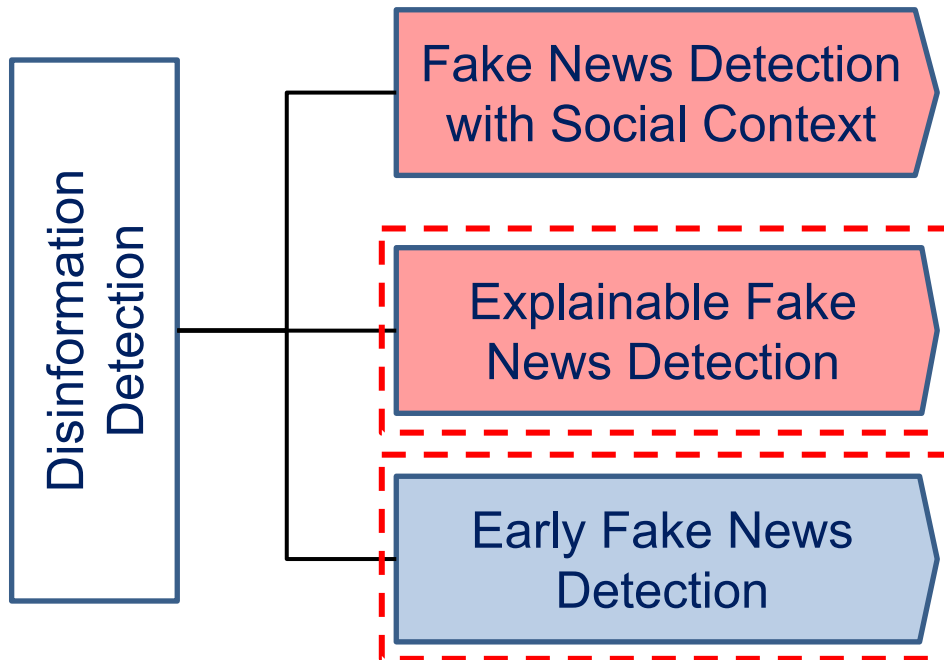
Explainability Performance

- **Contents:** dEFEND can achieve better performance to capture check-worthy sentences
- **Comments:** dEFEND can better discover explainable comments than baselines



Summary

- We study a **new** problem of explainable fake news detection
- dEFEND can **Identify** explainable news sentences and user comments for understanding why news is detected as fake
- dEFEND **achieves high accuracy** in comparison with the state-of-the-art fake news detection methods



[Kai Shu](#), Guoqing Zheng, Yichuan Li, Subhabrata Mukherjee, Ahmed Awadallah, Scott Ruston, and Huan Liu. "Early Detection of Fake News with Multi-source Weak Social Supervision", [ECML-PKDD 2020](#), September 14-18, 2020. Ghent, Belgium.

Early Fake News Detection

- Fake news can spread farther, faster, deeper, and more widely than true news
 - **Goal:** detect fake news at an **early** stage with limited labeled data
-
- Would ***social engagements*** be helpful to detect fake news early?
 - How to ***learn from*** weak social supervision for early fake news detection?

Weak Social Supervision (WSS) Can Help

- User engagements in social media provide **different** sources to derive **weak social supervision**

- **Sentiment:**

conflicting sentiments
indicate high probability
of fake news

- **Bias and Credibility:**

more biased and less
credible users are
more likely to share fake news

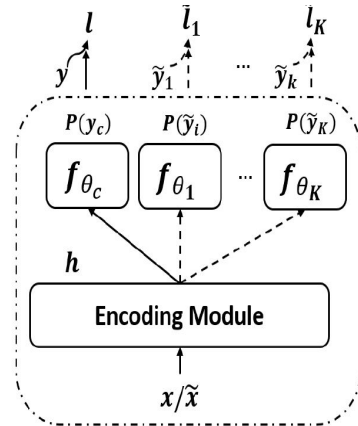


Learning with Multi-source of WSS

- **Goal: jointly** learn the **correlation** and **distinction** from clean and weak labels
 - *Shared encoder* for representation learning
 - *Separate functions* for mapping representations to clean or weak labels

$$\begin{aligned}
 \mathcal{L} = & \min_{\theta_E, \theta_c, \theta_1, \dots, \theta_K} \mathbb{E}_{(x, y) \in \mathcal{D}} \overbrace{\ell(y, f_{\theta_c}(h_{\theta_E}(x)))}^{\text{clean labels}} \\
 & + \underbrace{\sum_{k=1}^K \mathbb{E}_{(x, \tilde{y}) \in \tilde{\mathcal{D}}^{(k)}} \omega_{\alpha}(h_{\theta_E}(x), \tilde{y}) \ell(\tilde{y}, f_{\theta_k}(h_{\theta_E}(x)))}_{\text{weak labels from K sources}}
 \end{aligned}$$

label weight function



Experimental Evaluation

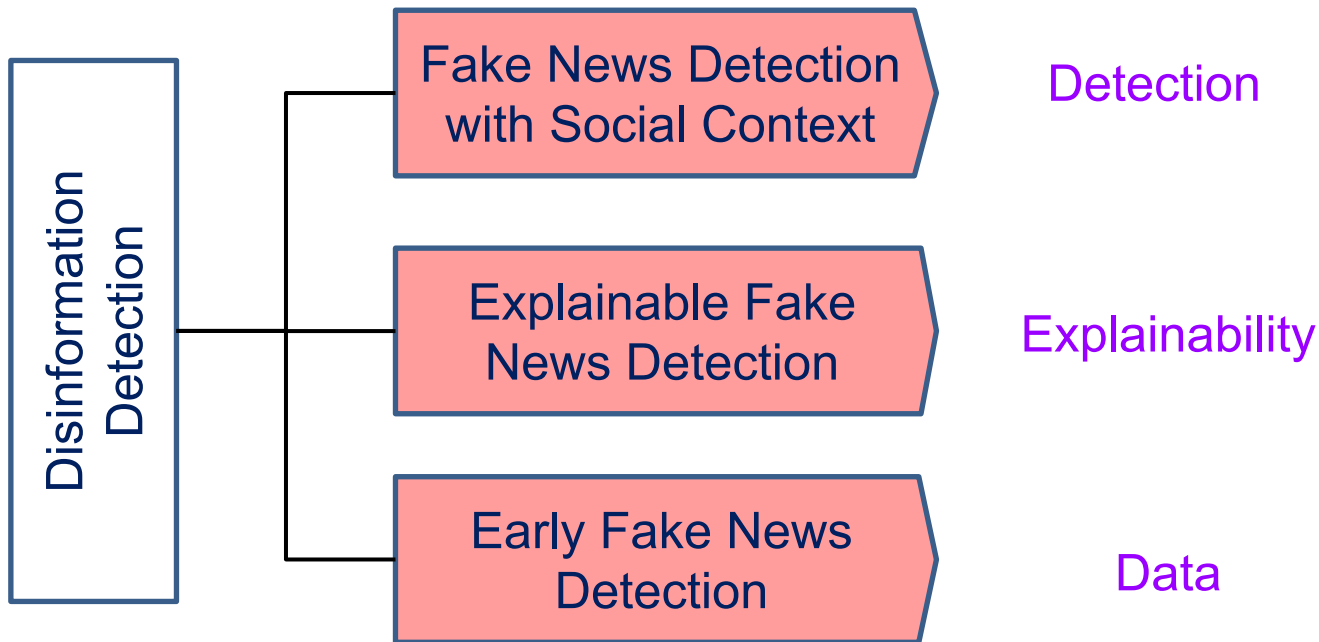
- In general, our model **MWSS** achieves the **best** performance consistently
- Training only on **clean** data performs **better** than only on **weak** data
- MWSS with **multiple** weak sources achieves **better** performance compared to that of a **single** weak source

Methods	GossipCop		PolitiFact	
	F1	Accuracy	F1	Accuracy
TCNN-URG (Clean)	0.76	0.74	0.77	0.78
EANN (Clean)	0.77	0.74	0.78	0.81
CNN (Clean)	0.74	0.73	0.72	0.72
CNN (Weak)	0.73	0.65	0.33	0.60
CNN (Clean+Weak)	0.76	0.74	0.73	0.72
CNN- <i>Snorkel</i> (Clean+Weak)	0.76	0.75	0.78	0.73
CNN- <i>L2R</i> (Clean+Weak)	0.77	0.74	0.79	0.78
CNN-MWSS (Clean+Weak)	0.79	0.77	0.82	0.82

Dataset	Sentiment	Bias	Credibility	All Sources
GossipCop	0.75/0.69	0.78/0.75	0.77/0.73	0.79/0.77
PolitiFact	0.75/0.75	0.77/0.77	0.75/0.73	0.78/0.75

Summary

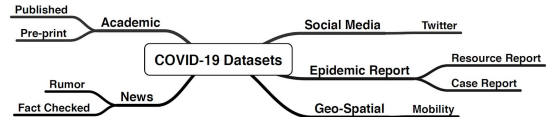
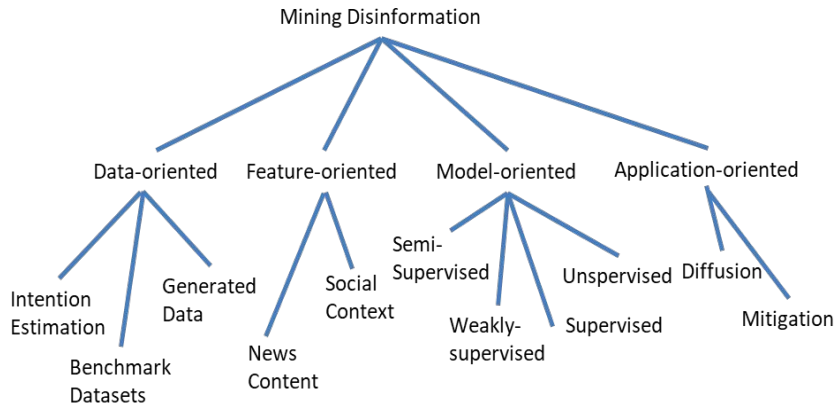
- A **novel** problem of early fake news detection with weak social supervision
- MWSS can **jointly** model **little** labeled data and **multi-source** of weak labels for early fake news detection
- Different sources of weak social supervision contain **complementary** information



Future Work

- Integrating **theories** from different disciplines
 - Learning with weak social supervision
- How can advanced ***AI and machine learning techniques*** help combat disinformation?
 - Multi-modality: text, images, videos, etc
 - Neural-generated: e.g., deepfakes
- Responsible AI for social good
 - Building explanation-enabled AI systems
 - AI with limited data
 - Illuminating AI with interdisciplinary research

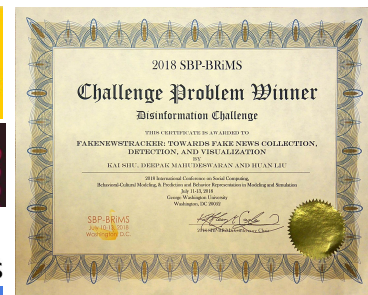
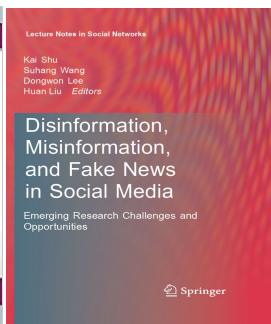
Thank You All



<https://github.com/bigheiniu/awesome-coronavirus19-dataset>



<https://www.fake-news-tutorial.com/>



<http://blogtrackers.fulton.asu.edu:3000/#/>