# Fake News Research: Theories, Detection Strategies, and Open Problems

Reza Zafarani, Xinyi Zhou, Kai Shu, Huan Liu.

**Tutorial** | 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

Syracuse University
Arizona State University

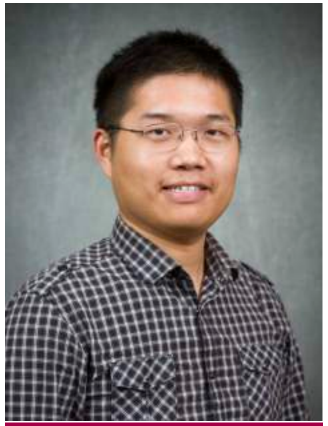
# Meet our Team

Reza Zafarani
Syracuse University

Assistant Professor
Data Lab,
EECS Department

Xinyi Zhou
Syracuse University

Ph.D. Candidate
Data Lab,
EECS Department



Kai Shu
Arizona State University

Ph.D. Candidate
Computer Science
and Engineering

Huan Liu
Arizona State University

Professor
Computer Science
and Engineering

CALL FOR PAPERS

## ACM Journal of Digital Threats: Research and Practice (DTRAP)

*Special Issue on Fake News Research*

**Guest editors**
Reza Zafarani, Syracuse University
Huan Liu, Arizona State University
Vir V. Phoha, Syracuse University
Javad Azimi, Facebook

Fake news, especially on social media, is now viewed as one of the main digital threats to democracy, journalism, and freedom of expression. Our economies are not immune to the spread of fake news either, with fake news being connected to stock market fluctuations and massive trades. The goal of this special issue is to promote exchange of research and studies that (1) aim to understand and characterize fake news and its patterns and how it can be differentiated from other similar concepts such as false/satire news, misinformation, disinformation, among others, which helps deepen our understanding of fake news; and (2) systematically detect fake news by determining its credibility, verifying its facts, assessing its style, or determining its propagation. To facilitate further research in fake news, this special issue especially welcomes research articles, new open access datasets, repositories, and benchmarks for fake news research, broadening research on fake news detection and its development.

**Topics -** The topics of interest of this special issue include but are not limited to:

- Patterns of Fake News
  - Internet measurements on Fake News
  - User behavior analysis with respect to Fake News
  - Patterns of Fake News Distribution/Consumption/Response
  - Tracing and characterizing the propagation of fake news and true news
- Fake News Detection
  - Supervised Fake News Detection
  - Semi-Supervised Fake News Detection
  - Unsupervised Fake News Detection
  - Early Detection of Fake News
  - Deep Nets for Fake News Detection
  - Representation for Fake News
- Mining of News Content
  - Text Mining of News Content
  - Analysis of Images, Videos, and Audio
- Fake Checking
  - Knowledge-based (e.g., Knowledge-graphs) analysis
  - Analyzing News Credibility/Credibility Assessment
  - Analyzing Source Credibility
- Malicious Entity Detection
  - Bot detection
- Fake News Benchmarks
- Fake News Datasets
- Fake News Open Repositories

**Important dates and timeline:**

| | |
|---|---|
| Initial submission: | Dec 1, 2019 |
| First review: | Mar 1, 2020 |
| Revised manuscripts: | May 1, 2020 |
| Second review: | July 1, 2019 |
| Source Files Due: | Aug 1, 2020 |
| Publication: | Sep 2020 |

http://dtrap.acm.org/authors.cfm

**Expected contributions -** We welcome two types of research contributions:

- Research manuscripts reporting novel methodologies and results (up to 25 pages)
- Benchmark, Datasets, Repositories, and Demonstration Systems that enable further research and facilitate research on fake news. These papers should be of interest to the broad fake news research community (10 pages + links to such systems)
- To submit to this special issue, please select "Fake News Research" as paper type

Visit dtrap.acm.org to submit your manuscript

# Introduction

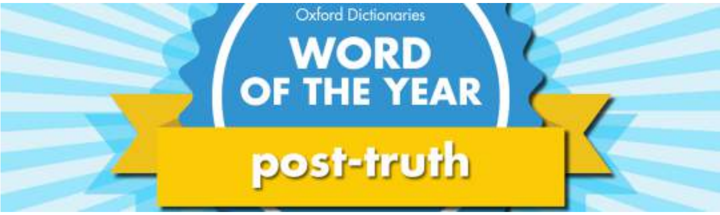
- Research Background
- What is Fake News?
- Related Concepts
- Fundamental Theories

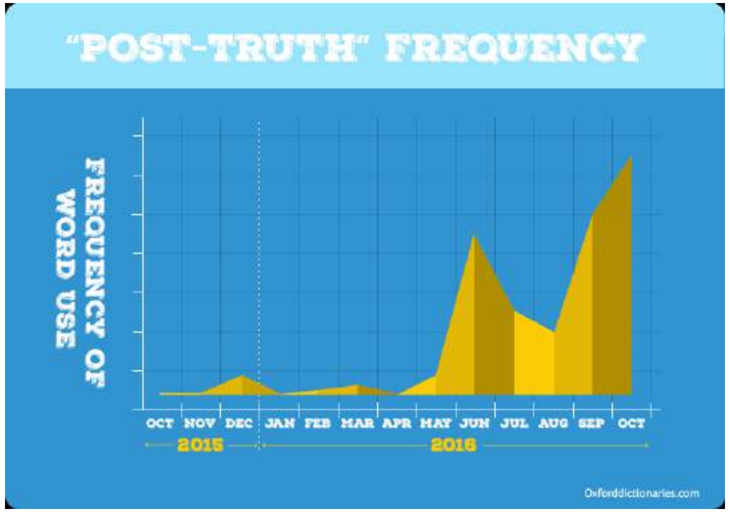
# Research Background

*Why Study Fake News?*

Fake news is now viewed as one of the greatest threats to **democracy**, **justice**, **public trust**, **freedom of expression**, **journalism** and **economy**.

- **Political** Aspects: May have had an impact on
  - "Brexit" referendum
  - 2016 U.S. presidential election
    - # Shares, reactions, and comments on Facebook.[1]
    - 8,711,000 for top 20 frequently-discussed **FAKE** election stories.
    - 7,367,000 for top 20 frequently-discussed **TRUE** election stories.
- Oxford Dictionaries international word of the year 2016:
  - **Post-Truth**: "Relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief."



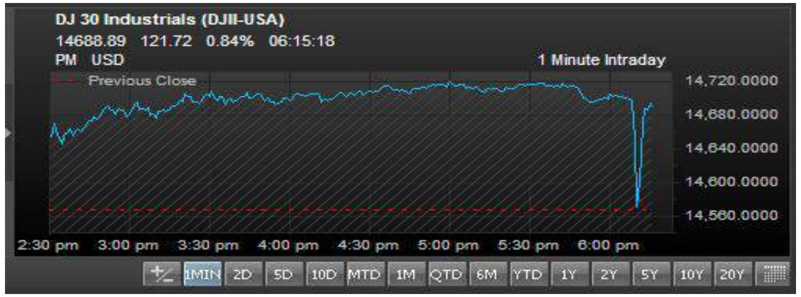

[1]C. Silverman. This analysis shows how viral fake election news stories outperformed real news on Facebook. BuzzFeed News, 2016.

# Research Background

*Why Study Fake News?*

- **Economic** Aspects:
  - “Barack Obama was injured in an explosion” wiped out $130 billion in stock value.[1]
  - Dozens of “well-known” teenagers in Veles, Macedonia[2]
    - Penny-per-click advertising
    - During U.S. 2016 presidential Elections
    - Earning at least $60,000 in six months
    - Far outstripping their parents’ income
    - Average annual wage in town: $4,800



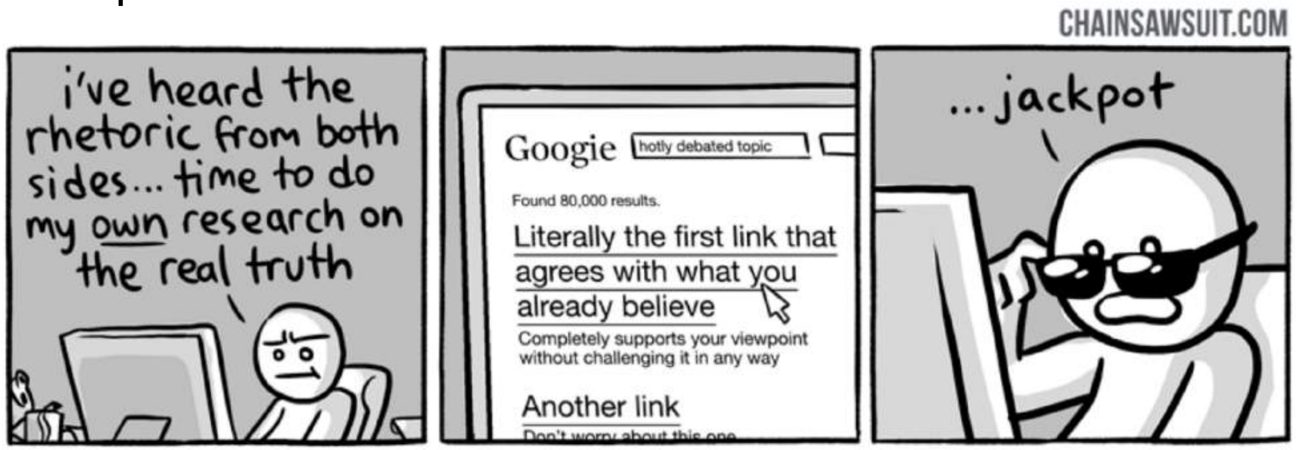



[1]K. Rapoza. Can ‘fake news’ impact the stock market? 2017.
[2] S. Subramanian, Inside the Macedonnian Fake News Complex https://www.wired.com/2017/02/veles-macedonia-fake-news/

# Research Background

*Why Study Fake News?*

- **Social/Psychological** Aspects:
  - Humans have been proven to be irrational/vulnerable when differentiating between truth/false news
    - Typical accuracy in the range of 55-58%
  - For fake news, it is relatively easier to obtain public trust
  - **Validity Effect:** individuals tend to trust fake news after repeated exposures
  - **Confirmation Bias:** individuals tend to believe fake news when it confirms their pre-existing knowledge
  - **Peer Pressure/Bandwagon Effect**






R. Zafarani, X. Zhou, K. Shu, H. Liu

# Research Background

*Why is Fake News attracting more public attention recently?*

- Fake news can now be created and *published faster* and *cheaper*
- The rise of Social Media and its popularity also plays an important role
  - As of Aug. 2017, 67% of Americans *get* their news from social media.[3]
- Social media accelerates *dissemination* of fake news.
  - It breaks the physical distance barrier among individuals.
  - It provides rich platforms to share, forward, vote, and review to encourage users to participate and discuss online news.
- Social media accelerates *evolution* of fake news.
  - **Echo chamber effect**: biased information can be amplified and reinforced within the social media.[4]
  - **Echo Chamber:** a situation in which beliefs are amplified or reinforced by communication and repetition inside a closed system


[3]http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/
[4]K. Jamieson and J. Cappella. Echo Chamber: Rush Limbaugh and the Conservative Media Establishment. Oxford University Press, 2008.


# What Is Fake News?

# Fake News & Related Concepts

*Definition of fake news*

*Fake news is **intentionally** and verifiably **false** news published by a **news** outlet.*

- *Intention*: Bad
- *Authenticity*: False
- *News or not?* News

A more broad definition:

- *Fake news is false news*



Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement

TOPICS: Pope Francis Endorses Donald Trump


BREAKING: Obama And Hillary Now Promising Amnesty To Any Illegal That Votes Democrat



R. Zafarani, X. Zhou, K. Shu, H. Liu

# Fake News & Related Concepts

*Distinguishing fake news from other related concepts*

| | Authenticity | Intention | News? |
|---|---|---|---|
| **Fake news** | False | Bad | Yes |
| **False news** | False | Unknown | Yes |
| **Satire news** | Unknown | Not bad | Yes |
| **Disinformation** | False | Bad | Unknown |
| **Misinformation** | False | Unknown | Unknown |
| **Rumor** | Unknown | Unknown | Unknown |

For example, disinformation is false information [news or non-news] with a bad intention aiming to mislead the public.

*Intention*: Badz



**Open Problems:**

- How similar are writing styles or propagation patterns?
- Can we use the same detection strategies?
- Can we distinguish between them? E.g., fake news from satire news


# Fundamental Theories

# Fundamental Theories

*Why is it necessary to study Fundamental Theories?*

**Fundamental human cognition and behavior theories** developed across various disciplines such as psychology, philosophy, social science, and economics provide invaluable nsights for fake news studies.

1. Provi opportunities for **qualitative and quantitative studies** of big fake news data;
2. Su building **well-justified nd explainable models** for fake news detection and int tion; and
3. elop da ound t esearch

*[Udo] Undeutsch hypothesis:*
A **statement** based on a factual experience differs in **content and quality** from that of fantasy.

Verification:
Is a **fake news** article differs in **content and quality** from the truth?

Utilizing:
How to **detect fake news** based on its **content style and quality**?



R. Zafarani, X. Zhou, K. Shu, H. Liu

# Style-Based Fundamental Theories

*Studying fake news from a style perspective, i..e, how it's written*

| | Term | Phenomenon |
|---|---|---|
| **Style-based** | *Undeutsch hypothesis* | A statement based on a factual experience differs in **content and quality** from that of fantasy |
| | *Reality monitoring* | Actual events are characterized by higher levels of **sensory-perceptual** information. |
| | *Four-factor theory* | Lies are expressed differently in terms of arousal, behavior control, **emotion**, and thinking from truth. |

# Propagation-based Fundamental Theories

*Studying fake news based on how it spreads*

| | Term | Phenomenon |
|---|---|---|
| **Propagation-based** | *Backfire effect* | Given evidence against their beliefs, individuals can reject it even more strongly |
| | *Conservatism bias* | The tendency to revise one's belief insufficiently when presented with new evidence. |
| | *Semmelweis reflex* | Individuals tend to reject new evidence as it contradicts with established norms and beliefs. |

***"Fake news is incorrect but hard to correct"*** [5]

It is difficult to correct users' perceptions after fake news has gained their trust.

**Fake News Early Detection!**

**Providing a solid foundation for epidemic models**

[5] A. Roets, et al. 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. Intelligence, 2017.

# User-based Fundamental Theories

*Studying fake news from a perspective of users:*
*How users engage with fake news and the role users play (or can play) in fake news creation, propagation, or intervention*

| | | Term | Phenomenon |
|---|---|---|---|
| **User-based** (User's Engagement and Role) | **Social influence** | *Attentional bias* | **Exposure frequency** - individuals tend to believe information is correct after repeated exposures. |
| | | *Validity effect* | |
| | | *Echo chamber effect* | |
| | | *Bandwagon effect* | **Peer pressure** - individuals do something primarily because others are doing it and to conform to be liked and accepted by others. |
| | | *Normative influence theory* | |
| | | *Social identity theory* | |
| | | *Availability cascade* | |
| | **Self-influence** | *Confirmation bias* | **Preexisting knowledge** - individuals tend to trust information that confirms their preexisting beliefs or hypotheses, which they perceive to surpass that of others. |
| | | *Illusion of asymmetric insight* | |
| | | *Naïve realism* | |
| | | *Overconfidence effect* | |
| | **Benefit influence** | *Prospect theory* | **Loss and gains preference** - people make decisions based on the value of losses and gains rather than the outcome, and they tend to overestimate the likelihood of gains happening rather than losses. |
| | | *Valence effect, i.e., wishful thinking* | |
| | | *Contrast effect* | |

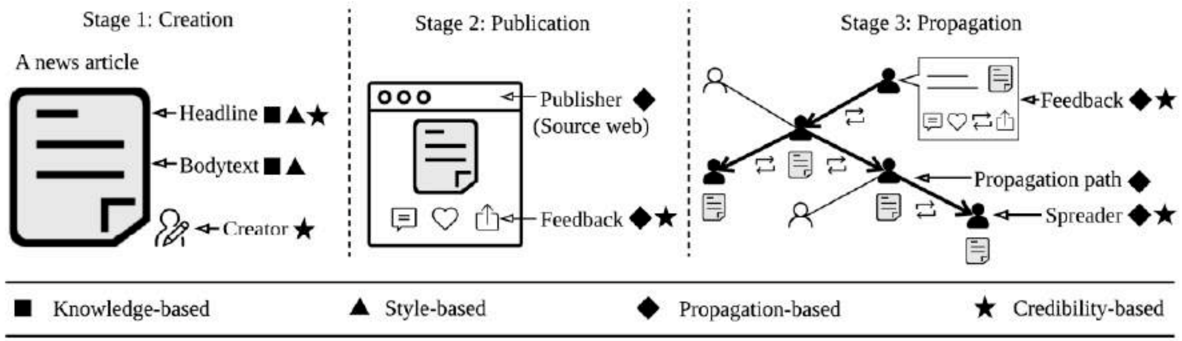
R. Zafarani, X. Zhou, K. Shu, H. Liu

# Fake News Detection

- Knowledge-based Fake News Detection
- Style-based Fake News Detection
- Propagation-based Fake News Detection
- Credibility-based Fake News Detection
- Fake News Datasets & Tools

# Fake News Detection

- Knowledge-based Fake News Detection
- Style-based Fake News Detection
- Propagation-based Fake News Detection
- Credibility-based Fake News Detection
- Fake News Datasets & Tools



# Knowledge-based Fake News Detection

*Overview*

Knowledge-based fake news detection aims to assess **news authenticity** by comparing the **knowledge** extracted from to-be-verified **news content** with known facts (i.e., true knowledge).

It is also known as **fact-checking**.

- *Manual Fact-checking* – providing ground truth.
- *Automatic Fact-checking* – a better choice for scalability.

# Manual Fact-checking

*Classification and comparison*

| | **Expert-based manual fact-checking** | **Crowd-sourced manual fact-checking** |
|---|---|---|
| Fact-checker(s) | One or several domain-expert(s) | A large population of regular individuals |
| Easy to manage? | Yes | No |
| Credibility | High | Comparatively low |
| Scalability | Poor | Comparatively high |
| Current resources (e.g., websites) | Rich | Comparatively poor |

E.g., political bias and conflicting annotations of fact-checkers

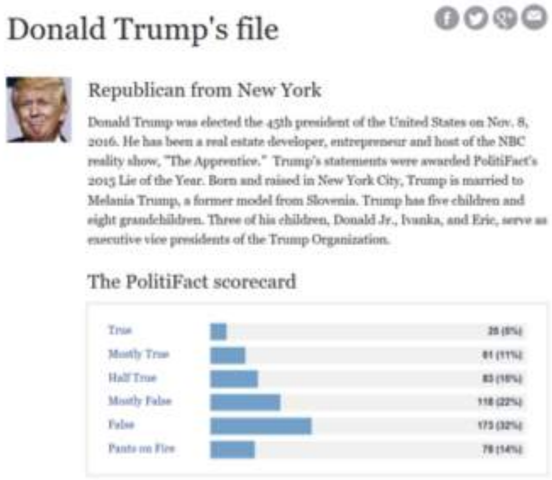
# Expert-based Manual Fact-checking

*Current resources*

Multilabel classification

Binary classification

| | Topics Covered | Content Analyzed | Assessment Labels |
|---|---|---|---|
| **PolitiFact** | American politics | Statements | True; Mostly true; Half true; Mostly false; False; Pants on fire |
| **The Washington Post Fact Checker** | American politics | Statements and claims | One pinocchio; Two pinocchio; Three pinocchio; Four pinocchio; The Geppetto checkmark; An upside-down Pinocchio; Verdict pending |
| **FactCheck** | American politics | TV ads, debates, speeches, interviews and news | True; No evidence; False |
| **Snopes** | Politics and other social and topical issues | News articles and videos | True; Mostly true; Mixture; Mostly false; False; Unproven; Outdated; Miscaptioned; Correct attribution; Misattributed; Scam; Legend |
| **TruthOrFiction** | Politics, religion, nature, aviation, food, medical, etc. | Email rumors | Truth; Fiction; etc. |
| **FullFact** | Economy, health, education, crime, immigration, law | Articles | Ambiguity (no clear labels) |
| **HoaxSlayer** | Ambiguity | Articles and messages | Hoaxes, scams, malware, bogus warning, fake news, misleading, true, humour, spams, etc. |




across domains

Multi-modal


https://reporterslab.org/fact-checking/

# Expert-based Manual Fact-checking

*Current resources*

*Reporters Lab – Duke University*

# Crowd-sourced Manual Fact-checking

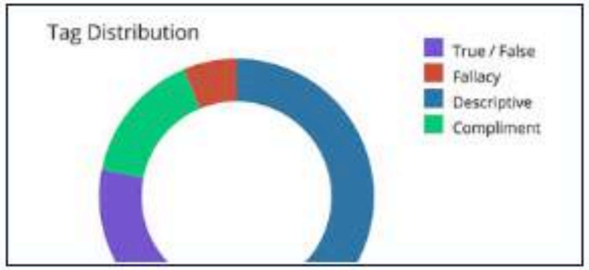
*Current resources*


1 Take an online article that you want to comment on, copy and paste the link into Fiskkit. This allows you to input the article into our system for you to comment on.

OR


Click on an article you find interesting.


2 Rate any sentence inside the article by clicking on a sentence & choosing tags that best describe it. Add comments to support your arguments.

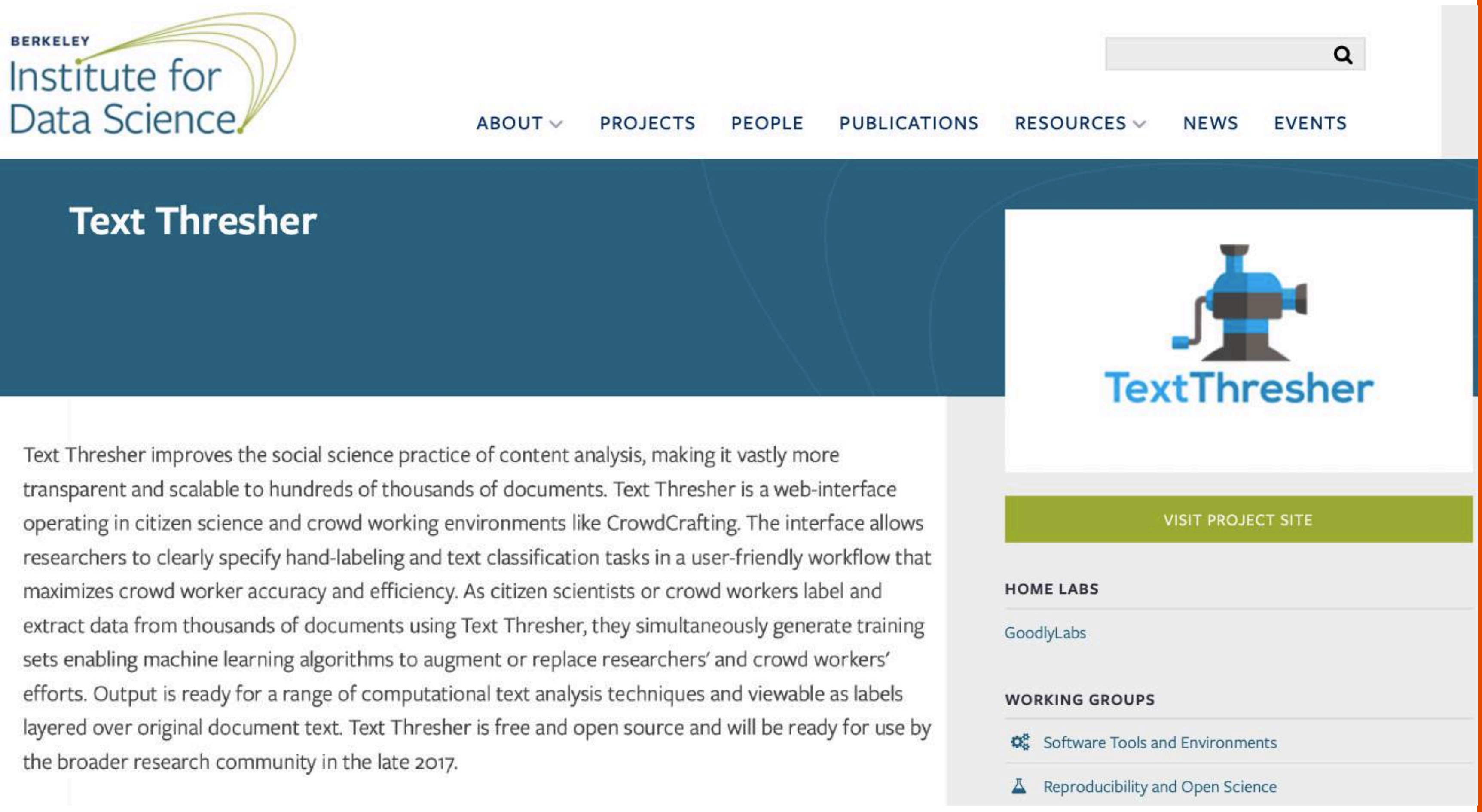



3 See how the article has been rated by other people through our insights page. Share the article so that your friends can come comment too.

http://www.fiskkit.com/

# Crowd-sourced Manual Fact-checking

*Current resources*




A. Zhang, et al. A structured response to misinformation: Defining and annotating credibility indicators in news articles. WWW'18 Companion

# Knowledge-based Fake News Detection

*Overview*

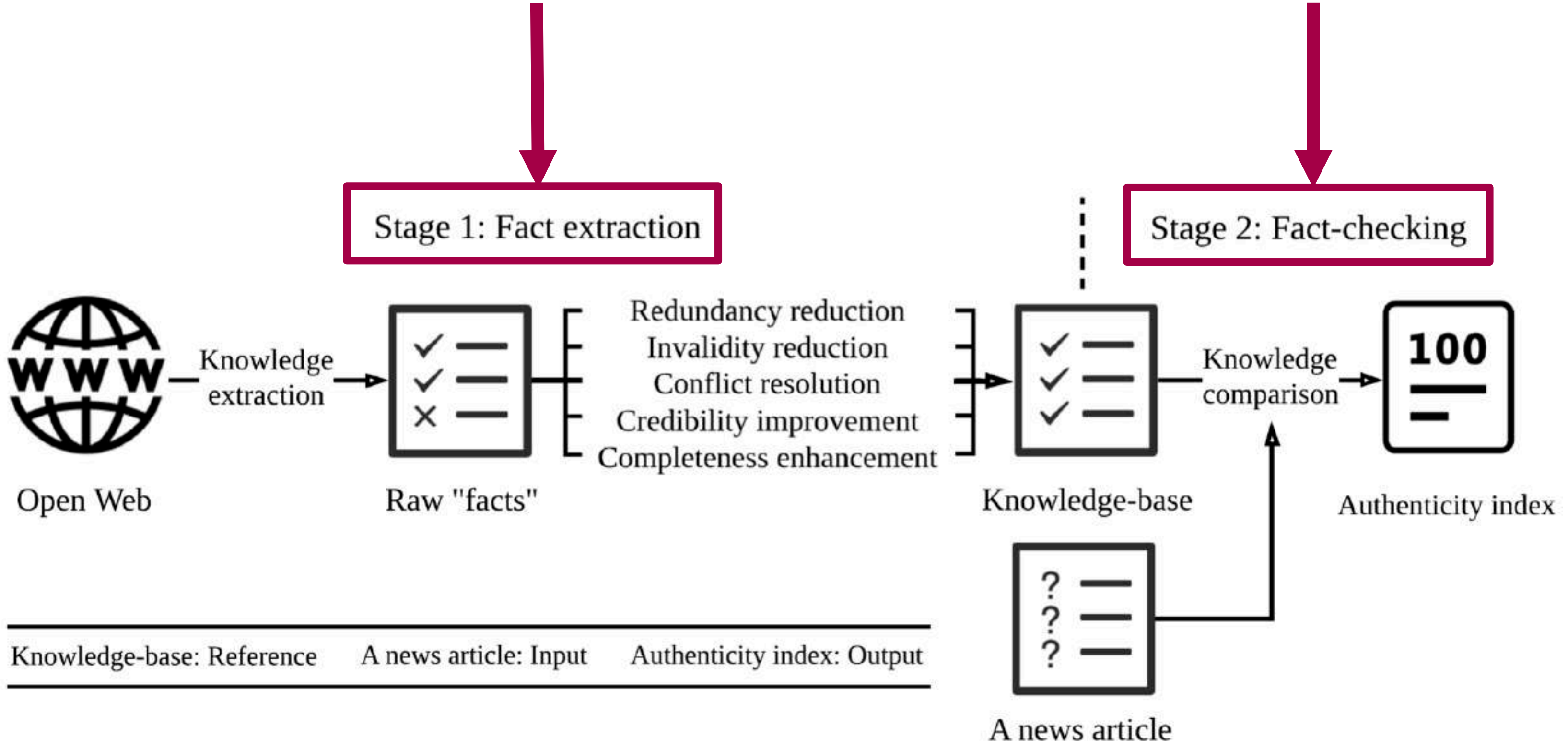
Knowledge-based fake news detection aims to assess **news authenticity** by comparing the **knowledge** extracted from to-be-verified **news content** with known facts (i.e., true knowledge).

It is also known as **fact-checking**.

- *Manual Fact-checking* – providing ground truth.
- *Automatic Fact-checking* – a better choice for scalability.

It aims to assess news authenticity by comparing the knowledge extracted from to-be-verified news content with known facts (i.e., true knowledge).

- How to represent "**knowledge**"?
- How to obtain **the known facts** (i.e., ground truth)?
- How to **compare** the knowledge extracted with known facts?




# Automatic Fact-checking

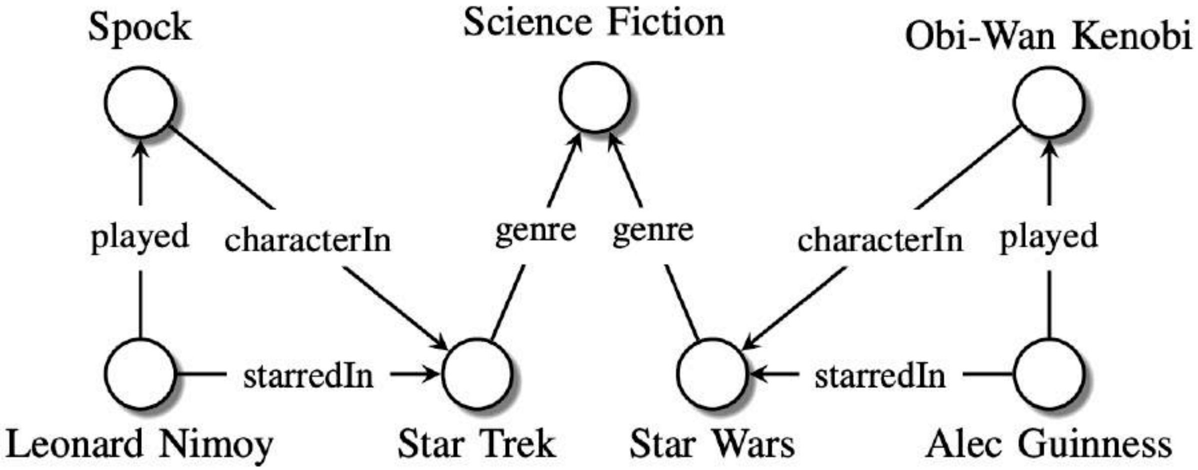
*Overview*

# Knowledge Representation

Knowledge is represented as **a set of (Subject, Predicate, Object) (SPO) triples** extracted from the given information. For example,

*"Leonard Nimoy was an actor who played the character Spock in the science-fiction movie Star Trek"*

| *subject* | *predicate* | *object* |
|---|---|---|
| *(LeonardNimoy,* | *profession,* | *Actor)* |
| *(LeonardNimoy,* | *starredIn,* | *StarTrek)* |
| *(LeonardNimoy,* | *played,* | *Spock)* |
| *(Spock,* | *characterIn,* | *StarTrek)* |
| *(StarTrek,* | *genre,* | *ScienceFiction)* |




The illustration is from: M. Nickel, et al. A Review of Relational Machine Learning for Knowledge Graphs, 2016

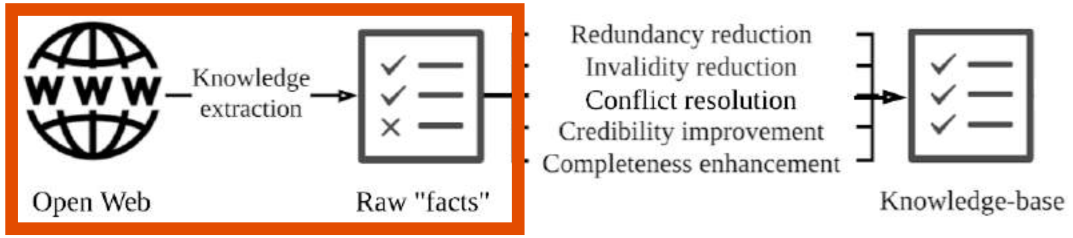
# Stage 1. Fact Extraction

*Constructing knowledge graph to obtain the known facts*

Types of Web content that contain relational information and can be utilized for knowledge extraction by different extractors: **text, tabular data, structured pages** and **human annotations.**[6]

Source(s):

- Single-source knowledge extraction
  - Rely on one comparatively reliable source (e.g., Wiki)
  - Efficient ⬆, Knowledge completeness ⬇
- Open-source knowledge extraction
  - Fuse knowledge from distinct knowledge
  - Efficient ⬇, Knowledge completeness ⬆



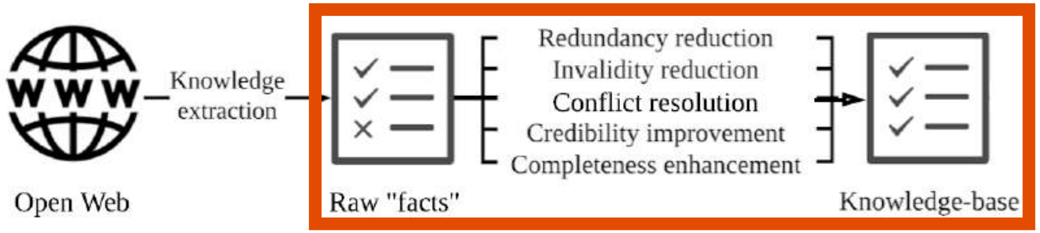

[6]X. Dong, et al.. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. KDD'14



R. Zafarani, X. Zhou, K. Shu, H. Liu

## Stage 1. Fact Extraction

*Constructing knowledge graph to obtain the known facts*




T1: **Entity Resolution (deduplication/record linkage)** to reduce redundancy

- To identify mentions that refer to the same real-world entity, e.g., *(DonaldJohnTrump,profession, President)* & *(DonaldTrump, profession, President)* should be a redundant pair.
- Current techniques are often distance- or dependence-based.
- Often expensive (requires pairwise distance) computation
- Blocking/Indexing can be used to deal with complexity

T2: **Time Recording** to remove outdated knowledge

- E.g., *(Britain, joinIn, EuropeanUnion)* has been outdated.
- Use Compound Value Type (CVT): facts having beginning and end dates
- Timeliness studies are limited

T3: **Knowledge Fusion** to handle conflicts (often in open-source knowledge extraction)

- E.g., *(DonaldTrump, bornIn, NewYorkCity)* & *(DonaldTrump, bornIn, LosAngeles)* are a conflicting pair.
- Fix by having support values for facts (e.g., website credibility), or using ensemble methods
- Often correlated to (T4).

T4: **Credibility Evaluation** to improve the credibility of knowledge

- E.g., The knowledge extracted from The Onion[7].
- Often focus on analyzing the source website(s).

[7]A https://www.theonion.com/

T5: ***Knowledge Inference/Link Prediction*** to infer new facts based on known ones

- Knowledge extracted from online resources, particularly, using a single source, are far from complete.

**Relation machine learning**

**Latent Feature Models,** e.g., **RESCAL**

Assume the existence of knowledge-base triples is conditionally independent given latent features and parameters

**Graph Feature Models**, e.g., **PRA**

Assume the existence of triples is conditionally independent given observed graph features and parameters

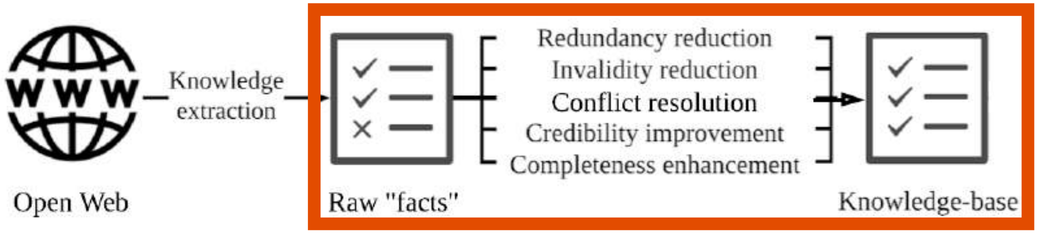
**Markov Random Field (MRF) Models**

Assume the existing triples have local interactions

M. Nickel, et al. A Review of Relational Machine Learning for Knowledge Graphs, Proceedings of the IEEE, 2016

# Stage 1. Fact Extraction

*Constructing knowledge graph to obtain the known facts*



# Stage 1. Fact Extraction

*Existing Knowledge Graphs*

| Name |
|---|
| *Knowledge Vault (KV)* |
| DeepDive [32] |
| NELL [8] |
| PROSPERA [30] |
| YAGO2 [19] |
| Freebase [4] |
| Knowledge Graph (KG) |

**Table 1: Comparison of Freebase and KG rely o facts means with a prob**

[a]Ce Zhang (U Wisconsin), private communication
[b]Bryan Kiesel (CMU), private communication
[c]Core facts, `http://www.mpi-inf.mpg.de/yago-naga/yago/downloads.html`
[d]This is the number of non-redundant base triples, excluding reverse predicates and "lazy" triples derived from flattening CVTs (complex value types).
[e]`http://insidesearch.blogspot.com/2012/12/get-smarter-answers-from-knowledge_4.html`

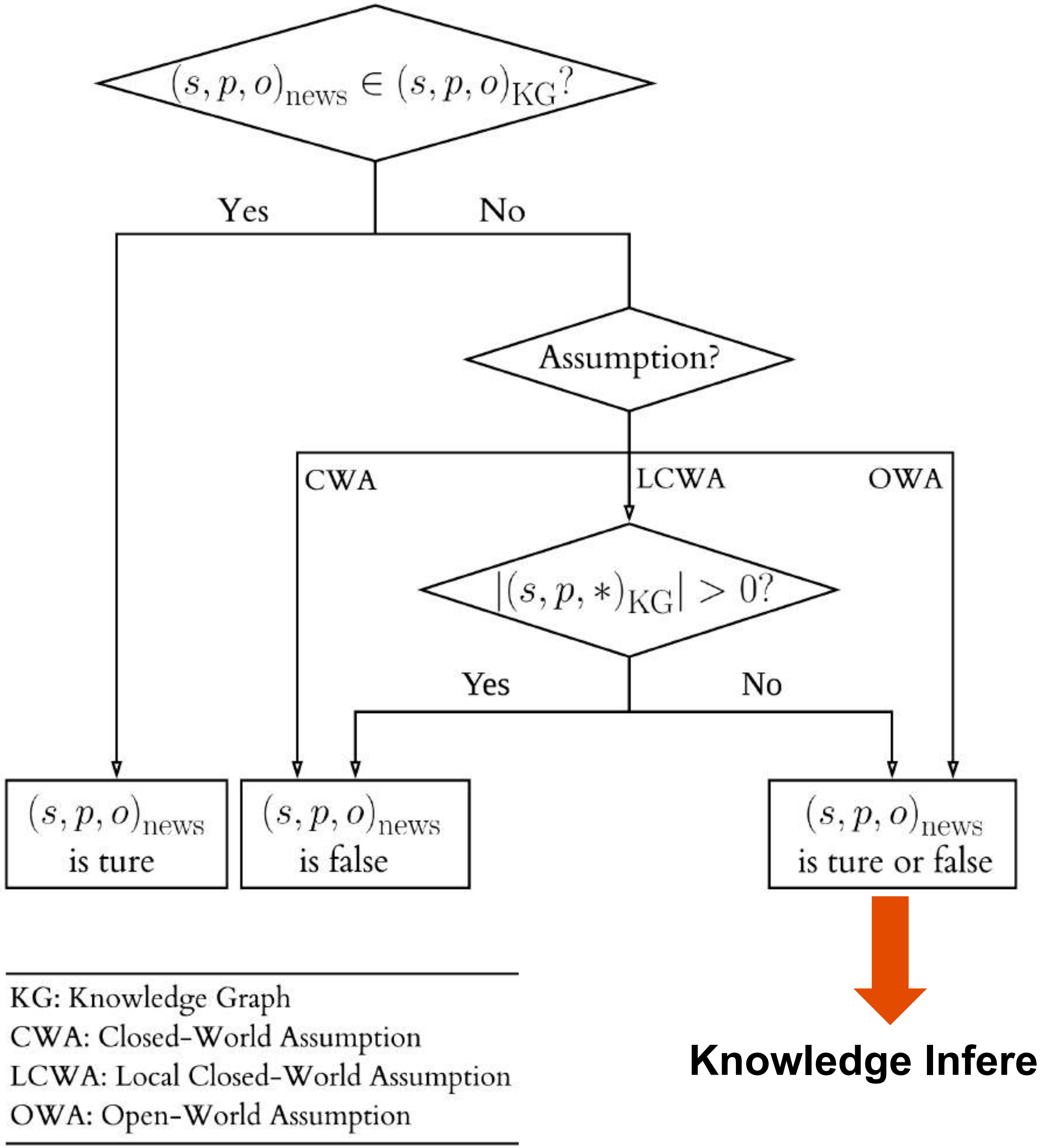
Open issues:

1. **Timeliness & Completeness of Knowledge Graphs**
2. **Domain-specific Knowledge Graphs for Fake News Detection**
*Related tutorial*: X. Ren, et al., Scalable Construction and Querying of Massive Knowledge Bases, WWW tutorial, 2018.

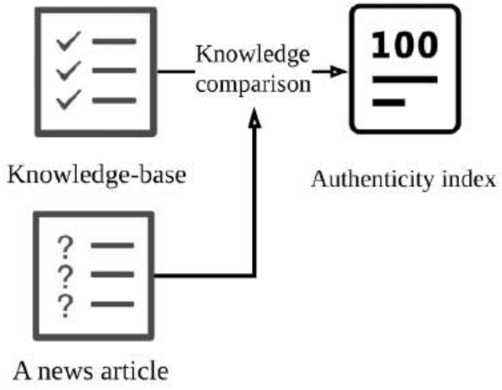
Source: X. Dong, et al.. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. KDD'14

# Stage 2. Fact-checking

*Comparing knowledge between news articles and knowledge graphs*




**Knowledge Inference**



## Stage 2. Fact-checking

*Knowledge Inference for unknown SPO triples: Illustrated studies*

*Shortest path-based method*:

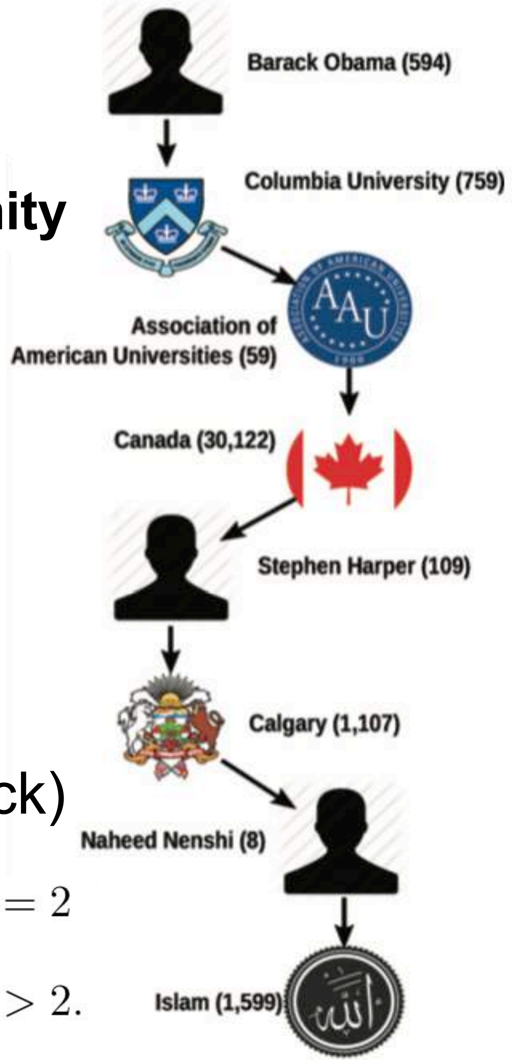
By finding the **shortest path** between concept nodes under properly defined **semantic proximity** metrics on knowledge graphs

$$\tau(e) = \max \mathcal{W}(P_{s,o}).$$

$$\mathcal{W}(P_{s,o}) = \mathcal{W}(v_1 \dots v_n) = \left[1 + \sum_{i=2}^{n-1} \log k(v_i)\right]^{-1}$$

An alternative formulation (widest bottleneck)

$$\mathcal{W}_u(P_{s,o}) = \mathcal{W}_u(v_1 \dots v_n) = \begin{cases} 1 & n = 2 \\ \left[1 + \max_{i=2}^{n-1} \{\log k(v_i)\}\right]^{-1} & n > 2. \end{cases}$$

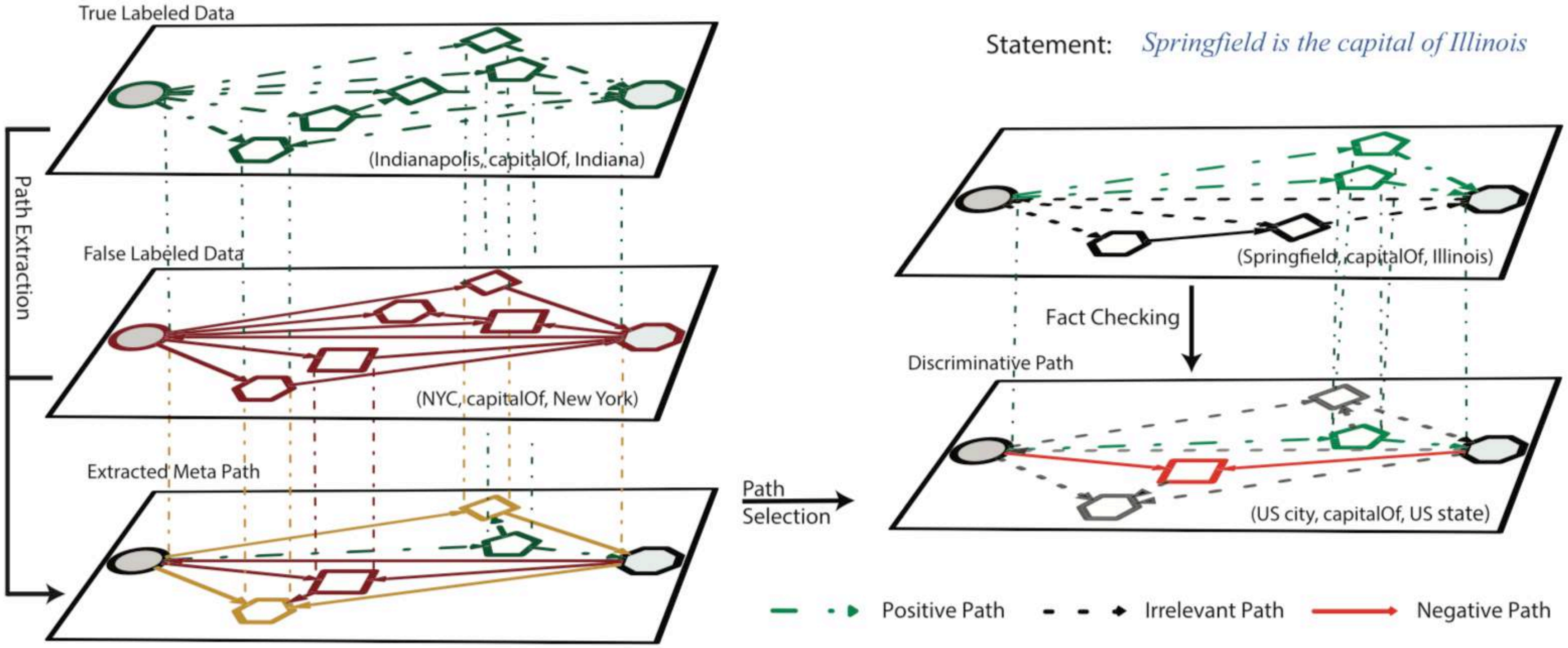



G. Ciampaglia, et al. Computational Fact Checking from Knowledge Networks, 2016

## Stage 2. Fact-checking

*Knowledge Inference for unknown SPO triples: Illustrated studies*

*Discriminative path-based method*:




B. Shi and T. Weninger, Discriminative predicate path mining for fact checking in knowledge graphs, 2015

# Knowledge Inference

*Comparison*

Knowledge inference can be conducted on both Stage I, when constructing knowledge graphs, and Stage II for fact-checking.

| **Stage** / **Operation** | **Knowledge Graph Construction** | **Fact-checking** |
|---|---|---|
| **Entity/Node** | *Few* operations on entities | Generally requires *additional* operations on entities, e.g., entity matching |
| **Relationship/Edge** | Inference targets relationships between *each pair of* given entities | Inference only targets relationships among *partial* entities |

# Fake News Detection

Xinyi Zhou, Ph.D. Candidate
Data Lab, EECS Department, Syracuse University
zhouxinyi@data.syr.edu www.xzhou.net

# Fake News: A Survey of Research, Detection Methods, and Opportunities

Xinyi Zhou and Reza Zafarani
Data Lab, EECS Department, Syracuse University

# Style-based Fake News Detection

Xinyi Zhou, Ph.D. Candidate
Data Lab, EECS Department, Syracuse University
zhouxinyi@data.syr.edu www.xzhou.net

# Style-based
# Fake News Detection

Xinyi Zhou, Ph.D. Candidate
Data Lab, EECS Department, Syracuse University
zhouxinyi@data.syr.edu www.xzhou.net

**IT'S OVER: Hillary's ISIS Email Just Leaked & It's Worse Than Anyone Could Have Imagined...**


— Hillary Clinton, Friend of the Syria people? Like the USA is friends of the people of Iraq, Afghanistan, Pakistan, Libya, Somalia, Yemen...?

**Today Wikileaks released what is, by far, the most devastating leak of the entire campaign. This makes Trump's dirty talk video looks like an episode of Barney and Friends.**

Even though when Trump called Hillary the 'founder' of ISIS he was telling the truth and 100% accurate, the media has never stopped ripping him apart over it.

Today the media is forced to eat their hats because the newest batch of leaked emails show Hillary, in her own words, admitting to doing just that, funding and running ISIS.

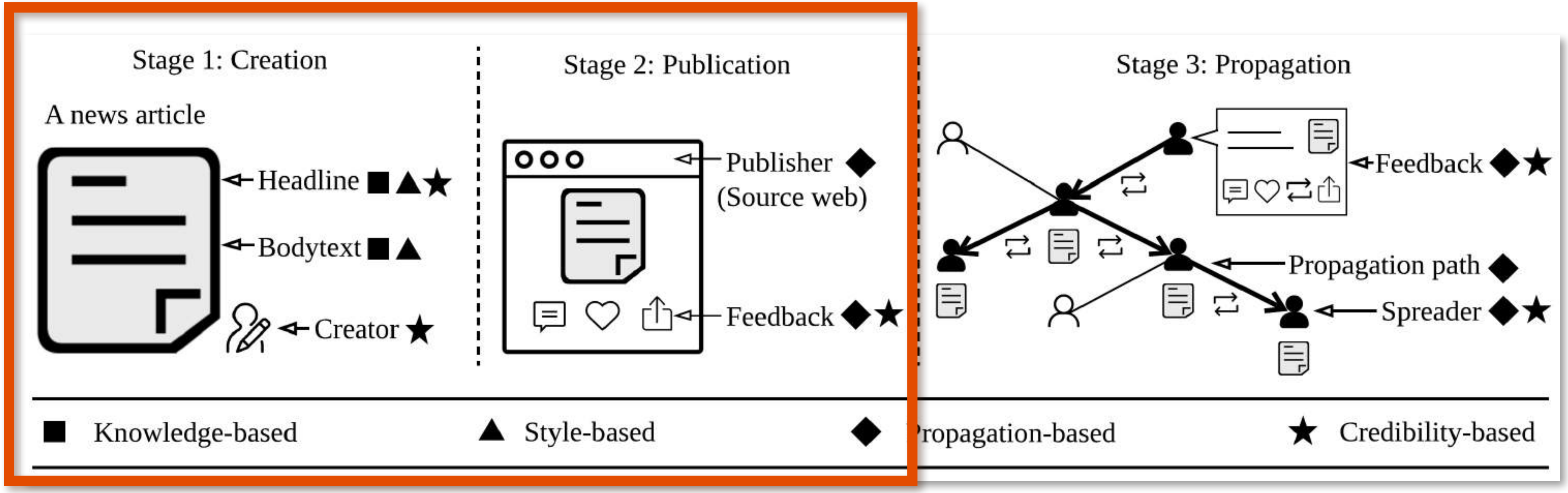
John Podesta, Hillary's campaign chair, who was also a counselor to President Obama at the time, was the recipient of the 2014 email which was released today.

Assange promised his latest batch of leaks would lead to the indictment of Hillary, and it looks like he was not kidding. The email proves Hillary knew and was complicit in the funding and arming of ISIS by our 'allies' Saudi Arabia and Qatar!

# THE GOOD

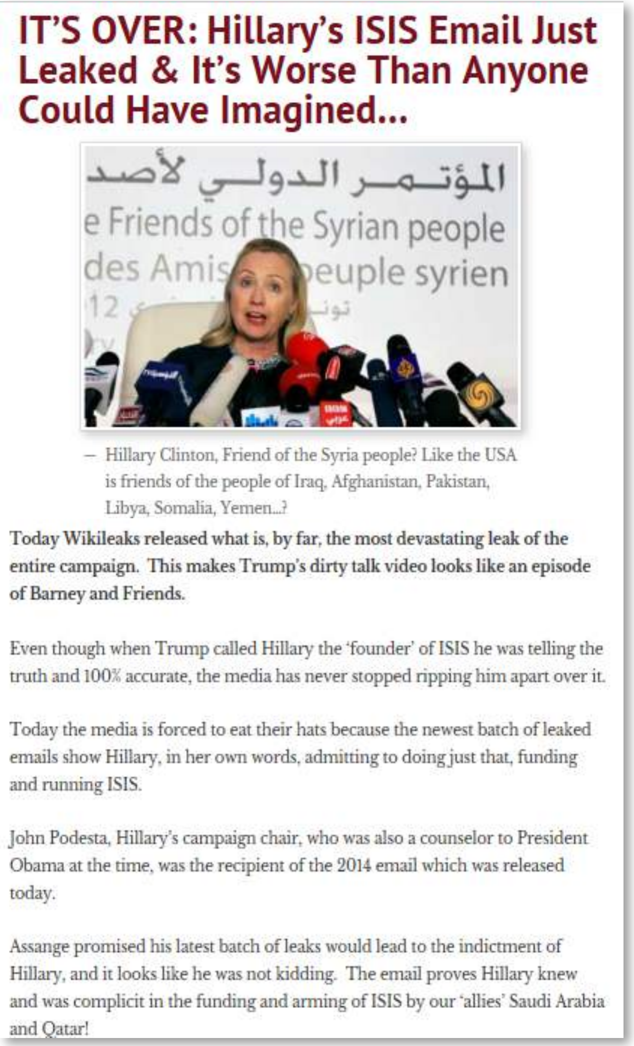
It can detect fake news before propagation…
It can detect “real” fake news…

| News Content |
|---|
| ~~Knowledge-based~~ |
| Style-based |




| ~~Social Context~~ |
|---|
| ~~Propagation-based~~ |
| ~~Credibility-based~~ |

# THE WAY TO DETECT



IT'S OVER: Hillary's ISIS Email Just Leaked & It's Worse Than Anyone Could Have Imagined...



– Hillary Clinton, Friend of the Syria people? Like the USA is friends of the people of Iraq, Afghanistan, Pakistan, Libya, Somalia, Yemen...?

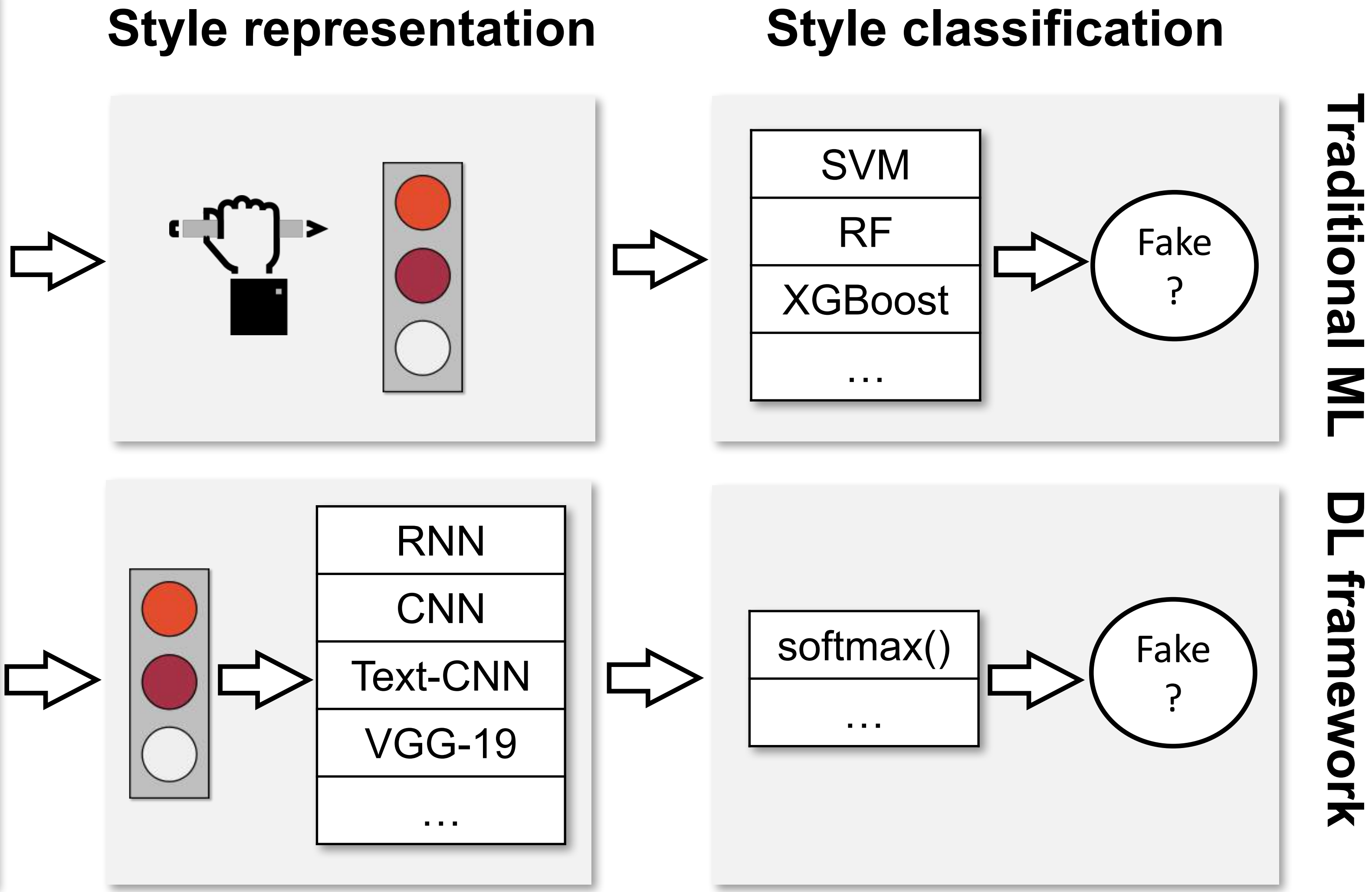
Today Wikileaks released what is, by far, the most devastating leak of the entire campaign. This makes Trump's dirty talk video looks like an episode of Barney and Friends.

Even though when Trump called Hillary the 'founder' of ISIS he was telling the truth and 100% accurate, the media has never stopped ripping him apart over it.

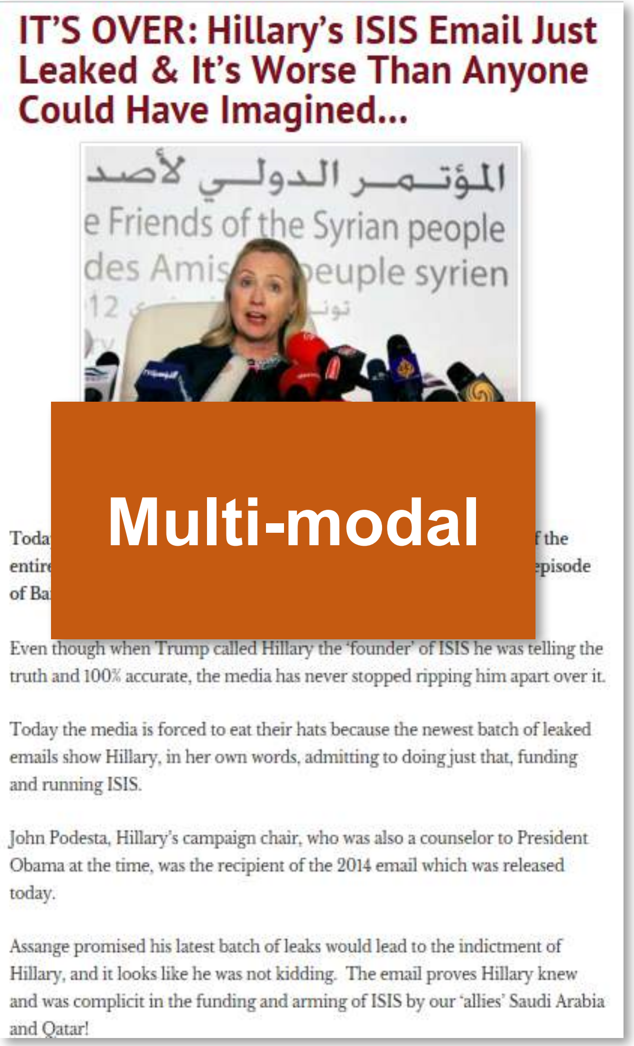
Today the media is forced to eat their hats because the newest batch of leaked emails show Hillary, in her own words, admitting to doing just that, funding and running ISIS.

John Podesta, Hillary's campaign chair, who was also a counselor to President Obama at the time, was the recipient of the 2014 email which was released today.

Assange promised his latest batch of leaks would lead to the indictment of Hillary, and it looks like he was not kidding. The email proves Hillary knew and was complicit in the funding and arming of ISIS by our 'allies' Saudi Arabia and Qatar!



# THE WAY TO DETECT



IT'S OVER: Hillary's ISIS Email Just Leaked & It's Worse Than Anyone Could Have Imagined...

Toda[...] f the entire[...] episode of Ba[...]

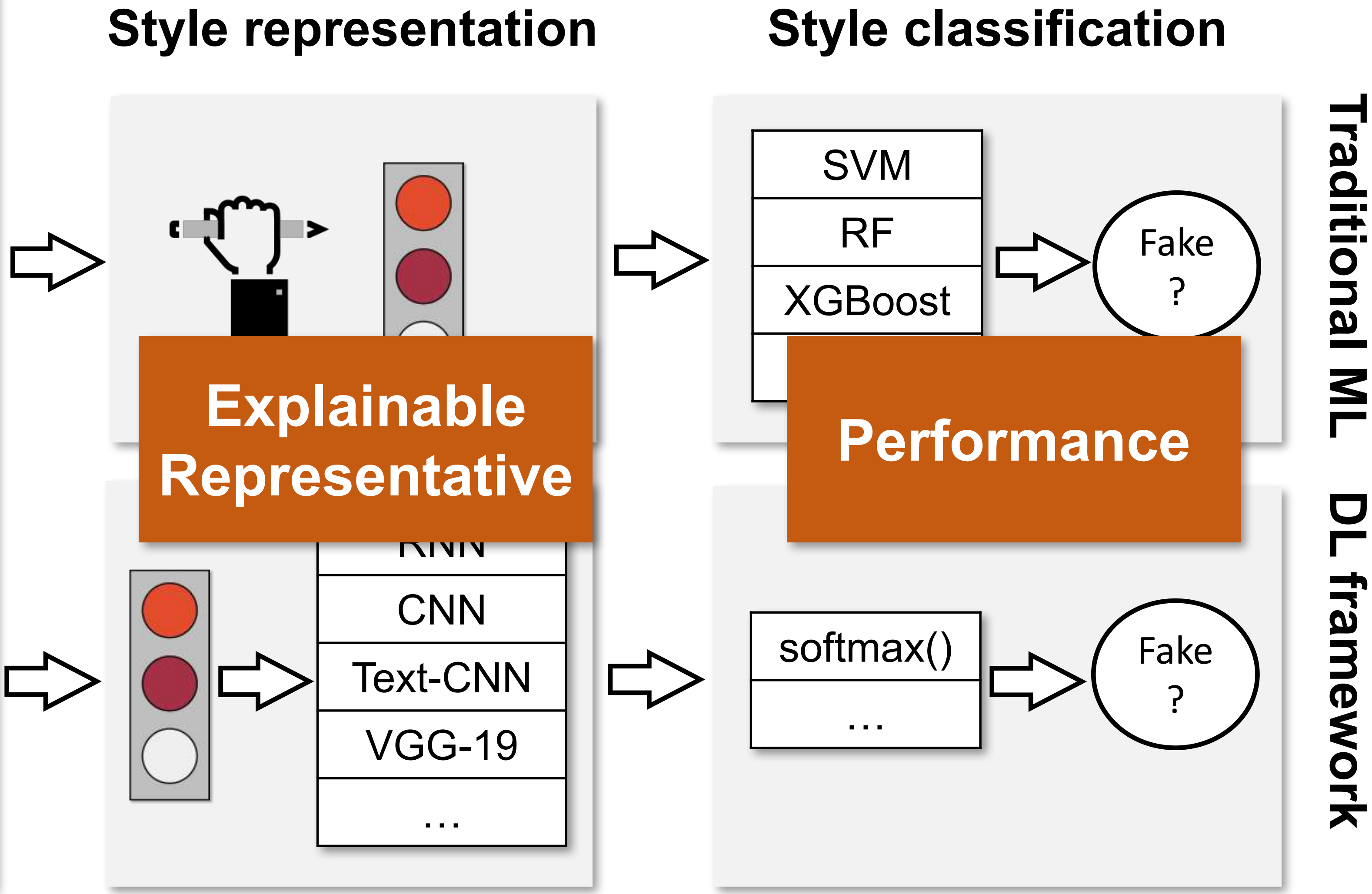
Even though when Trump called Hillary the 'founder' of ISIS he was telling the truth and 100% accurate, the media has never stopped ripping him apart over it.

Today the media is forced to eat their hats because the newest batch of leaked emails show Hillary, in her own words, admitting to doing just that, funding and running ISIS.

John Podesta, Hillary's campaign chair, who was also a counselor to President Obama at the time, was the recipient of the 2014 email which was released today.

Assange promised his latest batch of leaks would lead to the indictment of Hillary, and it looks like he was not kidding. The email proves Hillary knew and was complicit in the funding and arming of ISIS by our 'allies' Saudi Arabia and Qatar!

**Multi-modal**




**Explainable Representative**

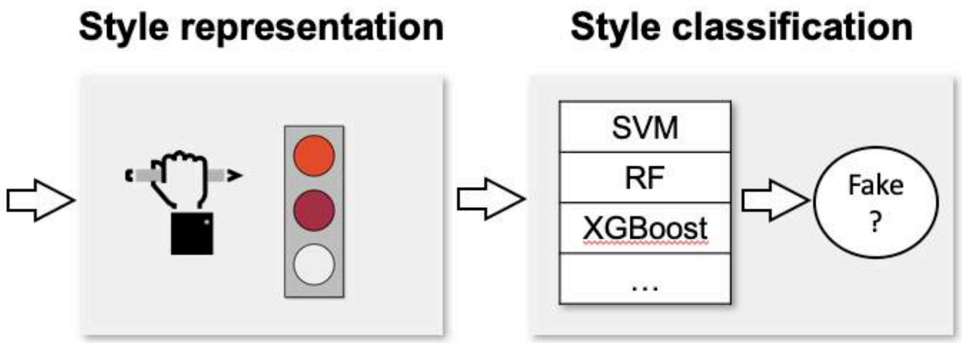
**Performance**

# Fake News Early Detection: A Theory-driven Model

Xinyi Zhou, Atishay Jain, Vir V. Phoha, Reza Zafarani
EECS Department, Syracuse University

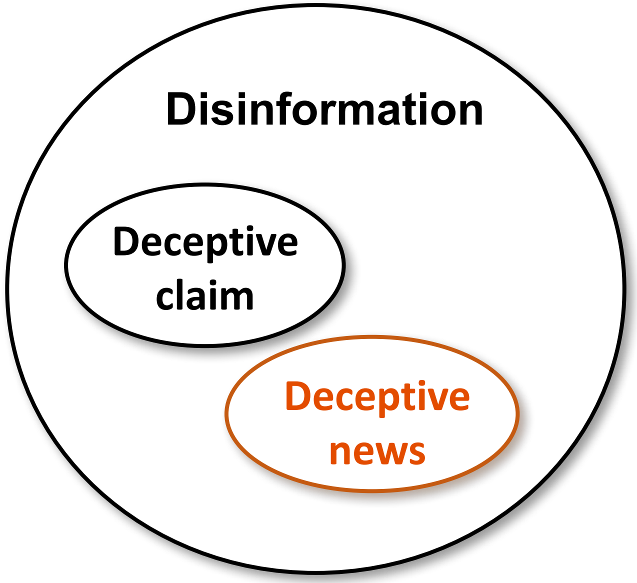
# Fake News Early Detection: A Theory-driven Model

Xinyi Zhou, Atishay Jain, Vir V. Phoha, Reza Zafarani


- **Interpretability**
- **Empirical relations**

| | |
|---|---|
| *Undeutsch hypothesis* | **Deceptive statements** differ in content **style and quality** from the truth. |
| *Reality monitoring* | **Deceptive claims** are characterized by higher levels of **sensory-perceptual** information. |
| *Four-factor theory* | **Lies** are expressed differently in **emotion** and **cognitive process** from the truth. |
| *Info. Manipu-lation theory* | Extreme information **quantity** often exists in **deception**. |



# Fake News Early Detection: A Theory-driven Model

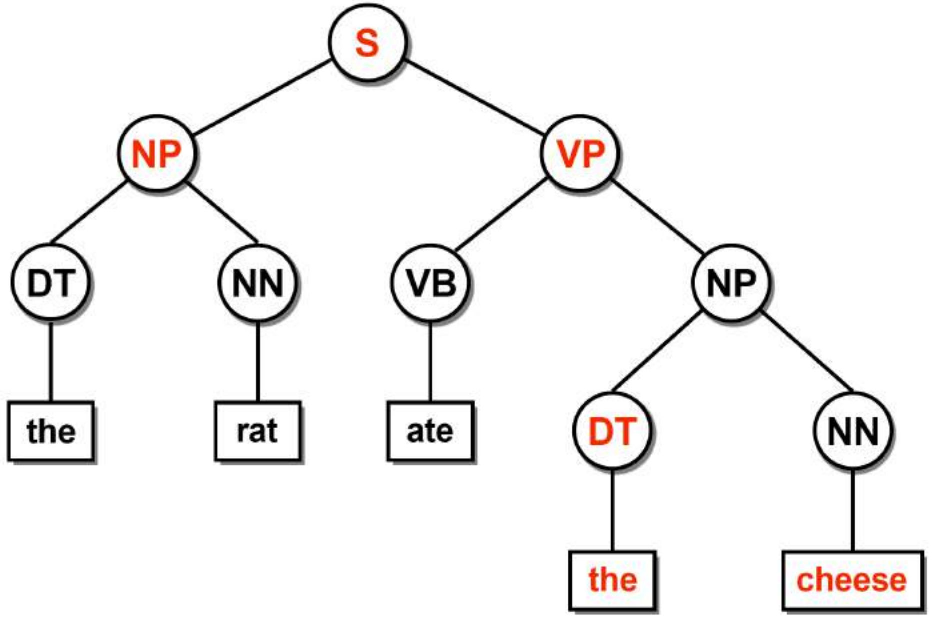
Xinyi Zhou, Atishay Jain, Vir V. Phoha, Reza Zafarani

| | | $N_1$ | $N_2$ | $N_3$ |
|---|---|---|---|---|
| Lexicon | 'rat' | 1 | x | x |
| | 'cheese' | 1 | x | x |
| POS | noun | 2 | x | x |
| | verb | 1 | x | x |
| CFG | **S → NP VP** | **1** | x | x |
| | **DT → 'the'** | **2** | x | x |
| RR | **Evidence** | **1** | x | x |
| | **Condition** | **2** | x | x |

## I. Writing Style

| Level | Feature(s) |
|---|---|
| Lexicon | BOWs |
| Syntax | POS Tags |
| | CFGs |
| Discourse | RRs |

**Frequency**: Absolute? Standardized? Relative by using TF-IDF?




**Fake News Early Detection: A Theory-driven Model**

Xinyi Zhou, Atishay Jain, Vir V. Phoha, Reza Zafarani

# II. Content Quality

| | Feature(s) | Example | Tool & Ref. |
|---|---|---|---|
| **Informality** | #/% Swear Words | "damn" | Linguistic Inquiry and Word Count (LIWC) |
| | #/% Netspeak | "btw" | |
| | #/% Assent | "OK" | |
| | #/% Nonfluencies | "umm" | |
| | #/% Fillers | "you know" | |
| | Overall #/% Informal Words | / | |
| **Subjectivity** | #/% Biased Lexicons | "attack" | [1] |
| | #/% Report Verbs | "announce" | |
| | #/% Factive Verbs | "observe" | [2] |
| **Diversity** | #/% Unique Words | / | / |
| | #/% Unique Content Words | "car" | LIWC |
| | #/% Unique Nouns | / | POS Taggers |
| | #/% Unique Verbs | / | |
| | #/% Unique Adjectives | / | |
| | #/% Unique Adverbs | / | |

[1] Marta Recasens, et al. Linguistic Models for Analyzing and Detecting Biased Language. ACL, 2013.

[2] J Hooper. On Assertive Predicates in Syntax and Semantics, New York, 1975.

# Fake News Early Detection: A Theory-driven Model

Xinyi Zhou, Atishay Jain, Vir V. Phoha, Reza Zafarani

## III. Sentiment

| | |
|---|---|
| #/% Positive Words | LIWC |
| #/% Negative Words | |
| #/% Anxiety Words | |
| #/% Anger Words | |
| #/% Sadness Words | |
| Overall #/% Emotional Words | |
| Avg. Sentiment Score of Words | NLTK |

## IV. Quantity

| |
|---|
| # Characters |
| # Words |
| # Sentences |
| # Paragraphs |
| Avg. # Characters Per Word |
| Avg. # Words Per Sentence |
| Avg. # Sentences Per Paragraph |

## V. Cognitive Process

| | | |
|---|---|---|
| #/% Insight | "think" | LIWC |
| #/% Causation | "because" | |
| #/% Discrepancy | "should" | |
| #/% Tentative | "perhaps" | |
| #/% Certainty | "always" | |
| #/% Differentiation | "but" | |
| Overall #/% Cognitive Processes | | |

## VI. Perceptual Process

| | |
|---|---|
| #/% See | LIWC |
| #/% Hear | |
| #/% Feel | |
| Overall #/% Perceptual Processes | |

**Fake News Early Detection: A Theory-driven Model**

Xinyi Zhou, Atishay Jain, Vir V. Phoha, Reza Zafarani

# Within/Across-level Performance

| | Language Level | Feature Group | PolitiFact | | | | BuzzFeed | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | XGBoost | | RF | | XGBoost | | RF | |
| | | | Acc. | F1 | Acc. | F1 | Acc. | F1 | Acc. | F1 |
| Within Levels | **Lexicon** | BOW | .856 | .858 | .837 | .836 | .823 | .823 | .815 | .815 |
| | **Shallow Syntax** | POS | .755 | .755 | .776 | .776 | .745 | .745 | .732 | .732 |
| | **Deep Syntax** | CFG | .877 | .877 | .836 | .836 | .778 | .778 | .845 | .845 |
| | **Semantic** | DIA+CBA | .745 | .748 | .737 | .737 | .722 | .750 | .789 | .789 |
| | **Discourse** | RR | .621 | .621 | .633 | .633 | .658 | .658 | .665 | .665 |
| Across Two Levels | **Lexicon+Syntax** | BOW+POS+CFG | .858 | .860 | .822 | .822 | .845 | .845 | .871 | .871 |
| | **Lexicon+Semantic** | BOW+DIA+CBA | .847 | .820 | .839 | .839 | .844 | .847 | .844 | .844 |
| | **Lexicon+Discourse** | BOW+RR | .877 | .877 | .880 | .880 | .872 | .873 | .841 | .841 |
| | **Syntax+Semantic** | POS+CFG+DIA+CBA | .879 | .880 | .827 | .827 | .817 | .823 | .844 | .844 |
| | **Syntax+Discourse** | POS+CFG+RR | .858 | .858 | .813 | .813 | .817 | .823 | .844 | .844 |
| | **Semantic+Discourse** | DIA+CBA+RR | .855 | .857 | .864 | .864 | .844 | .841 | .847 | .847 |
| Across Three Levels | **All-Lexicon** | All-BOW | .870 | .870 | .871 | .871 | .851 | .844 | .856 | .856 |
| | **All-Syntax** | All-POS-CFG | .834 | .834 | .822 | .822 | .844 | .844 | .822 | .822 |
| | **All-Semantic** | All-DIA-CBA | .868 | .868 | .852 | .852 | .848 | .847 | .866 | .866 |
| | **All-Discourse** | All-RR | **.892** | **.892** | **.887** | **.887** | **.879** | **.879** | .868 | .868 |
| | | Overall | .865 | .865 | .845 | .845 | .855 | .856 | .854 | .854 |

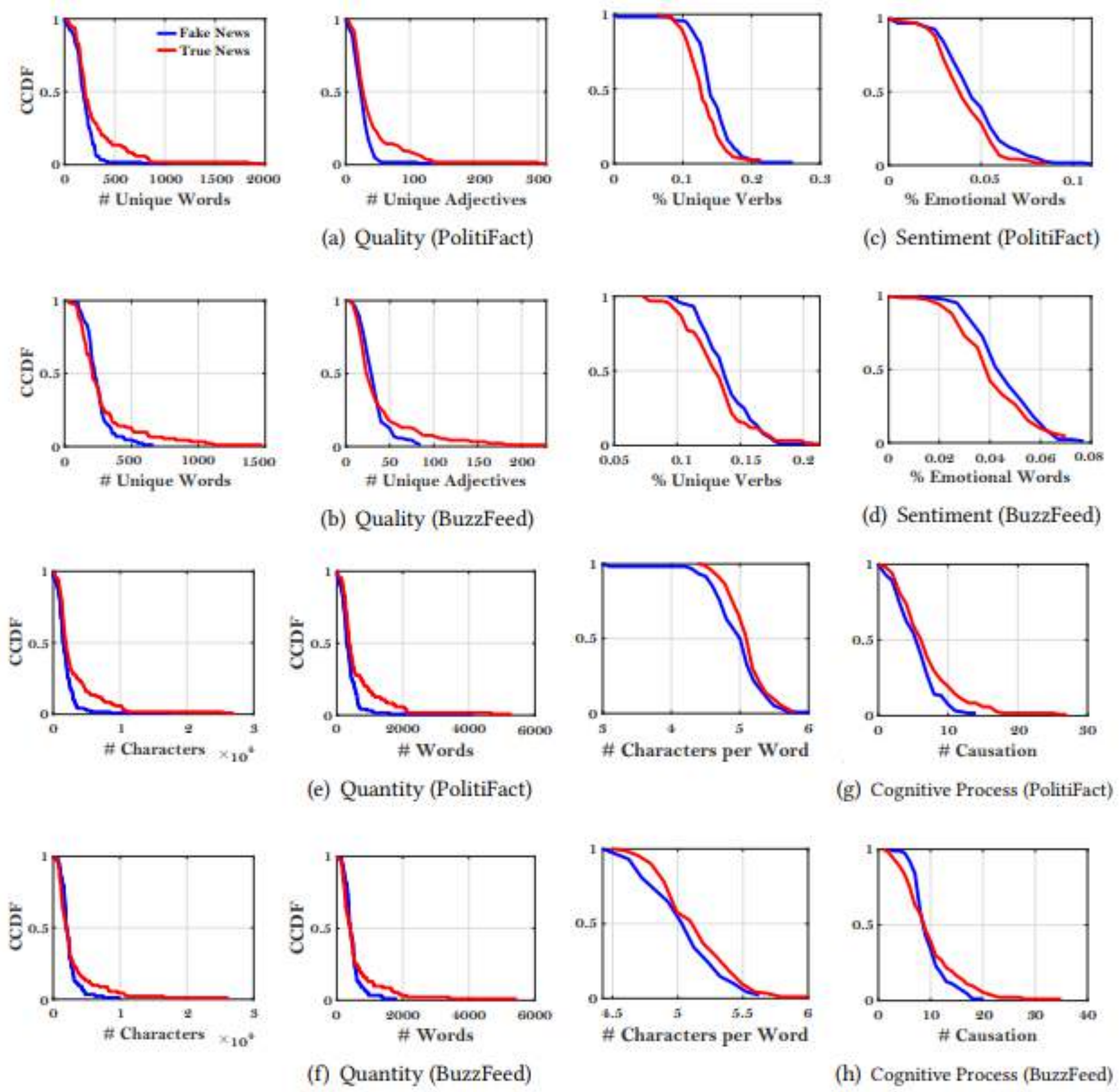
Within-level
1. **Lexicon / Deep Syntax**
   (80%~90%)
2. **Semantic / Shallow Syntax**
   (70%~80%)
3. **Discourse**
   (60%~70%)

**Across-level > Within-level**
(exclude RRs)

# Fake News & Deception

| Supportive Theory | Deception | Fake News |
|---|---|---|
| *Undeutsch hypothesis* | Differs in content **style** and **quality** from truth | 😀 Consistent |
| *Reality monitoring* | Has a higher levels of **sensory-perceptual information** than truth | 😒 Similar levels to the truth |
| *Four-factor theory* | Differs in **cognitive process** from the truth | 😀 Carries poorer cognitive information than truth |
| *Info. Manip-ulation theory* | Often refers to extreme information **quantity** | 😲 More words in headlines while less in body-text. |

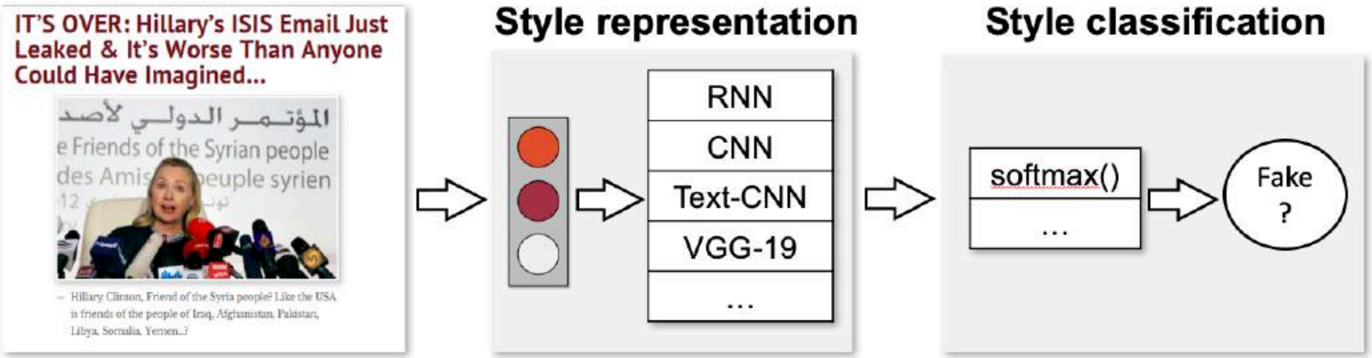
*p*-value<0.1



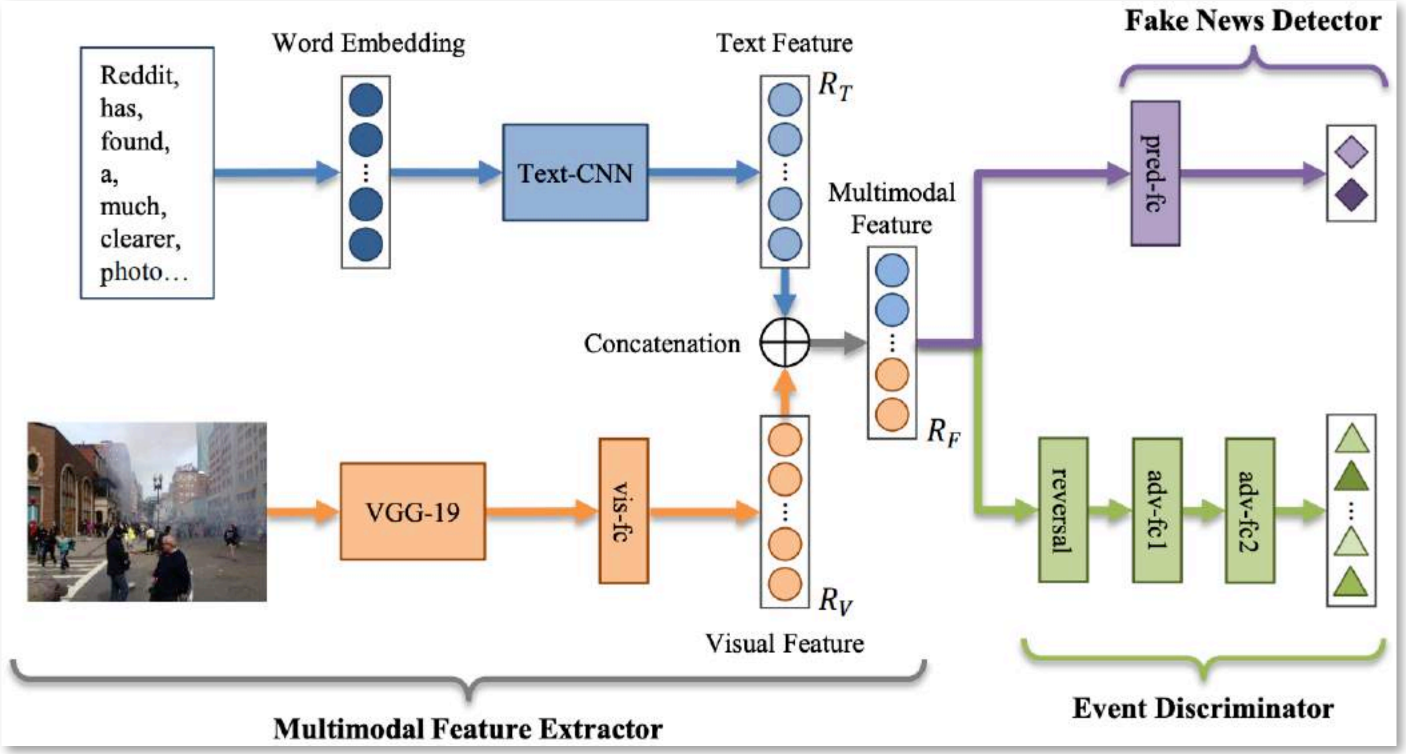

(a) Quality (PolitiFact)
(b) Quality (BuzzFeed)
(c) Sentiment (PolitiFact)
(d) Sentiment (BuzzFeed)
(e) Quantity (PolitiFact)
(f) Quantity (BuzzFeed)
(g) Cognitive Process (PolitiFact)
(h) Cognitive Process (BuzzFeed)

# EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection

Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, Jing Gao




- **Multi-modal**
- **Event-invariant**




$$(\hat{\theta}_f, \hat{\theta}_d) = arg \min_{\theta_f, \theta_d} L_{final}(\theta_f, \theta_d, \hat{\theta}_e),$$

$$\hat{\theta}_e = arg \max_{\theta_e} L_{final}(\hat{\theta}_f, \theta_e).$$

# THE CHALLENGES

I. Algorithm transparency
  - Writing style can be manipulated…

II. Golden datasets with reliable labels
  - Multi-labels, domains, languages, modals, …

III. Different types of fake news
  - Mining relationships between text and images

IV. Model explain-ability
  - Introducing fundamental theories to guide learning process in NNs

# THE WEBSITE https://www.fake-news-tutorial.com/

Fake News: A Survey of Research, Detection Methods, and Opportunities. Xinyi Zhou, Reza Zafarani. arXiv, 2018.

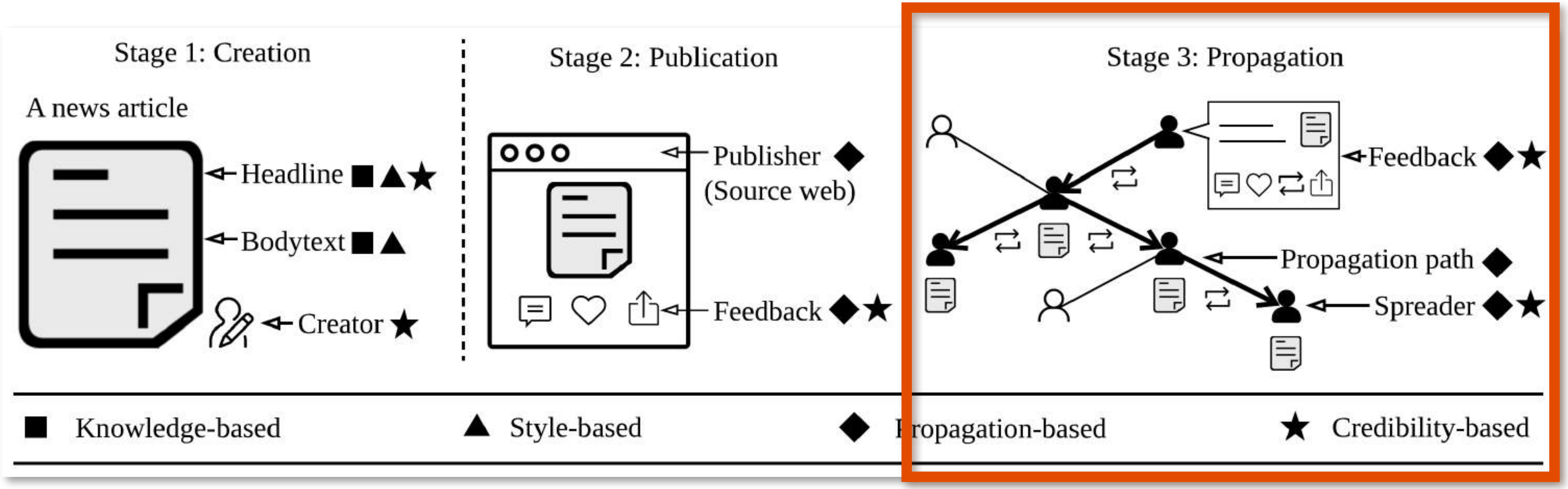
# Fake News Detection

www.fake-news-tutorial.com

Xinyi Zhou, Ph.D. Candidate
Data Lab, EECS Department, Syracuse University
zhouxinyi@data.syr.edu www.xzhou.net

# THE **GOOD** 

Massive auxiliary information can be utilized for comprehensive evaluation.



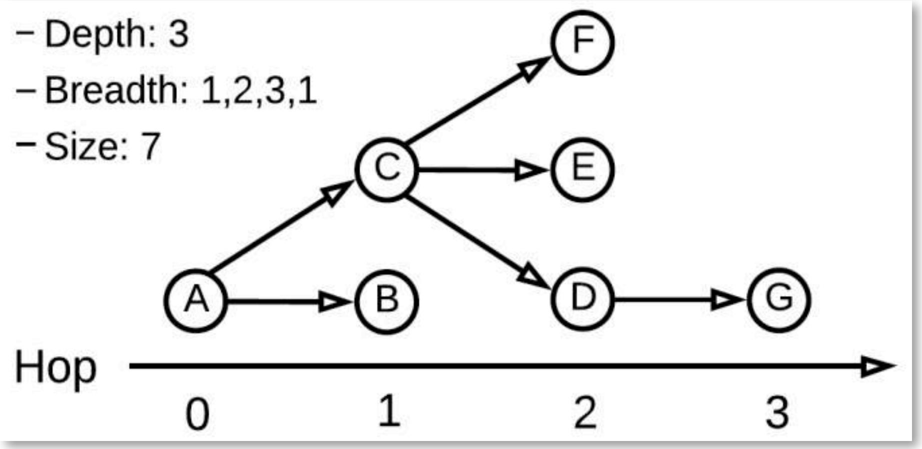

| News Content |
|---|
| ~~Knowledge-based~~ |
| ~~Style-based~~ |

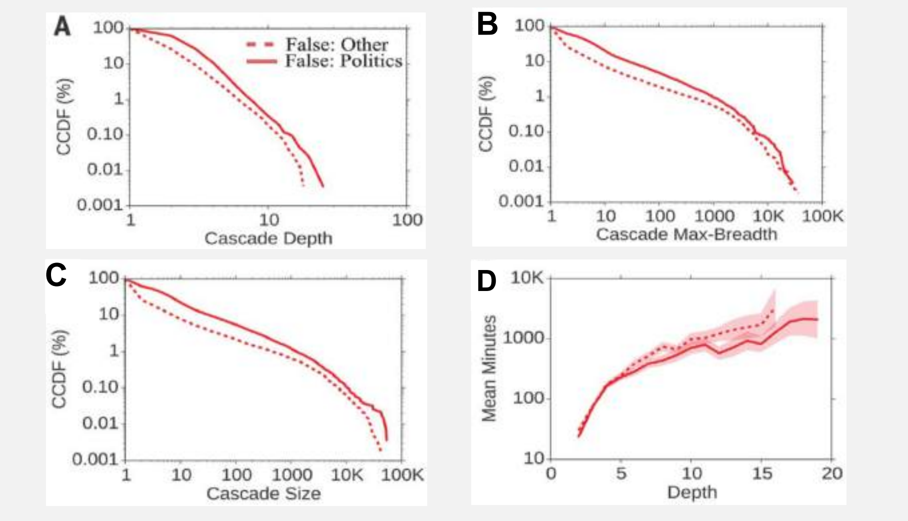
| Social Context |
|---|
| **Propagation-based** |
| **Credibility-based** |

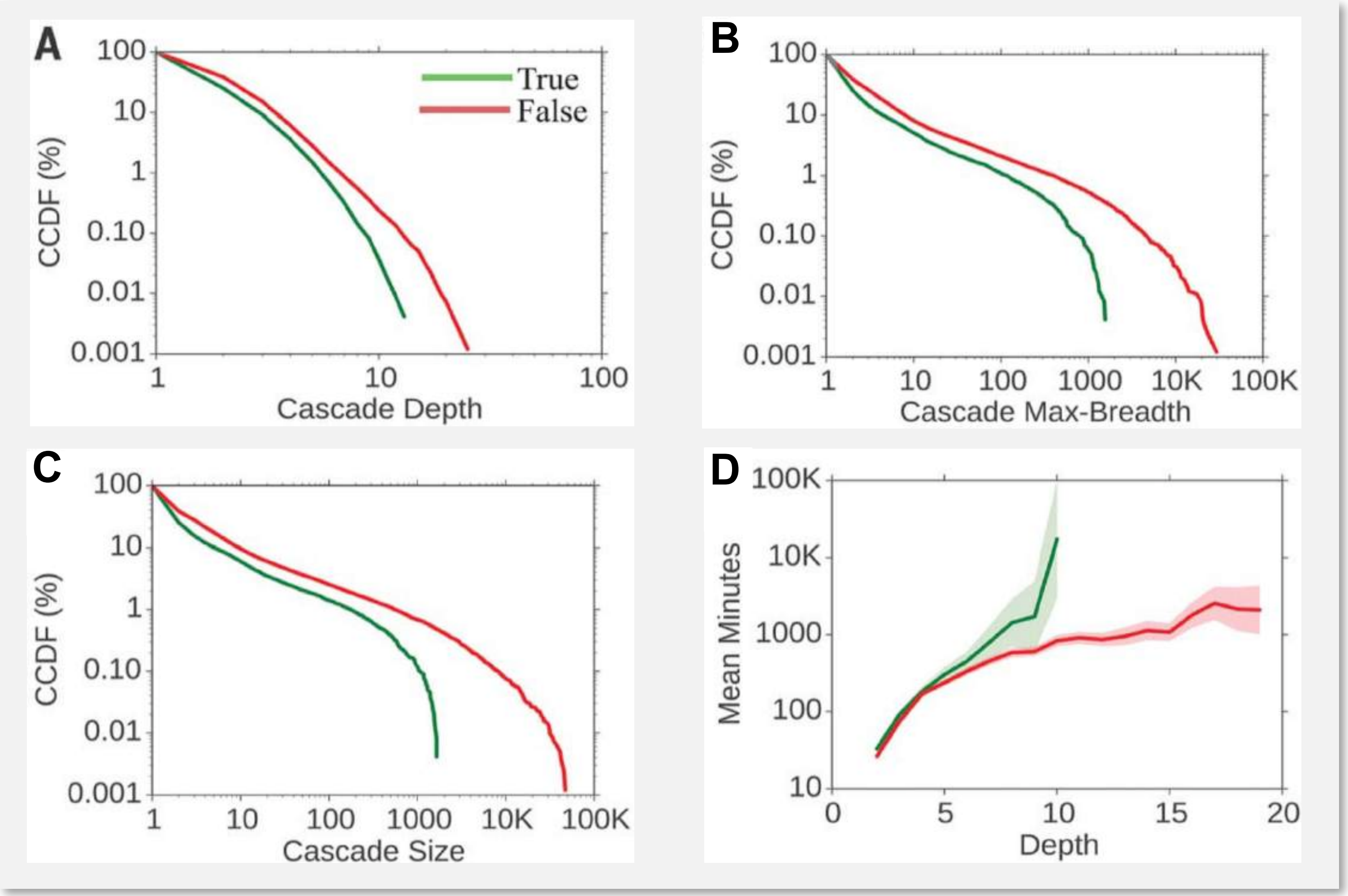
# Propagation-based Fake News Detection

Xinyi Zhou, Ph.D. Candidate
Data Lab, EECS Department, Syracuse University
zhouxinyi@data.syr.edu www.xzhou.net

# NEWS CASCADE



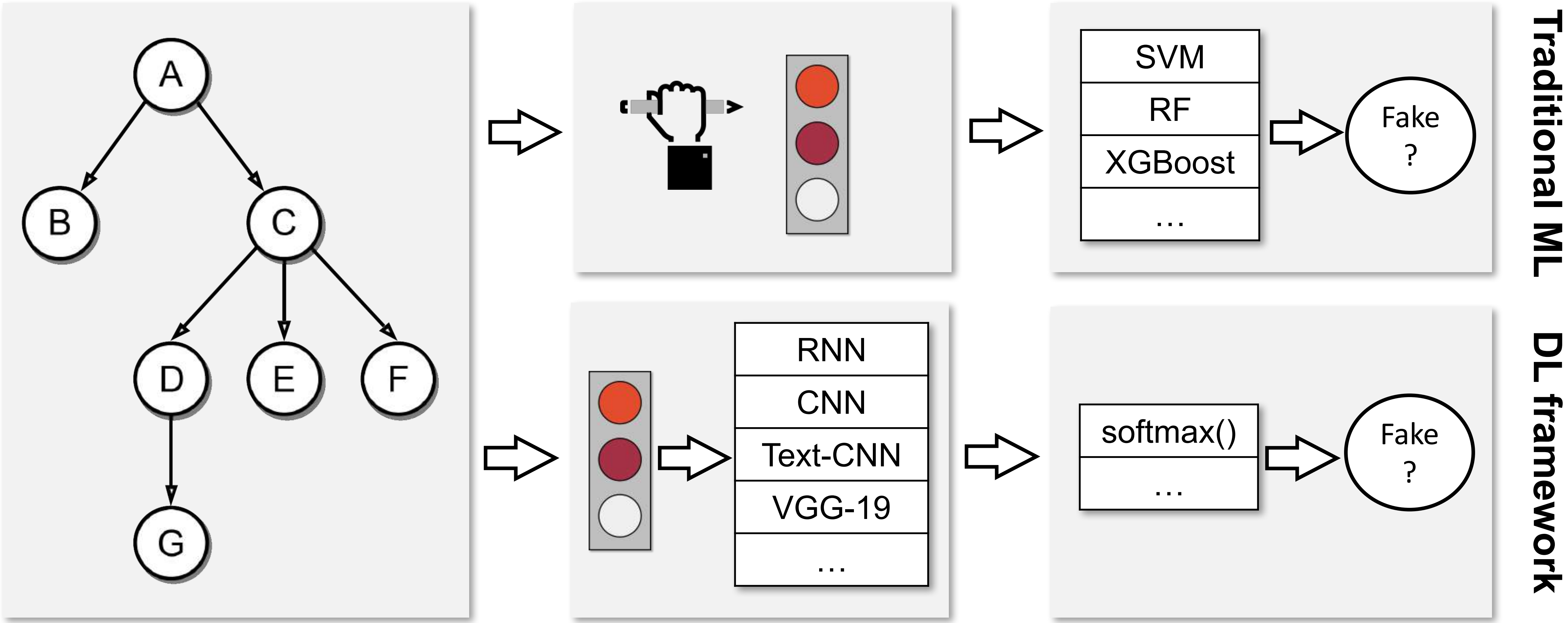



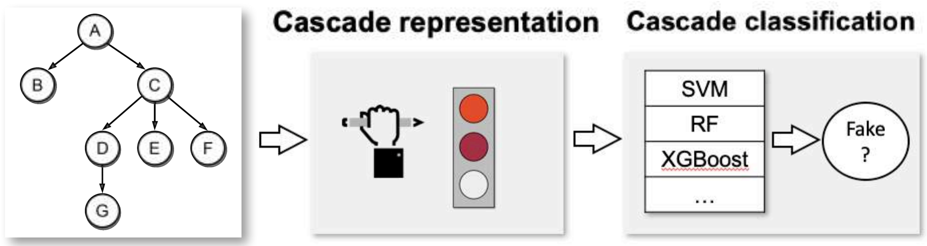



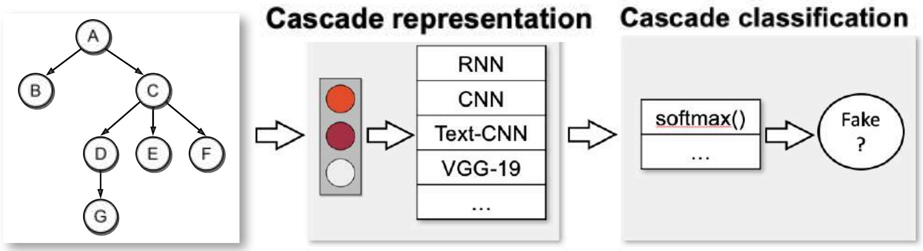

R. Zafarani, X. Zhou, K. Shu, H. Liu

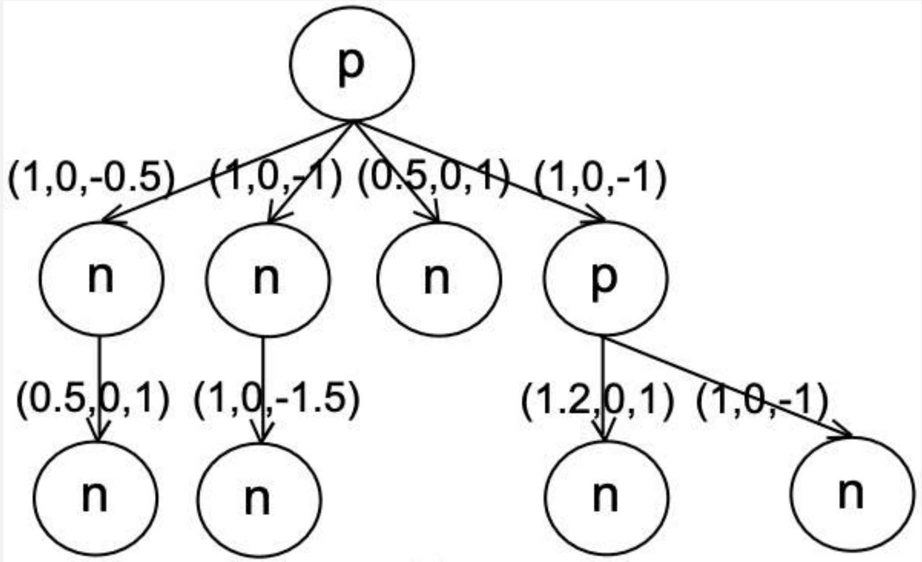
## Cascade representation

## Cascade classification



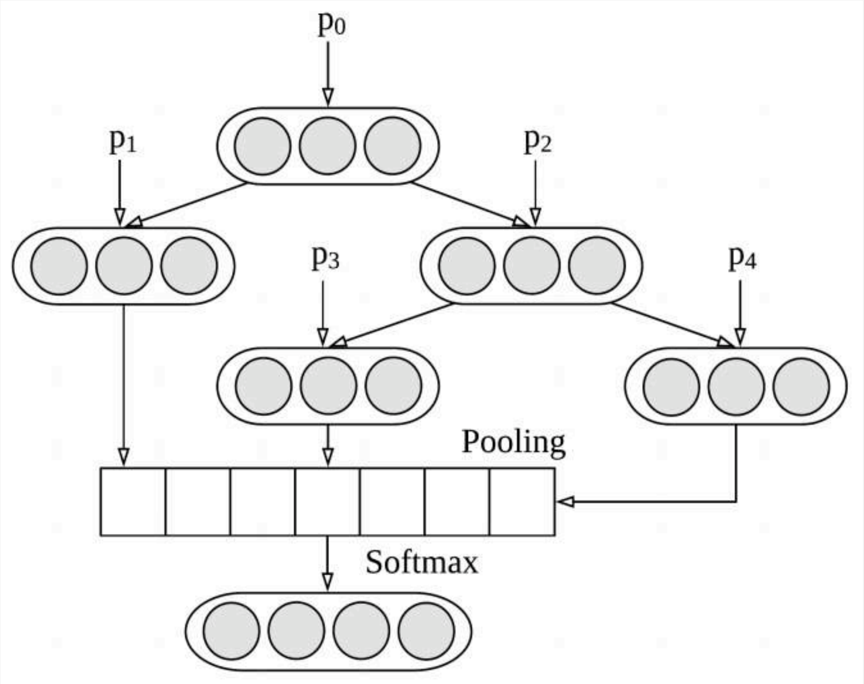



**Computational expense** → **Prune**

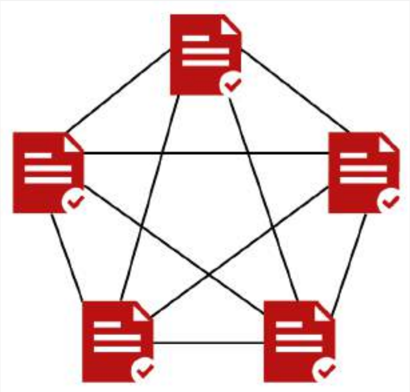
K. Wu, et al. False Rumors Detection on Sina Weibo by Propagation Structures. ICDE'15

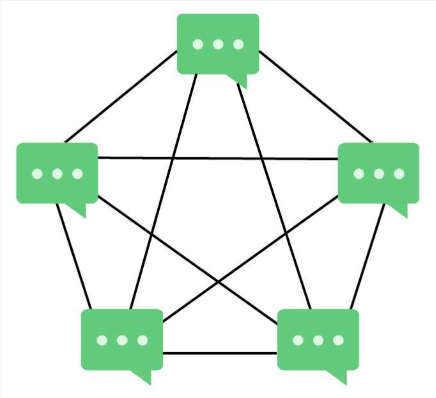



J. Ma, et al. Rumor Detection on Twitter with Tree-structure Recursive Neural Networks. ACL'18

# HOMOGENOUS NETWORK

## Stance Network

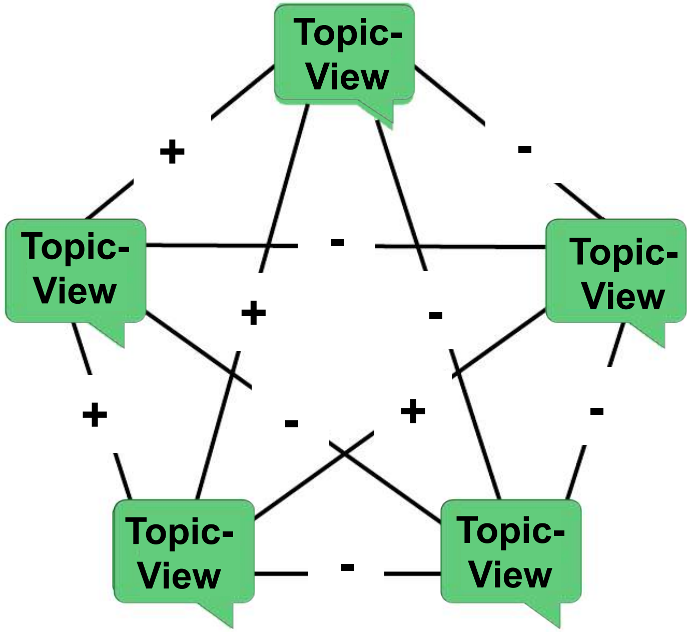




News article

User post

Similarity of text, stance, topic, etc.

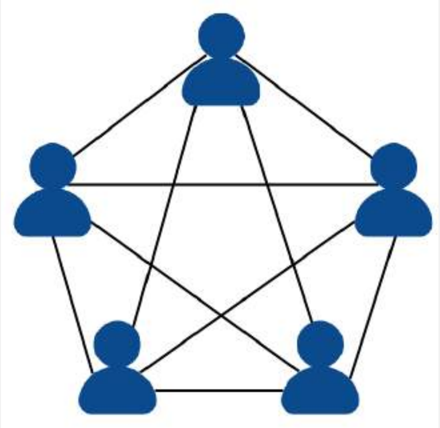



$$\arg\min_{\mathbf{c}} \underbrace{\mu||\mathbf{c}-\mathbf{c}_0||^2}_{\text{Fitting constraint}} \underbrace{+(1-\mu)\sum_{i,j=1}^{n} \mathbf{A}_{ij}\left(\frac{\mathbf{c}_i}{\sqrt{\mathbf{D}_{ii}}}-\frac{\mathbf{c}_j}{\sqrt{\mathbf{D}_{jj}}}\right)^2}_{\text{Smoothness constraint}}$$

Z. Jin, et al. News Verification by Exploiting Conflicting Social Viewpoints in Microblogs. AAAI’16

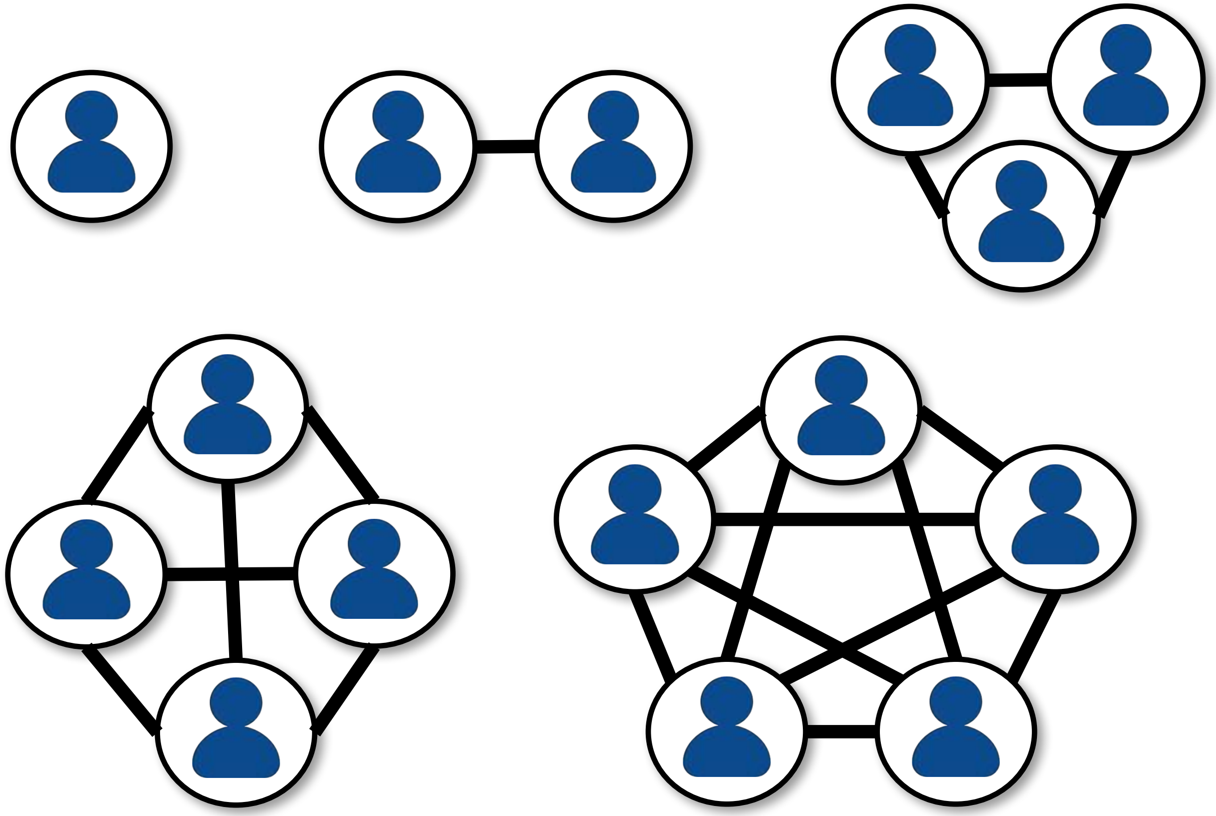
# HOMOGENOUS NETWORK

## Social Network



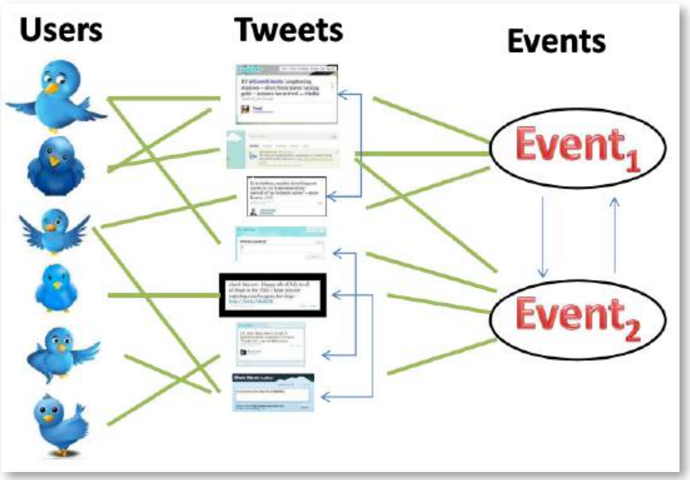
Spreader

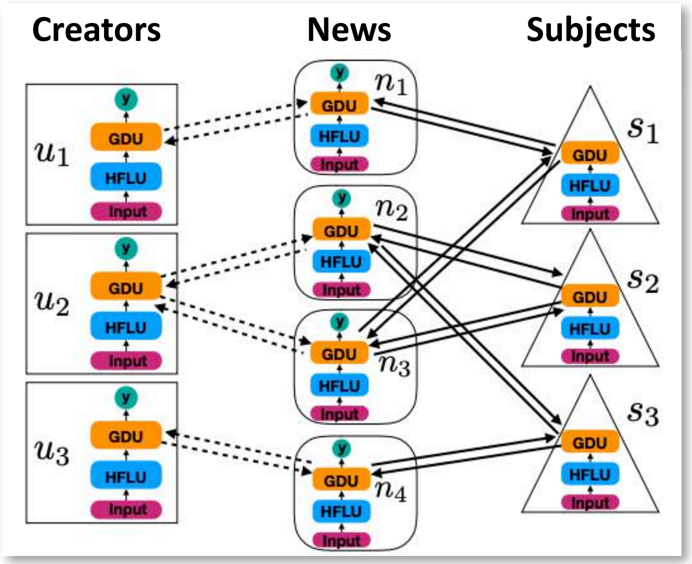
Social connection

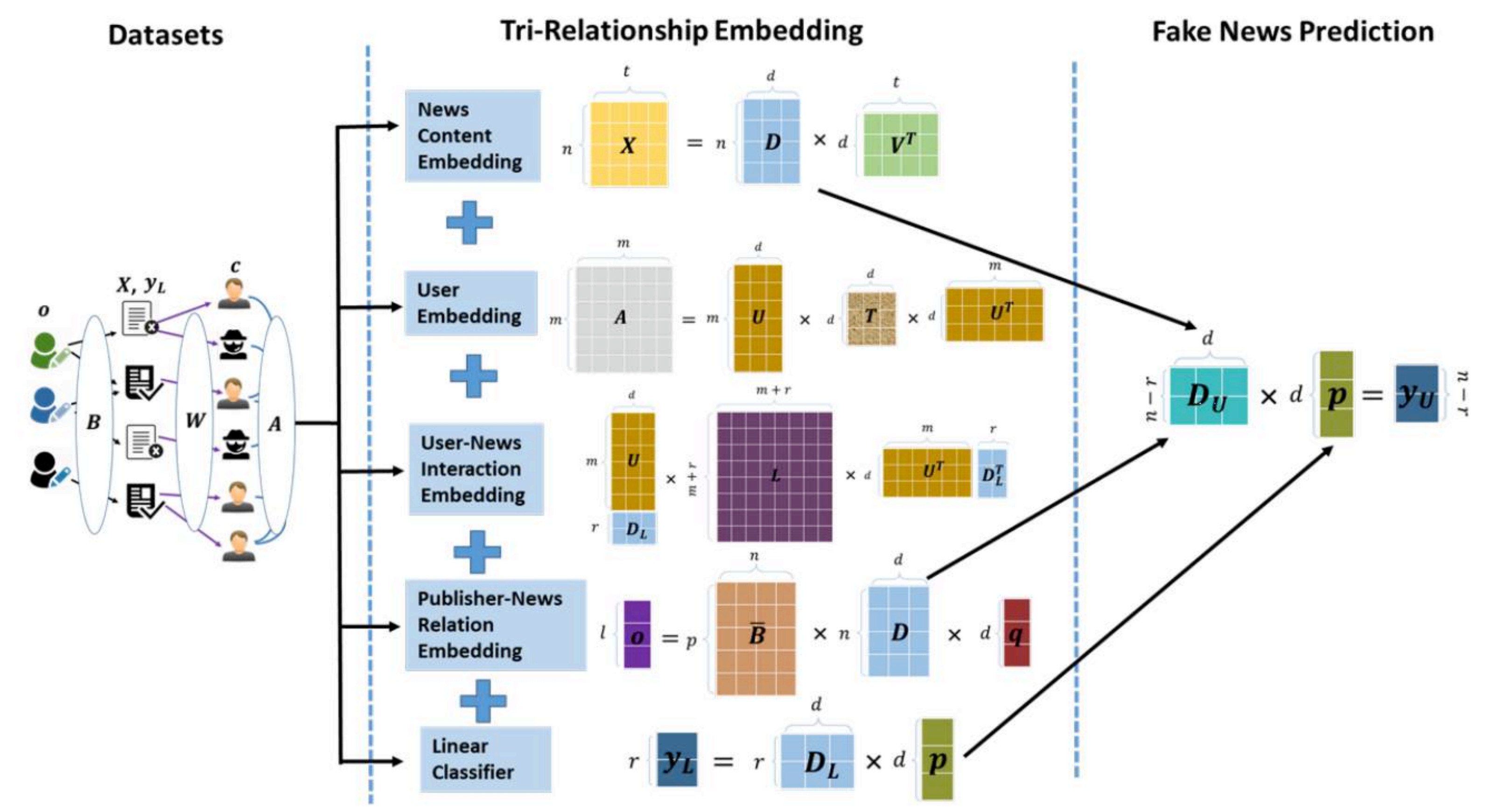


X. Zhou and R. Zafarani. Network-based Fake News Detection: A Pattern-driven Model. arXiv, 2019

# HETROGENEOUS NETWORK



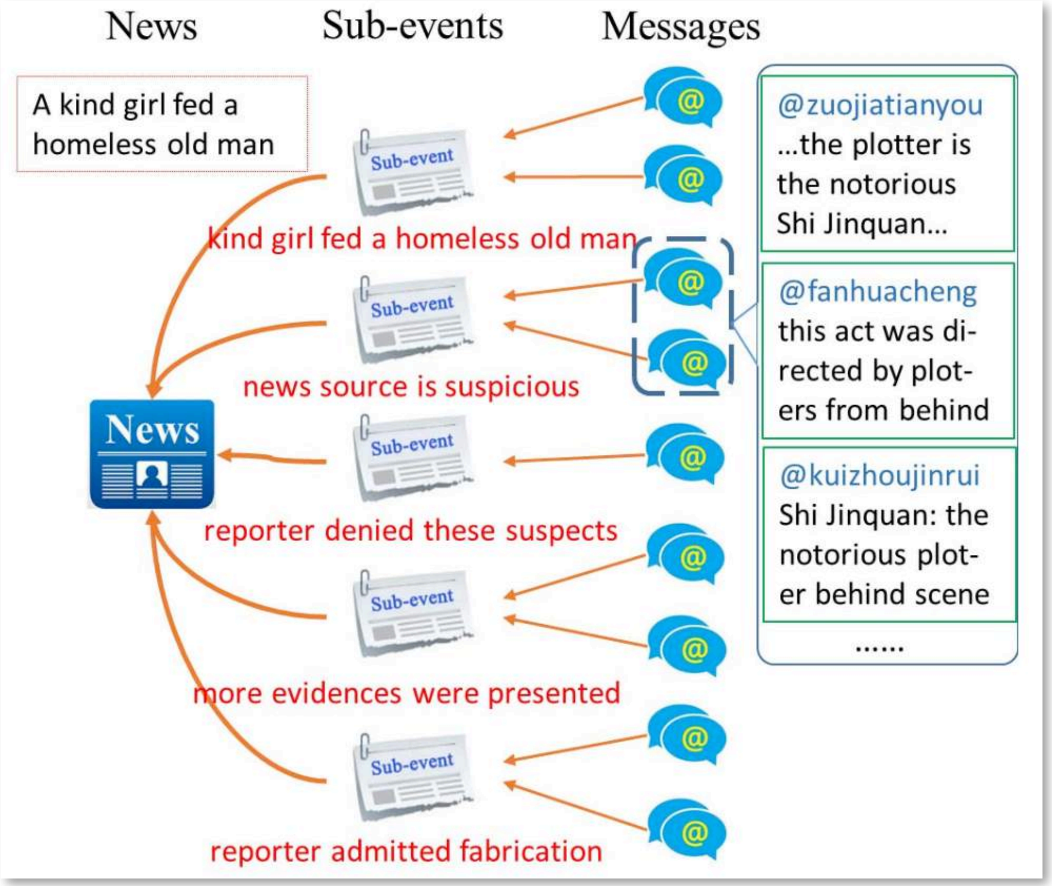



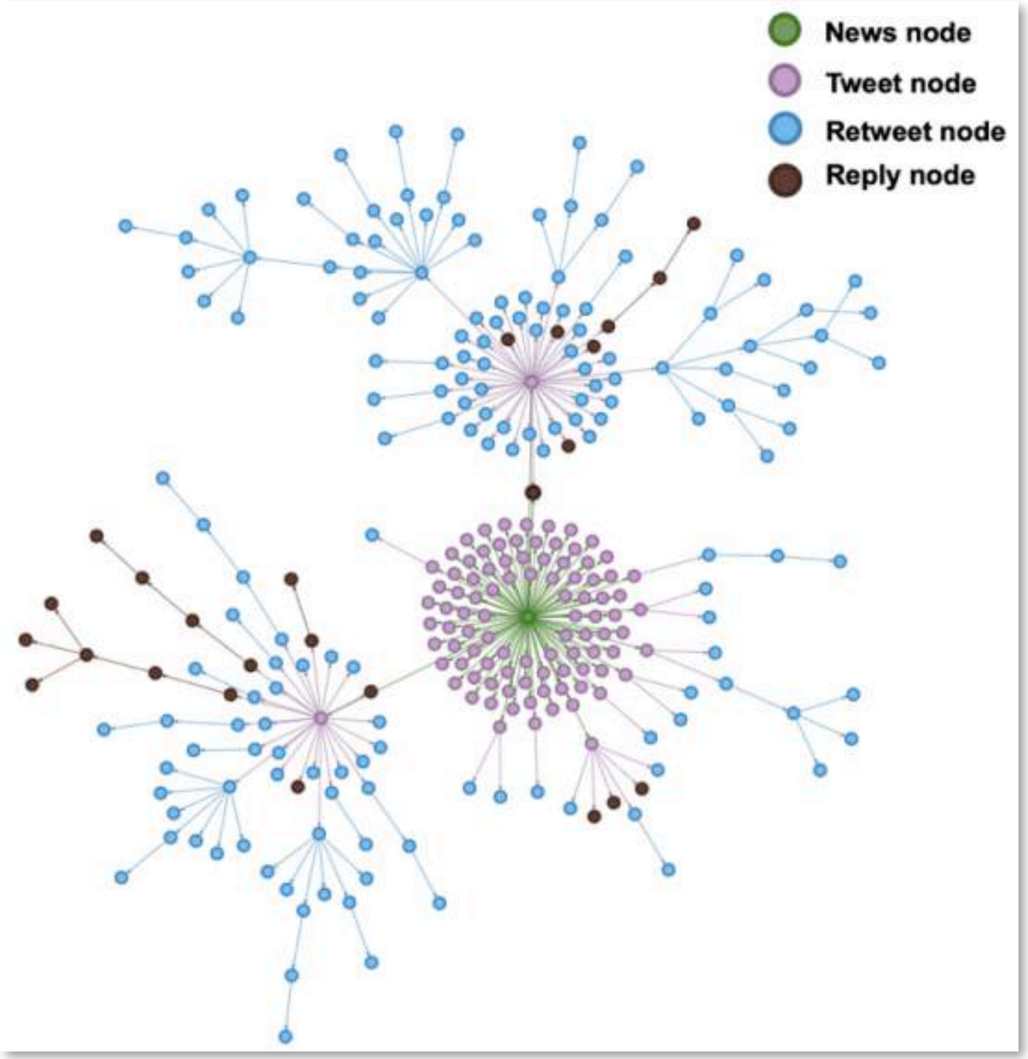




$$\min_{\mathbf{D},\mathbf{U},\mathbf{V},\mathbf{T}\geq 0,\mathbf{p},\mathbf{q}} \|\mathbf{X}-\mathbf{D}\mathbf{V}^T\|_F^2 + \alpha\|\mathbf{Y}\odot(\mathbf{A}-\mathbf{U}\mathbf{T}\mathbf{U}^T)\|_F^2$$
$$+\beta \mathrm{tr}(\mathbf{H}^T\mathbf{L}\mathbf{H}) + \gamma\|\mathbf{e}\odot(\bar{\mathbf{B}}\mathbf{D}\mathbf{q}-\mathbf{o})\|_2^2$$
$$+\eta\|\mathbf{D}_L\mathbf{p}-\mathbf{y}_L\|_2^2 + \lambda R$$

M. Gupta, et al. Evaluating Event Credibility on Twitter. SDM'12
K. Shu, et al. Beyond News Contents: The Role of Social Context for Fake News Detection. WSDM'19.
J. Zhang, et al. Fake News Detection with Deep Diffusive Network Model. arXiv, 2018

# HIERARCHICAL NETWORK



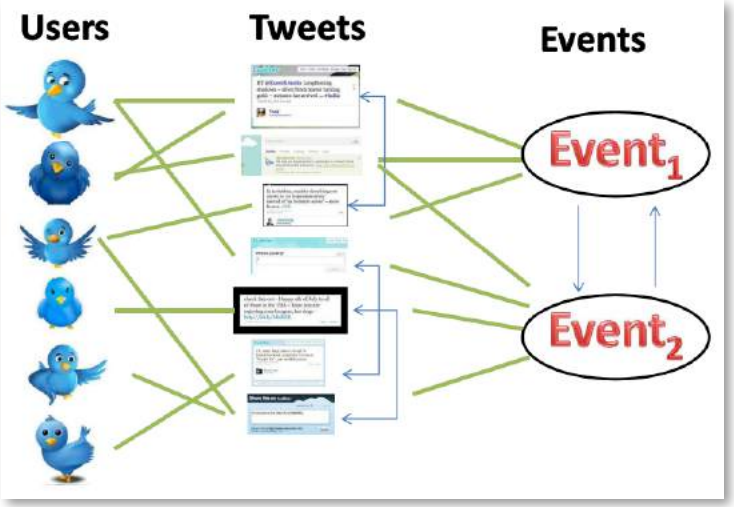



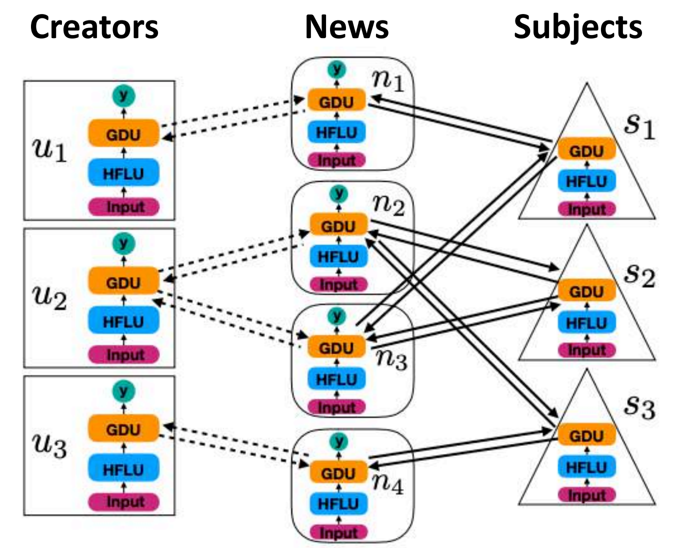

Z. Jin, et al. News Credibility Evaluation on Microblog with a Hierarchical Propagation Model, ICDM'14

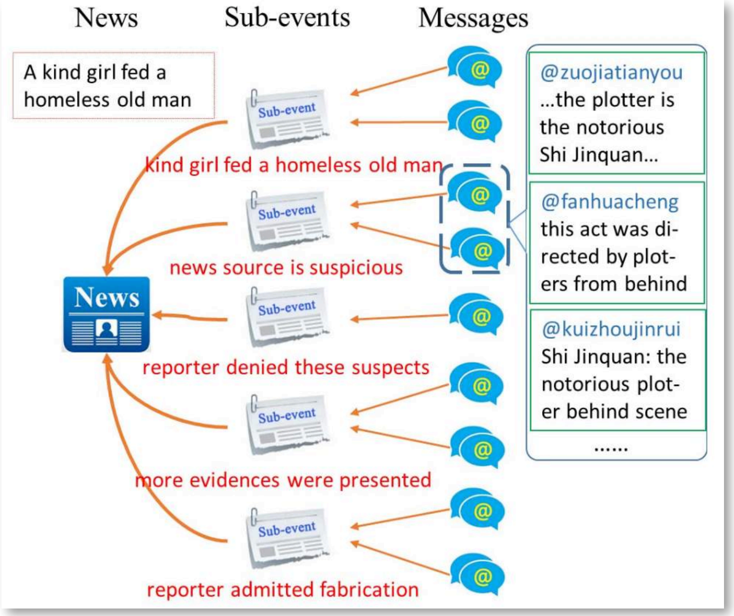
K. Shu, et al. Hierarchical Propagation Networks for Fake News Detection: Investigation and Exploitation. arXiv, 2019

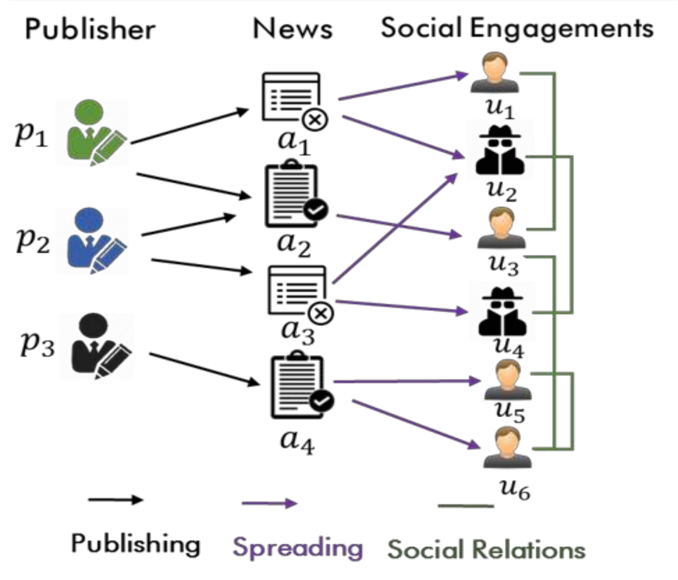
# Credibility-based Fake News Detection

Xinyi Zhou, Ph.D. Candidate
Data Lab, EECS Department, Syracuse University
zhouxinyi@data.syr.edu www.xzhou.net



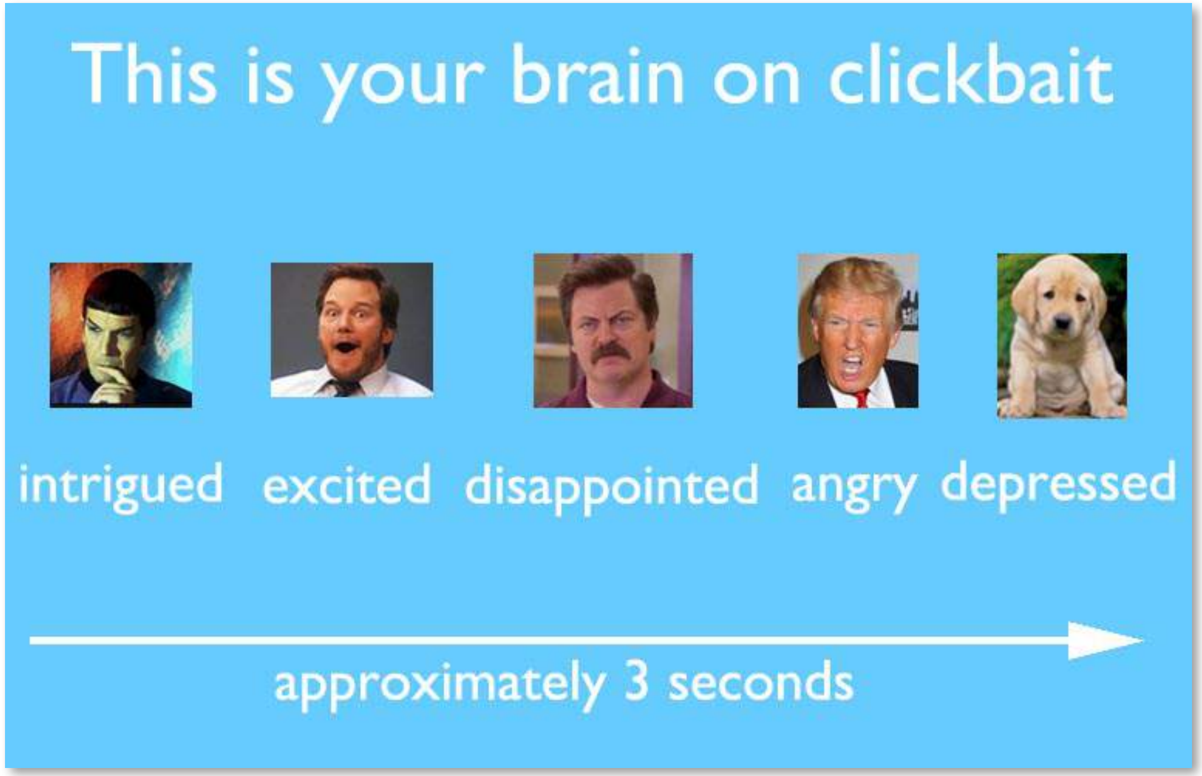



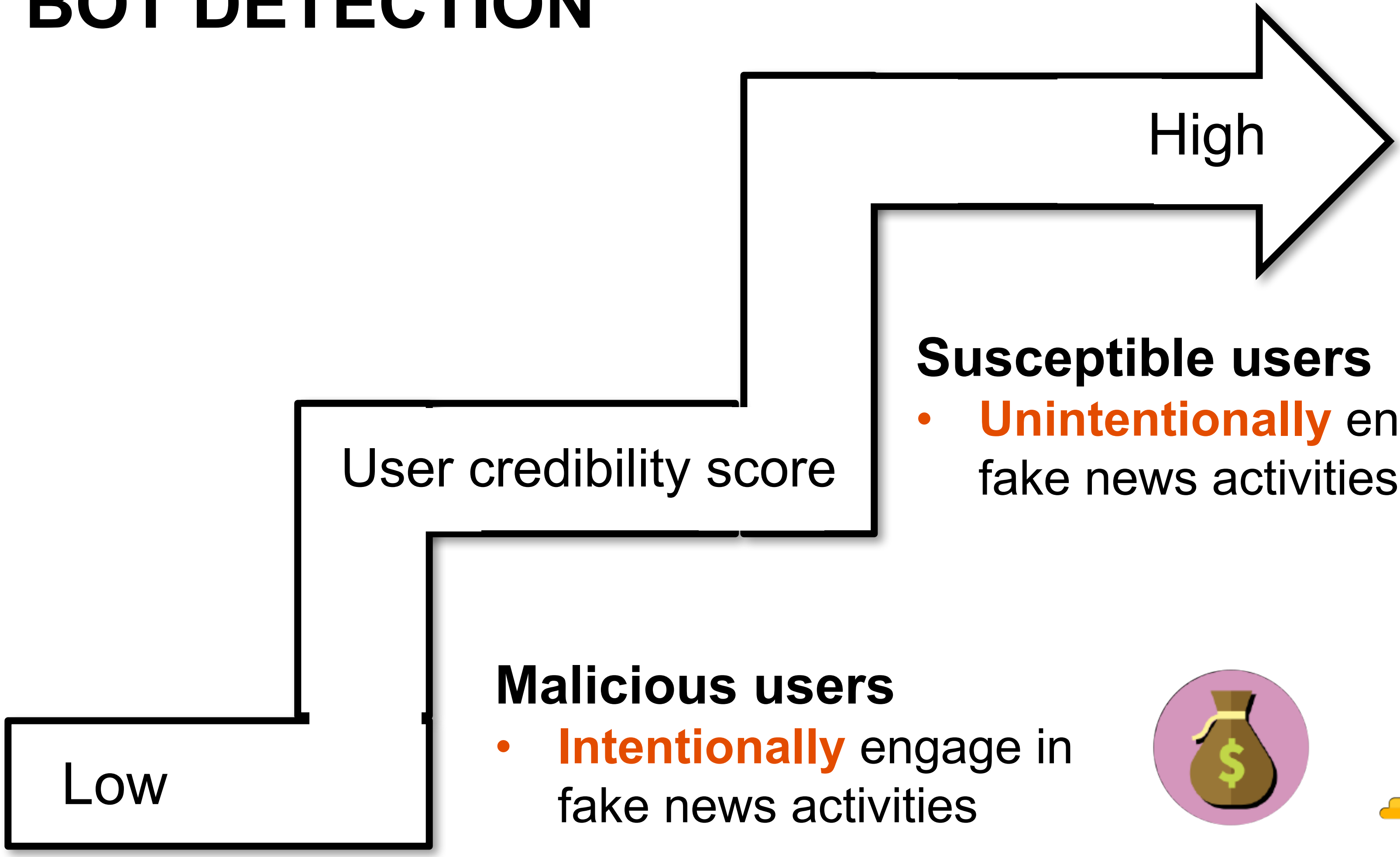

# It overlaps with propagation-based fake news detection…

# HEADLINE CREDIBILITY & CLICKBAIT DETECTION

**Fake News Early Detection: A Theory-driven Model**
Xinyi Zhou, Atishay Jain, Vir V. Phoha, Reza Zafarani

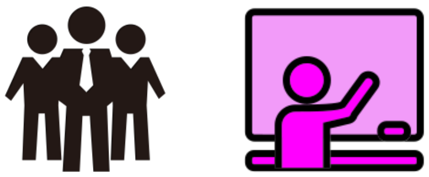



# USER CREDIBILITY & BOT DETECTION




**Insusceptible users**
- Immune to fake news

**Susceptible users**
- **Unintentionally** engage in fake news activities



**Malicious users**
- **Intentionally** engage in fake news activities

# THE CHALLENGES

I. Fake news early detection…
- Effectively detecting fake news when limited social context information is available

II. Empirical relationships between fake news and clickbait…
- Dataset containing the ground truth of both

III. Assessing user intention in fake news activities…

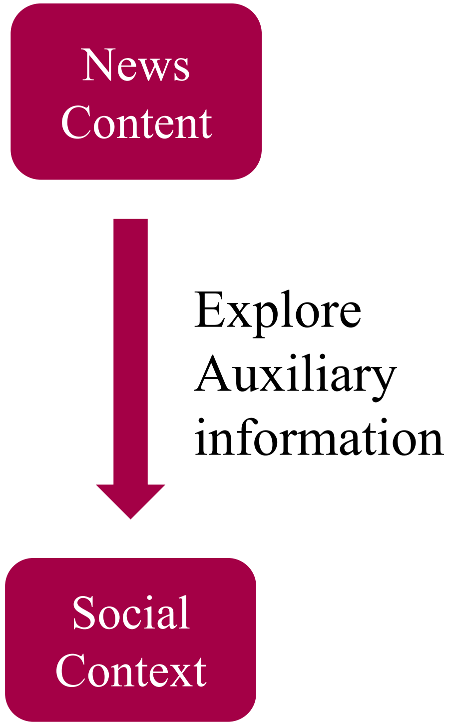
# Beyond News Contents: The Role of Social Context for Fake News Detection

**Kai Shu, Suhang Wang and Huan Liu**

**WSDM 2019**

# Fake News Detection on Social Media - Challenges

- News Content
  - Fake news pieces are intentionally written to mislead users
  - Diverse in terms of topics, styles, and media platforms
- Social Context
  - Social engagements are massive, incomplete, unstructured, and noisy
  - Effective methods are sought to differentiate credible users, extract useful post features, and exploit network interactions



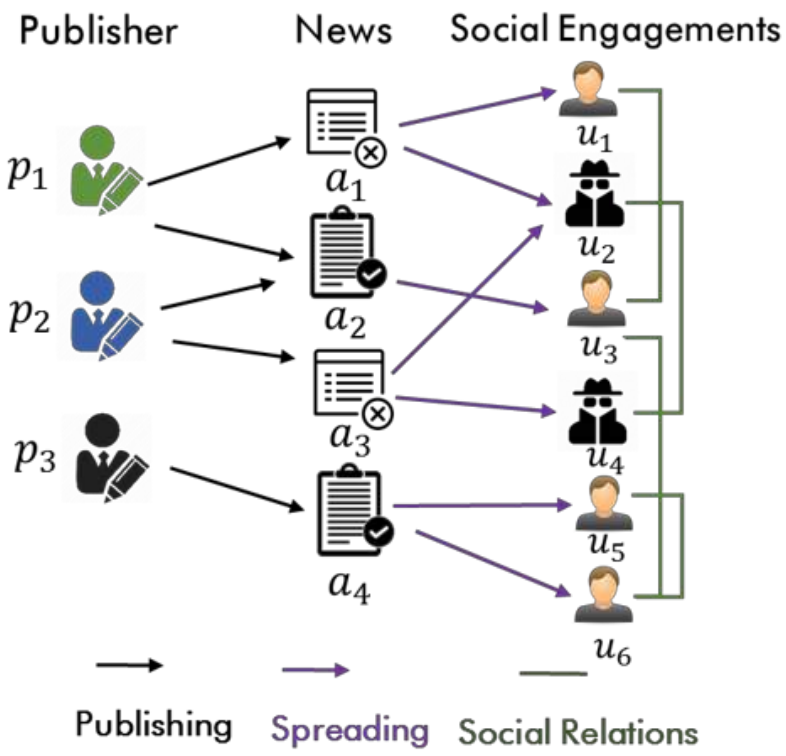
# Fake News Detection – Multi-Source

- A typical  news dissemination system on social media
  - Entities: publisher p, news a, and social media users u
  - Relations: **publishing**, spreading, social relations

➢ *Publishing* Publisher with partisan bias are more likely to post fake news

e.g., $p_1 \to a_1 \quad p_2 \to a_3$

$p_3 \to a_4$




➢ *spreading*

Low credibility users on social media are likely to share fake news, e.g., $a_1 \to u_2 \; a_3 \to u_2$

➢ *social*

Users form relationship with like-minded people

e.g., $u_2 \leftrightarrow u_4 \; u_3 \leftrightarrow u_1$

X. Zhou, R. Zafarani, K. Shu, H. Liu

# Tri-Relationship Embedding (TriFN)

- News content embedding
  - Content modeling
  - Publisher news relation embedding
- Social Context embedding
  - Basic user feature representation
  - User news engagement modeling
- We jointly combine news content embedding and social context embedding for fake news detection

$$\min_{\mathbf{D},\mathbf{V}\geq 0} \| \mathbf{X} - \mathbf{D}\mathbf{V}^T \|_F^2 + \lambda(\|\mathbf{D}\|_F^2 + \|\mathbf{V}\|_F^2)$$

$$\min \| \bar{\mathbf{B}}\mathbf{D}\mathbf{Q} - \mathbf{o} \|_2^2 + \lambda\|\mathbf{Q}\|_2^2$$

$$\min_{\mathbf{U},\mathbf{T}\geq 0} \|\mathbf{Y} \odot (\mathbf{A} - \mathbf{U}\mathbf{T}\mathbf{U}^T)\|_F^2 + \lambda(\|\mathbf{U}\|_F^2 + \|\mathbf{T}\|_F^2)$$

$$\min \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{r} \mathbf{W}_{ij}\mathbf{c}_i(1 - \frac{1+\mathbf{y}_{Lj}}{2})\|\mathbf{U}_i - \mathbf{D}_{L_j}\|_2^2}_{\text{True news}} + \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{r} \mathbf{W}_{ij}(1-\mathbf{c}_i)(\frac{1+\mathbf{y}_{Lj}}{2})\|\mathbf{U}_i - \mathbf{D}_{L_j}\|_2^2}_{\text{Fake news}}$$

X. Zhou, R. Zafarani, K. Shu, H. Liu

# Evaluation Setting

- Datasets: FakeNewsNet with information for news conte social context and ground truth labels from fact-checkin websites
- Compared baselines:
  - RST: rhetorical relations among the words in the text
  - LIWC: lexicons falling into psycholinguistic categories
  - Castillo: features from user profiles, social networks
  - RST+Castillo
  - LIWC+Castillo

News Content

Social Context

News Content + Social Context

Table 1: The statistics of FakeNewsNet dataset

| Platform | BuzzFeed | PolitiFact |
|---|---|---|
| # Users | 15,257 | 23,865 |
| # Engagements | 25,240 | 37,259 |
| # Social Links | 634,750 | 574,744 |
| # Candidate news | 182 | 240 |
| # True news | 91 | 120 |
| # Fake news | 91 | 120 |
| # Publisher | 9 | 91 |

# Evaluation Results - Detection Performance

- Social context based features are more effective than news content based features
- TriFN performs the best than other methods using both news content and social context information

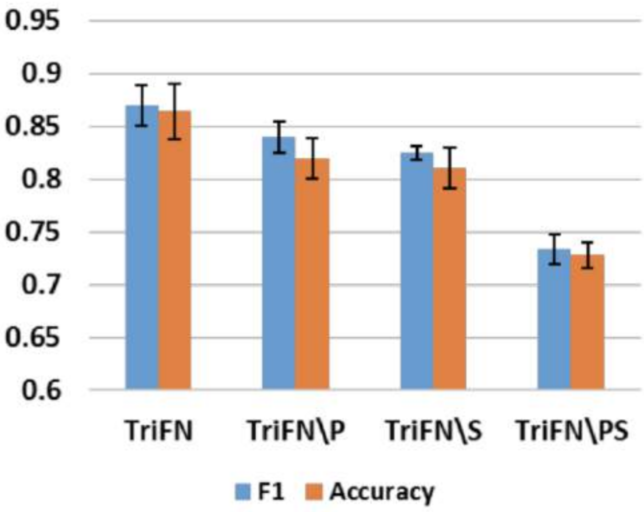
Table 2: Performance comparison for fake news detection

| Datasets | Metric | RST | LIWC | Castillo | RST+Castillo | LIWC+Castillo | TriFN |
|---|---|---|---|---|---|---|---|
| **BuzzFeed** | Accuracy | $0.610 \pm 0.023$ | $0.655 \pm 0.075$ | $0.747 \pm 0.061$ | $0.758 \pm 0.030$ | $0.791 \pm 0.036$ | $\mathbf{0.864 \pm 0.026}$ |
| | Precision | $0.602 \pm 0.066$ | $0.683 \pm 0.065$ | $0.735 \pm 0.080$ | $0.795 \pm 0.060$ | $0.825 \pm 0.061$ | $\mathbf{0.849 \pm 0.040}$ |
| | Recall | $0.561 \pm 0.057$ | $0.628 \pm 0.021$ | $0.783 \pm 0.048$ | $0.784 \pm 0.074$ | $0.834 \pm 0.094$ | $\mathbf{0.893 \pm 0.013}$ |
| | F1 | $0.555 \pm 0.057$ | $0.623 \pm 0.066$ | $0.756 \pm 0.051$ | $0.789 \pm 0.056$ | $0.802 \pm 0.023$ | $\mathbf{0.870 \pm 0.019}$ |
| **PolitiFact** | Accuracy | $0.571 \pm 0.039$ | $0.637 \pm 0.021$ | $0.779 \pm 0.025$ | $0.812 \pm 0.026$ | $0.821 \pm 0.052$ | $\mathbf{0.878 \pm 0.020}$ |
| | Precision | $0.595 \pm 0.032$ | $0.621 \pm 0.025$ | $0.777 \pm 0.051$ | $0.823 \pm 0.040$ | $0.856 \pm 0.071$ | $\mathbf{0.867 \pm 0.034}$ |
| | Recall | $0.533 \pm 0.031$ | $0.667 \pm 0.091$ | $0.791 \pm 0.026$ | $0.792 \pm 0.026$ | $0.767 \pm 0.120$ | $\mathbf{0.893 \pm 0.023}$ |
| | F1 | $0.544 \pm 0.042$ | $0.615 \pm 0.044$ | $0.783 \pm 0.015$ | $0.793 \pm 0.032$ | $0.813 \pm 0.070$ | $\mathbf{0.880 \pm 0.017}$ |

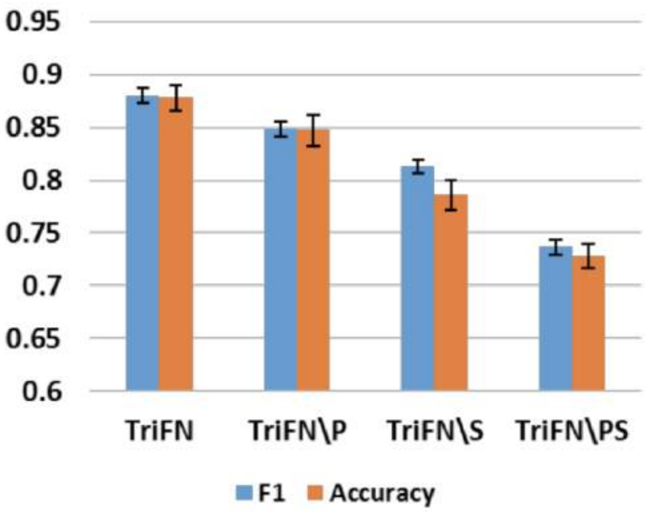
News Content Social Context News Content + Social Context

# Evaluation Results - Component Analysis and Early Detection

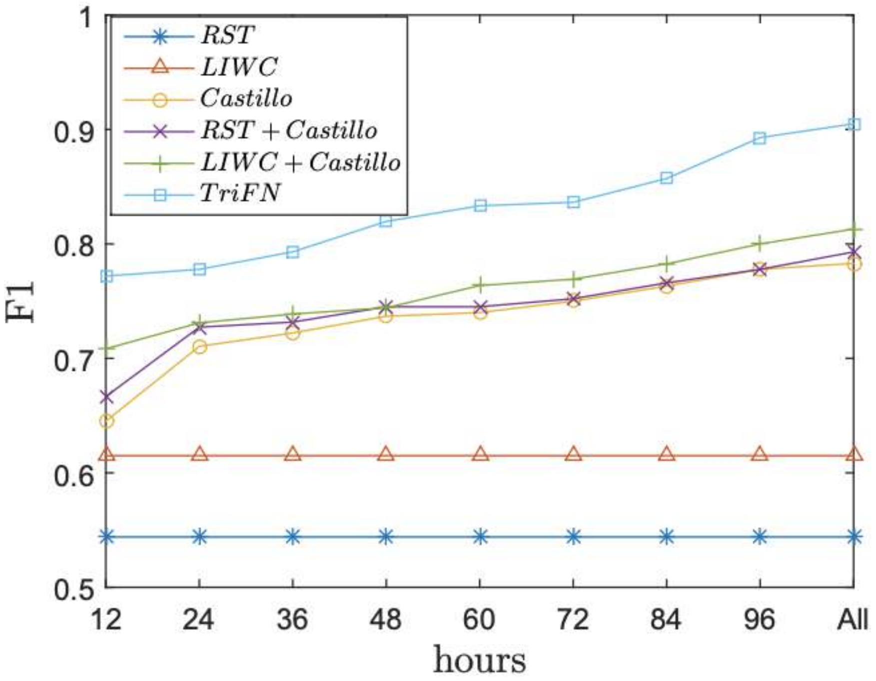
- Both publisher-news and news-user relations can contribute to the performance improvement of TriFN
- TriFN consistently achieves best performances in the early stage of news dissemination




(a) BuzzFeed




(b) PolitiFact



# 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

# Unsupervised Fake News Detection: A Generative Approach

Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu

AAAI 2019

# Unsupervised Fake News Detection

- Existing methods are mainly supervised, which require extensive amount of time and labor to build a reliably annotated dataset.
- We aim to build an unsupervised fake news detection method by modeling user opinions and user credibility


Agreeing the authenticity of the news


Doubting the authenticity of the news

# Unsupervised Fake News Detection - challenges

- User social engagements are usually unstructured, large-scale, and noisy
- User opinions may be conflicting and unreliable, as the users usually have different degrees of credibility in identifying fake news

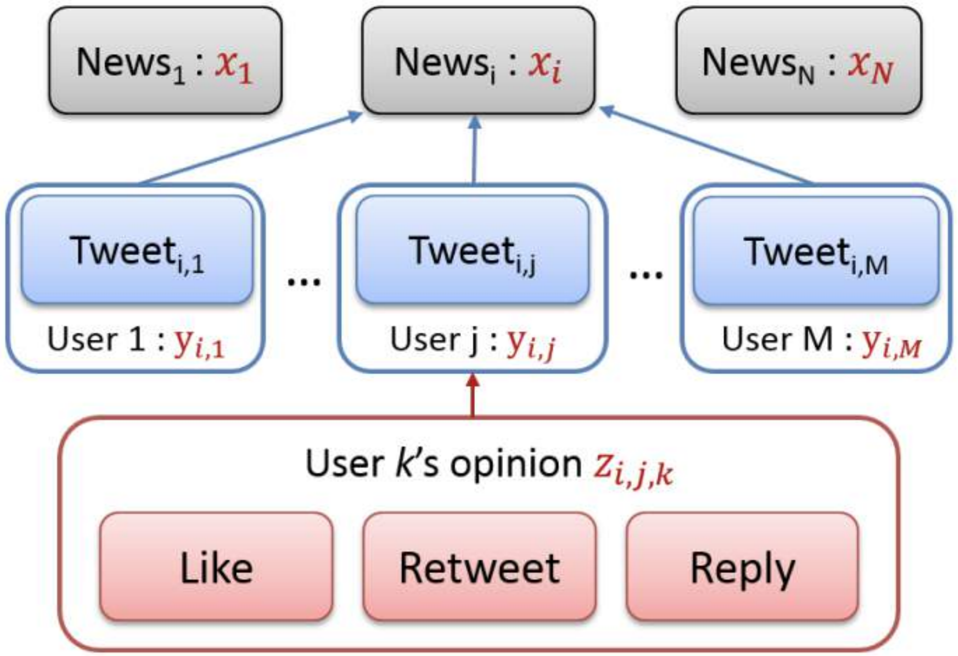
- The relationships among news, tweets, and users on social media form more complicated topologies
- Existing truth discovery methods mainly focus on "source-item" paths, and cannot be directly applied

# The hierarchical user engagement structure

- We build a hierarchical user engagement structure for each news
  - $x_i$ is a random variable denoting the label of $news_i$
  - $y_{i,j}$ denotes the opinion with sentiment of verified user $j$ to $news_i$
  - $z_{i,j,k}$ is the opinion of unverified user $k$ to $news_i$
    - Like: opinion same with $y_{i,j}$
    - Reply: sentiment score of the reply
    - Retweet: opinion same with $y_{i,j}$

Verified User

Unverified User



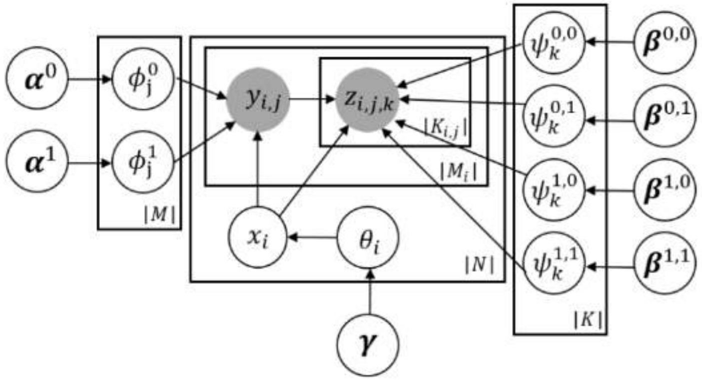
# The Proposed Probabilistic Model (UFD)

- For each news $i$, $x_i$ is generated from Bernoulli distribution

$$x_i \sim \text{Bernoulli}(\theta_i)$$

- For verified user $j$ $\quad y_{i,j} \sim \text{Bernoulli}(\phi_j^{x_i})$
  - $\phi_j^1$ ($\phi_j^0$) the probability that the user $j$ thinks a news piece is real given the truth estimation of the news is true and fake
- For unverified $k$, $\quad z_{i,j,k} \sim \text{Bernoulli}(\psi_k^{x_i, y_{i,j}})$
  - the opinion is likely to be influenced by the news itself and the verified users' opinions

$$\psi_k^{0,0} := p(z_{i,j,k} = 1 | x_i = 0, y_{i,j} = 0)$$
$$\psi_k^{0,1} := p(z_{i,j,k} = 1 | x_i = 0, y_{i,j} = 1)$$
$$\psi_k^{1,0} := p(z_{i,j,k} = 1 | x_i = 1, y_{i,j} = 0)$$
$$\psi_k^{1,1} := p(z_{i,j,k} = 1 | x_i = 1, y_{i,j} = 1)$$



# Evaluation Results - Detection Performance

- Majority voting achieves the worst performance since it equally aggregates the users' opinions without considering user's credibility degree
- The proposed framework UFD can achieve best performance comparing with other unsupervised truth discovery methods
- We can also discover the top-k creidible users, and these users are mostly expert journalists, professional news reporters

Table 2: Performance comparison on LIAR dataset

| Methods | Accuracy | True | | | Fake | | |
|---|---|---|---|---|---|---|---|
| | | Precision | Recall | F1-score | Precision | Recall | F1-scor |
| Majority Voting | 0.586 | 0.624 | 0.628 | 0.626 | 0.539 | 0.534 | 0.537 |
| TruthFinder | 0.634 | 0.650 | 0.679 | 0.664 | 0.615 | 0.583 | 0.599 |
| LTM | 0.641 | 0.654 | 0.691 | 0.672 | 0.624 | 0.583 | 0.603 |
| CRH | 0.639 | 0.653 | 0.687 | 0.669 | 0.621 | 0.583 | 0.601 |
| **UFD** | **0.759** | **0.766** | **0.783** | **0.774** | **0.750** | **0.732** | **0.741** |

Table 4: Top accurate verified users on two datasets

| User | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| amy_hollyfield | 1.0 | 1.0 | 1.0 |
| politico | 0.909 | 0.833 | 1.0 |
| loujacobson | 0.84 | 0.842 | 0.833 |
| dcexaminer | 0.833 | 0.818 | 0.857 |
| FoxNews | 0.818 | 0.714 | 1.0 |

# Summary

- We study the novel problem of unsupervised fake news detection, a much desired scenario in the real world
- We propose a probabilistic model to consider the user opinions and user credibility in a hierarchical engagement structure
- We demonstrate the effectiveness of the proposed framework in real-world datasets
- **Future work**
    - Incorporating user profiles and news contents into unsupervised models
    - Building semi-supervised models with limited engagements information

# Deep Headline Generation for Clickbait Detection

Kai Shu, Suhang Wang, Thai Le, Dongwon Lee, and Huan Liu

ICDM 2018

# Clickbaits

- Clickbaits are catchy social media posts or sensational headlines that attempt to lure the readers to click



- Clickbaits can have negative societal impacts
  - clickbaits may contain sensational and inaccurate information to mislead readers and spread fake news
  - clickbaits may be used to perform clickjacking attacks by redirecting users to phishing websites

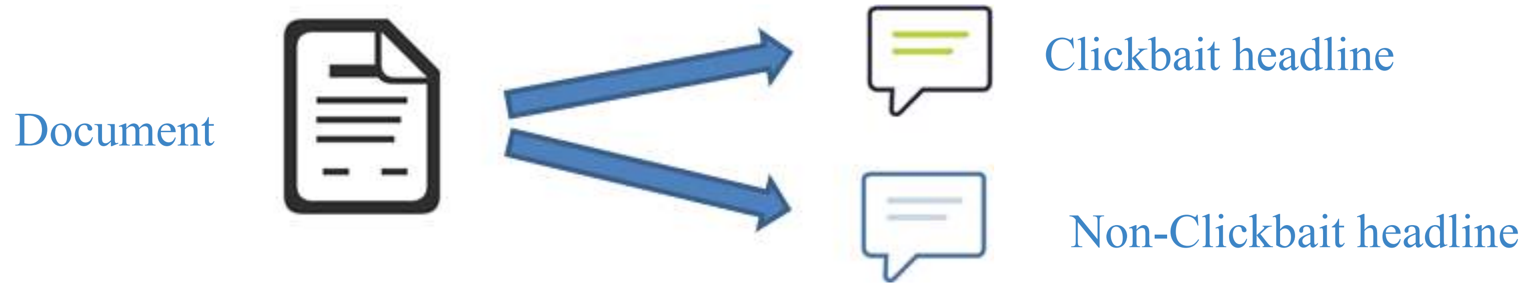
# Clickbait Detection

- Existing approaches mainly focus on extracting hand-crafted linguistic features (as traditionally done so) or building sophisticated predictive models such as deep neural networks
- However, these methods may face following limitations
    - Scale: datasets with labels are often limited
    - Distribution: imbalanced distribution of clickbaits and non-clickbaits

We aim to generate synthetic headlines with specific styles and exploit the utility to improve clickbait detection

# Headline Generation from Documents

- Goal: Generate stylized headlines that also preserve document contents




- Stylized headlines can help augment training data for clickbait detection
- Content preserved headlines make it possible to suggest a non-clickbait headline to readers after we detect a clickbait

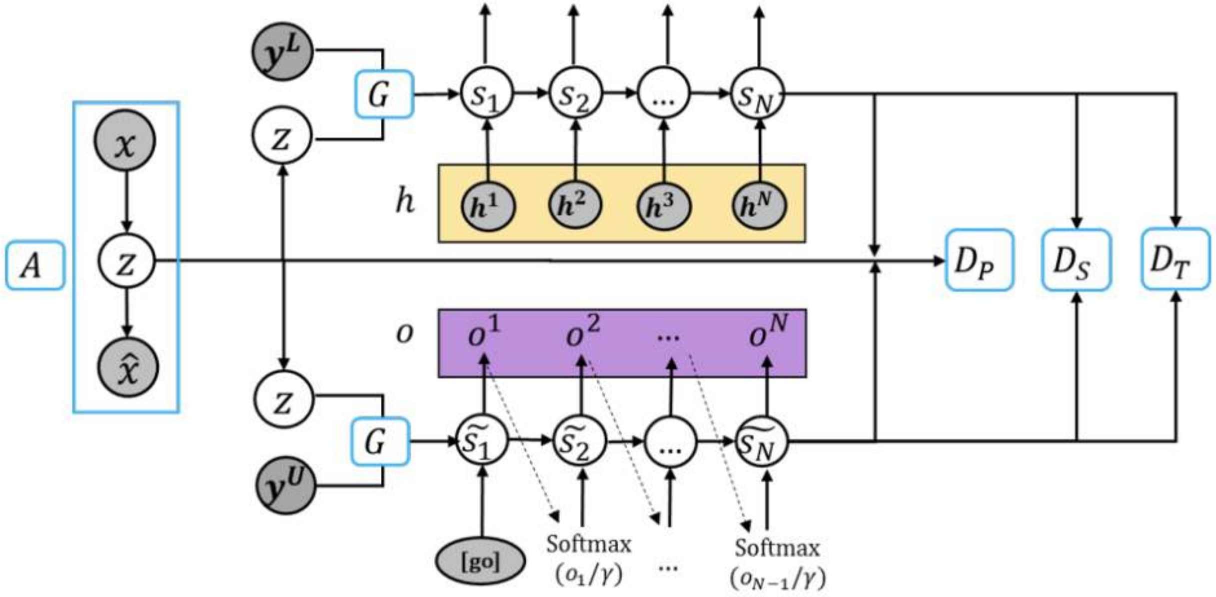
# Problem Definition

- Let $\{x_1, x_2, \ldots, x_m\}$, $\{h_1, h_2, \ldots, h_m\}$ and $\{y_1, y_2, \ldots, y_m\}$ denote the set of $m$ documents, and corresponding headlines and labels
- Giving $S = \{(x_i, h_i) | i = 1, \ldots, m\}$, learn a generator that can generate stylized headlines given a document and a style label, i.e., $o_i = f(x_i, y_i)$

- Challenges
  - How to generate realistic and readable headlines from original documents?
  - How to utilize generated headlines to augment training data for clickbait detection
  - How to generate new headlines that can preserve the content of documents and transfer the style of original headlines

# Stylized Headline Generation (SHG)

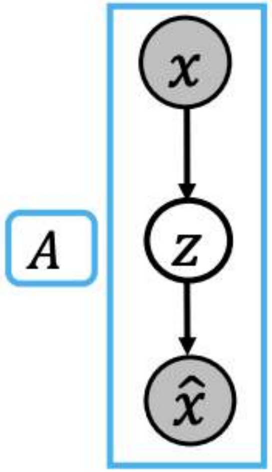
- We propose a deep learning model to generate both click-baits and non-clickbaits with style transfer
  - Generator Learning: a document autoencoder $A$, a headline generator $G$
  - Discriminator Learning: a transfer discriminator $D_T$, a style discriminator $D_S$, a pair discriminator $D_P$



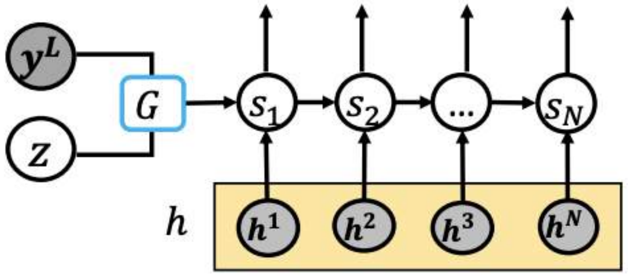
# Generator Learning

- Document autoencoder $\mathcal{A}$ extract document representation by minimizing the reconstruction error

$$\mathcal{L}_{rec}(\theta_e, \theta_d) = -\sum_{i=1}^{m} \log p(\hat{x}_i | x_i; \theta_d, \theta_e)$$

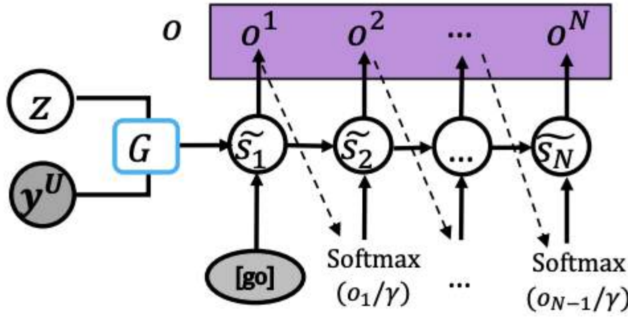



- Headline generator $G$
  - Generate stylized headline by minimizing the reconstructi error of original headline



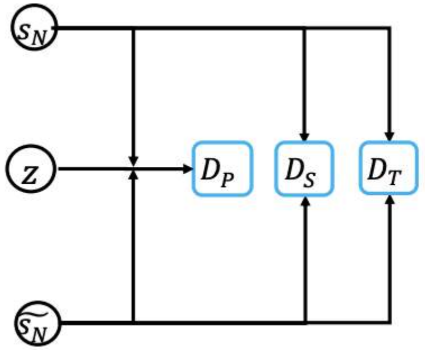

$$\mathcal{L}_G(\theta_G) = \mathbb{E}_{(x,h)\in\mathcal{S}}[-\log p_G(h|\mathbf{y}^L, \mathbf{z}))]$$

- Generate a set of new headlines $O$ with the styles $y^U$ opposite to the original headlines



# Discriminator Learning

- Discriminators regularize the representation learning of document $z$, original headline $S_N$, and generated headline $\tilde{S_N}$
- Transfer discriminator $D_T$: discriminate original data samples with generated data samples




Original clickbaits and generated non-clickbaits

$$\mathcal{L}_{D_T} = \boxed{\mathcal{L}_{D_T^{(1)}}(\theta_{D_T^{(1)}})} + \boxed{\mathcal{L}_{D_T^{(2)}}(\theta_{D_T^{(2)}})}$$

Original non-clickbaits and generated clickbaits

- Style discriminator $D_S$: assign a correct label of styles for both original headlines and generated headlines

Original clickbaits and original non-clickbaits

$$\mathcal{L}_{D_S}(\mathbf{W}, \mathbf{b}) = \boxed{\mathcal{L}_{D_S}^{(1)}} + \boxed{\mathcal{L}_{D_S}^{(2)}}$$

Generated clickbaits and generated non clickbaits

# Discriminator Learning

- Pair discriminator $D_P$ ensures that the correspondences of documents and headlines are maintained

Proximity function $$p(h_i, x_j) = \frac{1}{1 + \exp(-\underbrace{\mathbf{s}^{(i)}}_{\text{Headline representation}}\mathbf{Q}\underbrace{\mathbf{z}^{(j)}}_{\text{Document representation}})}$$

- Maximizing the proximity of (document, headline) pairs with negative sampling

$$\mathcal{L}_{D_P} = -\log \sigma(\mathbf{s}^{(i)}\mathbf{Q}\mathbf{z}^{(i)}) - \sum_{k=1}^{K} \mathbb{E}_{x_k \sim P_n(x)}[\log \sigma(-\mathbf{s}^{(i)}\mathbf{Q}\mathbf{z}^{(k)})]$$

# Experiments Setting

TABLE I: The statistics and descriptions of the datasets

| Dataset | Source | # Clickbaits | # Non-clickbaits |
|---|---|---|---|
| $P$ | **P**rofessional Writers | 5,000 | 16,933 |
| $M$ | Social **M**edia Users | 4,883 | 16,150 |

- Datasets
  - Professional writers (P):
    Reporters or editors generate clickbaits for their news pieces
  - Social media users (M):
    Clickbaits to lure people to click their posts on social media.
- Baselines
  - SeqGAN [AAAI'17] : Text generation using GAN with reinforcement learning
  - SVAE [CONLL'16]: Sentence generation using Variational AutoEncoder (VAE)
  - CrossA [NIPS'17]: Generating sentences across different styles

# Experiments - Evaluation questions

- **Consistency**: are generated clickbaits/non-clickbaits consistent with the original datasets?
- **Readability**: are generated headlines readable or not?
- **Similarity**: are generated headlines semantically similar to original documents?

**Data Quality**

- **Differentiability**: are generated clickbaits/non-clickbaits differentiable?
- **Accuracy**: can generated clickbaits/non-clickbaits help improve the detection performance?

**Data Utility**

# Experimental Results - Data Quality

- **Similarity:** evaluate the semantic similarity of headlines and documents
  - Bilingual Evaluation Understudy (BLEU) score
  - Uni_sim: similarity of universal text embedding
- SHG achieves better performances to preserve document content than CrossA

TABLE V: **EQ3**: The Average BLEU (BLEU-4) Score Comparison of Generated Headlines. $\mathcal{H}$ indicates original headlines, and $\mathcal{O}$ represents the generated headlines.

| Data | Headlines | Methods | Clickbait | Non-Clickbait |
|---|---|---|---|---|
| $P$ | $\mathcal{H}$ | | 0.555 | 0.527 |
| | $\mathcal{O}$ | CrossA | 0.407 | 0.432 |
| | | SHG | **0.453** | **0.446** |
| $M$ | $\mathcal{H}$ | | 0.541 | 0.534 |
| | $\mathcal{O}$ | CrossA | 0.432 | 0.437 |
| | | SHG | **0.451** | **0.442** |

TABLE VI: **EQ3**: The Average Uni_sim Value Comparison of Generated Headlines. $\mathcal{H}$ indicates original headlines, and $\mathcal{O}$ represents the generated headlines.

| Data | Headlines | Methods | Clickbait | Non-Clickbait |
|---|---|---|---|---|
| $P$ | $\mathcal{H}$ | | 0.63 | 0.81 |
| | $\mathcal{O}$ | CrossA | 0.20 | 0.22 |
| | | SHG | **0.37** | **0.40** |
| $M$ | $\mathcal{H}$ | | 0.64 | 0.81 |
| | $\mathcal{O}$ | CrossA | 0.26 | 0.34 |
| | | SHG | **0.34** | **0.38** |

# Experimental Results - Data Utility

- **Accuracy:** improvement comparison of original headlines on AUC
  - The headlines generated by SVAE, CrossA, and SHG can increase the performance of clickbait detection to some extent
  - SHG consistently outperforms SVAE and CrossA

| Data | Classifier | Org | SeqGAN | SVAE | CrossA | SHG |
|---|---|---|---|---|---|---|
| $P$ | LogReg | 0.928 | 0.900 (↓ 3.02%) | 0.933 (↑ 0.54%) | 0.932 (↑ 0.64%) | **0.936 (↑ 0.86%)** |
| | DTree | 0.894 | 0.882 (↓ 1.34%) | 0.908 (↑ 1.57%) | 0.900 (↑ 0.67%) | **0.910 (↑ 1.79%)** |
| | RForest | 0.900 | 0.893 (↓ 0.78%) | 0.912 (↑ 1.33%) | 0.916 (↑ 1.78%) | **0.925 (↑ 2.78%)** |
| | XGBoost | 0.919 | 0.914 (↓ 0.54%) | 0.923 (↑ 0.43%) | 0.926 (↑ 0.76%) | **0.928 (↑ 0.98%)** |
| | AdaBoost | 0.917 | 0.896 (↓ 2.29%) | 0.921 (↑ 0.44%) | 0.921 (↑ 0.44%) | **0.931 (↑ 1.64%)** |
| | SVM | 0.904 | 0.898 (↓ 0.66%) | 0.917 (↑ 1.44%) | 0.920 (↑ 1.77%) | **0.923 (↑ 2.10%)** |
| | GradBoost | 0.921 | 0.914 (↓ 0.76%) | 0.924 (↑ 0.33%) | 0.926 (↑ 0.54%) | **0.928 (↑ 0.76%)** |

# Summary

- We study the problem of generating clickbaits/nonclickbaits from original documents for clickbait detection
- We propose a novel deep generative model with adversarial learning

- **Future work**
  - Explore the generalization capacity of SHG on other styles such as positive-negative sentiment style and academic-news reporting style
  - Investigate the strategy of learning the disentangled representations of content and style

# FakeNewsTracker: A Tool for Fake News Collection, Detection, and Visualization

**Kai Shu, Deepak Mahudeswaran, and Huan Liu**


**SBP 2018**

SBP Disinformation Challenge Winner

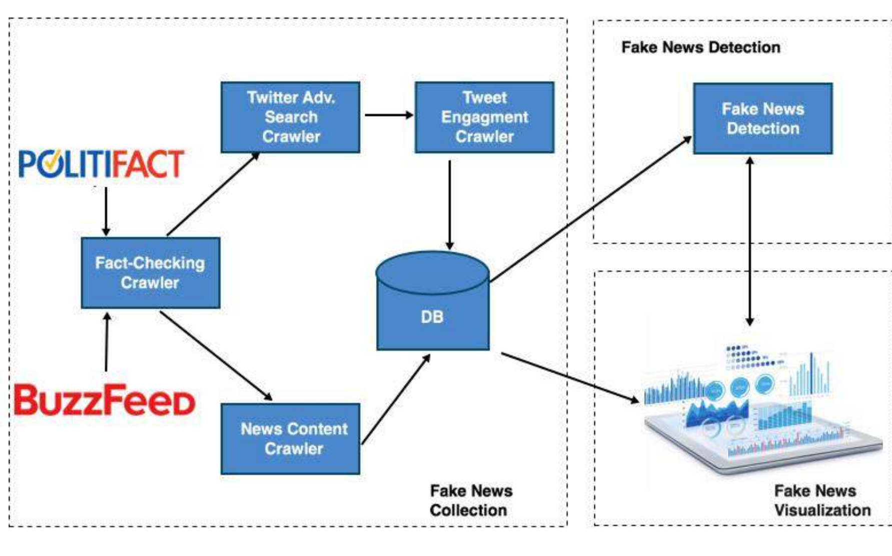
http://blogtrackers.fulton.asu.edu:3000

# An end-to-end framework for fake news collection, detection, and visualization

- **Data Collection:** collecting fake and real news articles from fact-checking websites and related social engagements from social media

- **Fake News Detection**: finding fake news with advanced machine learning methods, such as deep neural networks

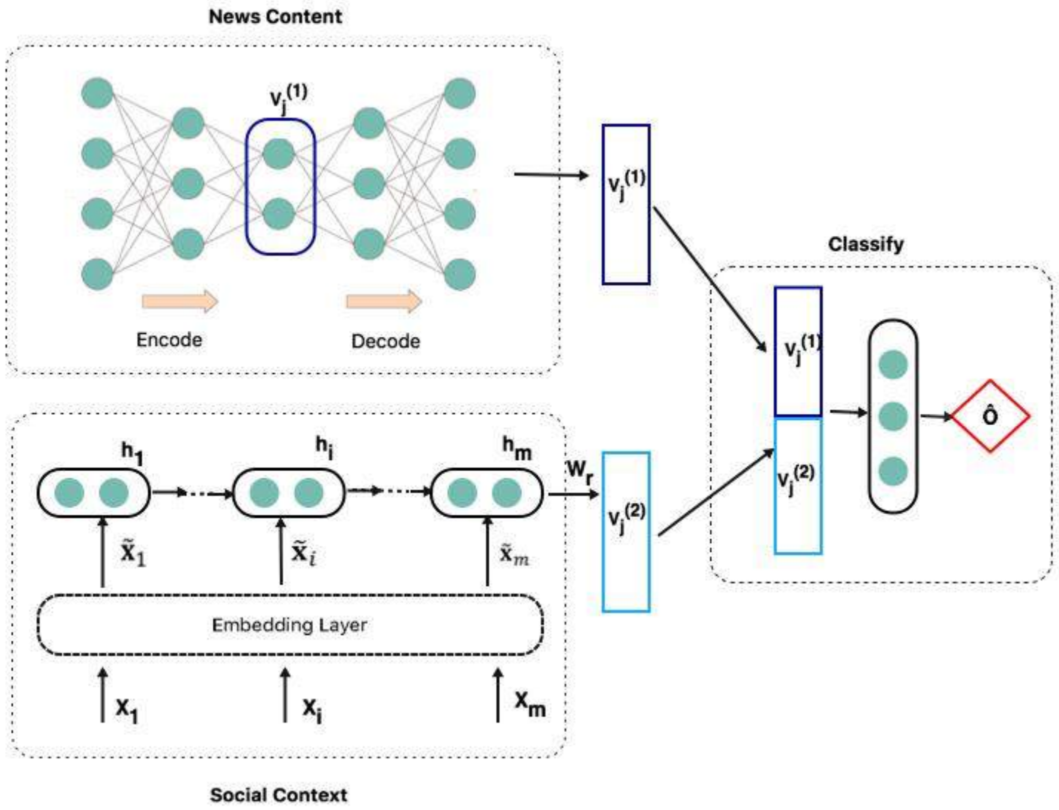
- **Fake News Visualization:** visualization on data attributes and model performance




X. Zhou, R. Zafarani, K. Shu, H. Liu

# Fake News Detection

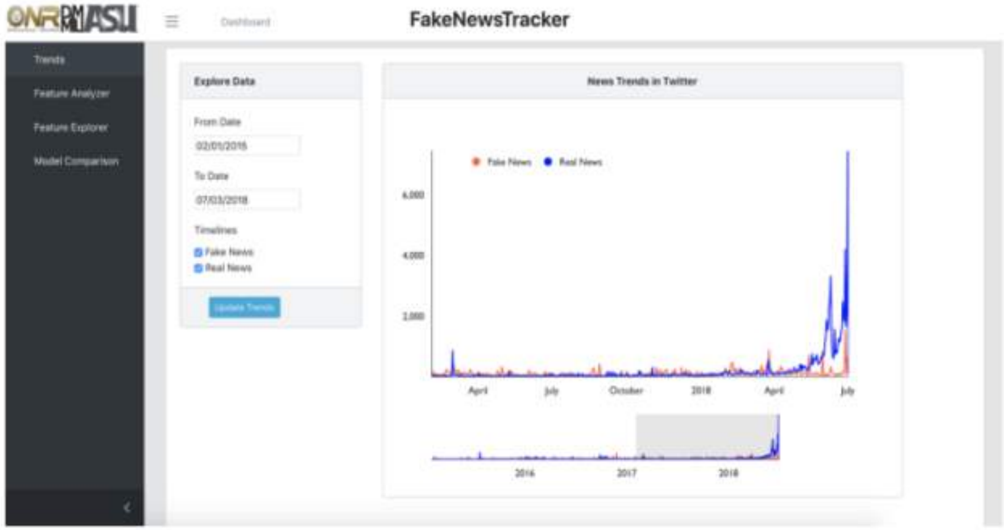
- Detect fake news with fusion of news content and social context
  - **News representation**: Represent news content using autoencoders
  - **Social engagement representation:** Represent social engagements using RNNs
  - **Social Article Fusion:** Combine both news and social engagement features to detect fake news




X. Zhou, R. Zafarani, K. Shu, H. Liu

# Fake News Visualization

## Trends on Twitter



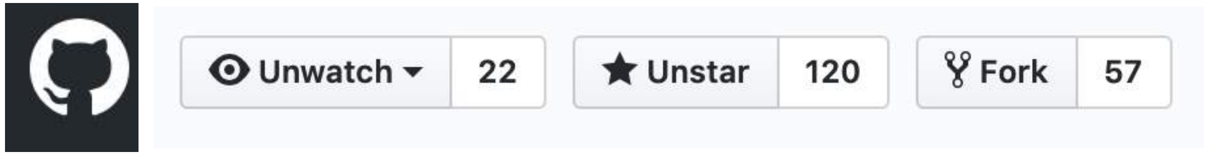

## Topics of Fake news vs Real News

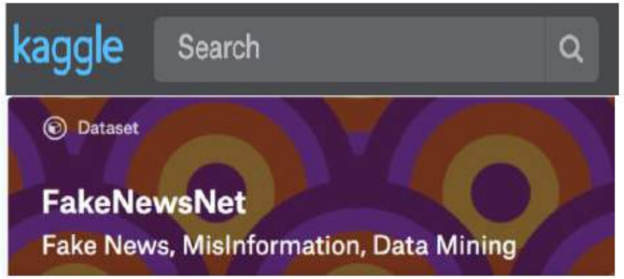

## Geolocation of Fake News vs Real News


## Social Network on Users Spreading Fake/Real news

# FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media

**Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, Huan Liu**




https://github.com/KaiDMML/FakeNewsNet https://www.kaggle.com/mdepak/fakenewsnet

# How unique is FakeNewsNet?

- A comprehensive data repository that contains news contents, social context, and spatiotemporal information

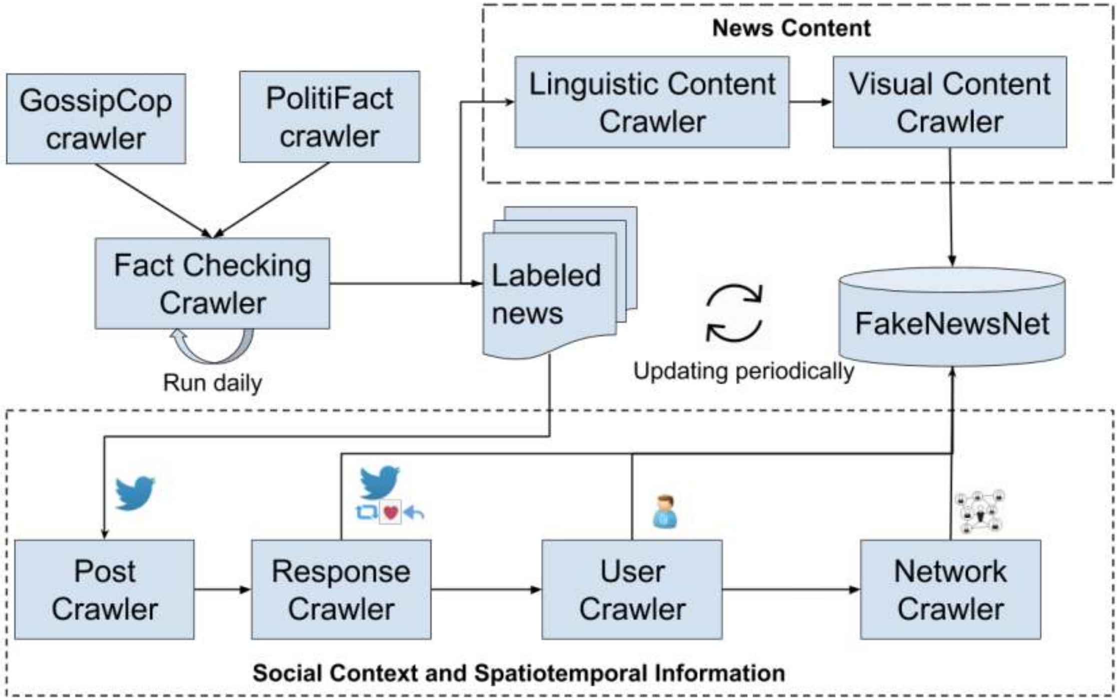
Table 1: Comparison with existing fake news detection datasets

| Features / Dataset | News Content | | Social Context | | | | Spatiotemporal Information | |
|---|---|---|---|---|---|---|---|---|
| | *Linguistic* | *Visual* | *User* | *Post* | *Response* | *Network* | *Spatial* | *Temporal* |
| BuzzFeedNews | ✓ | | | | | | | |
| LIAR | ✓ | | | | | | | |
| BS Detector | ✓ | | | | | | | |
| CREDBANK | ✓ | | ✓ | ✓ | | | ✓ | ✓ |
| BuzzFace | ✓ | | | ✓ | ✓ | | | ✓ |
| FacebookHoax | ✓ | | ✓ | ✓ | ✓ | | | |
| **FakeNewsNet** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

X. Zhou, R. Zafarani, K. Shu, H. Liu

# Data Integration

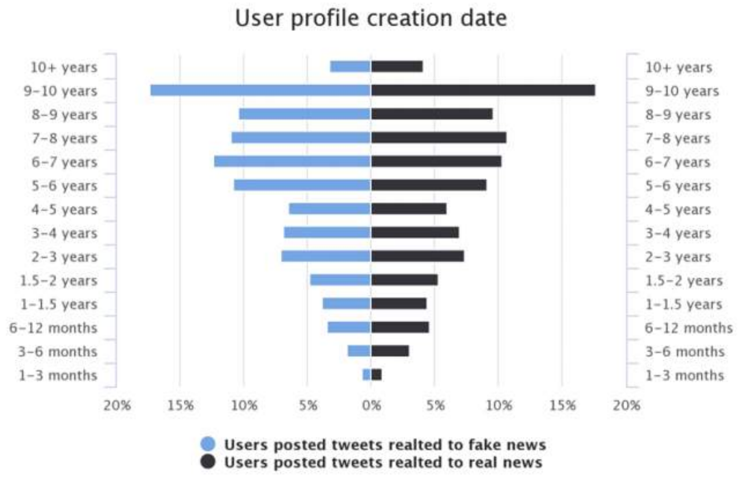
- **News Content**: we utilize fact-checking websites to obtain news contents for fake news and true news
- **Social Context**: collecting user engagements from Twitter using the headlines of news articles
- **Spatiotemporal Information:** spatial information and temporal data from meta data of Twitter

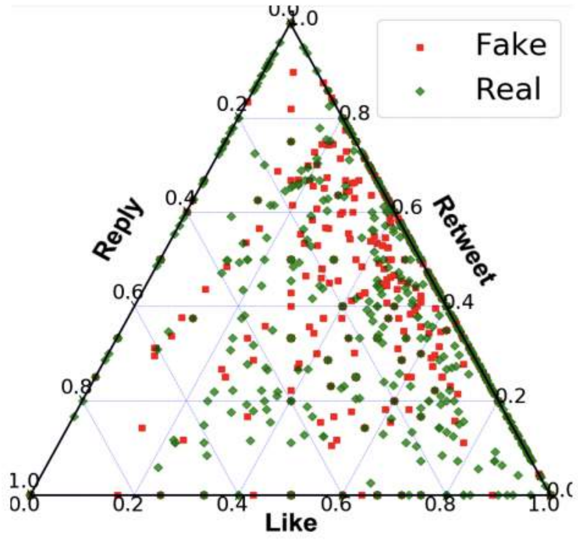


# Data Analysis

- **User profiles**: users who share real news pieces tend to have longer register time than those who share the fake news on average

- **User engagements**: fake news pieces tend to have fewer replies and more retweets; real news pieces have more ratio of likes than fake news pieces do

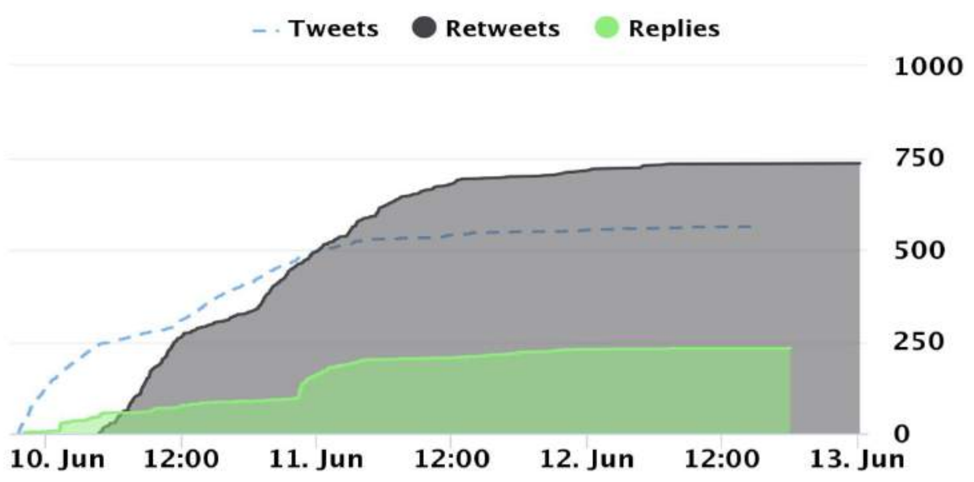


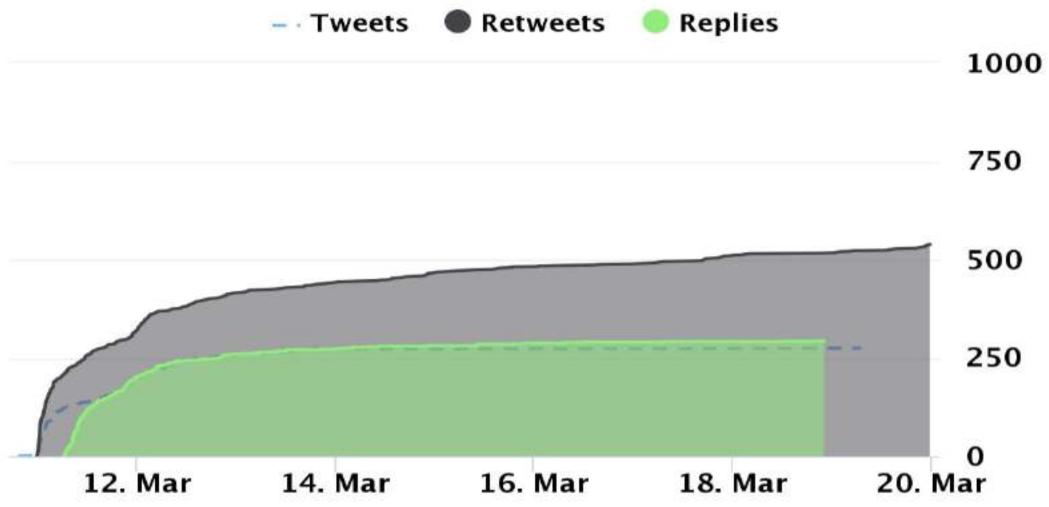



X. Zhou, R. Zafarani, K. Shu, H. Liu

- **A case study of temporal engagements for fake news and real news**
  - For fake news, a sudden increase in the number of retweets and remain constant beyond a short time
  - For real news, the number of retweets increases steadily
  - Fake news pieces tend to receive fewer replies than real news




Fake News




Real News

# Potential Applications for FakeNewsNet

- **Fake News Detection**
  - News content, social context based
  - Early fake news detection
- **Fake News Evolution**
  - Temporal, Topic, Network, evolution
- **Fake News Mitigation**
  - Provenances, persuaders, clarifiers
  - Influence minimization, mitigation campaign
- **Malicious Account Detection**
  - Detecting bots that spread fake news

# dEFEND: Explainable Fake News Detection

**Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu**

**KDD 2019**

# Explainable Fake News Detection

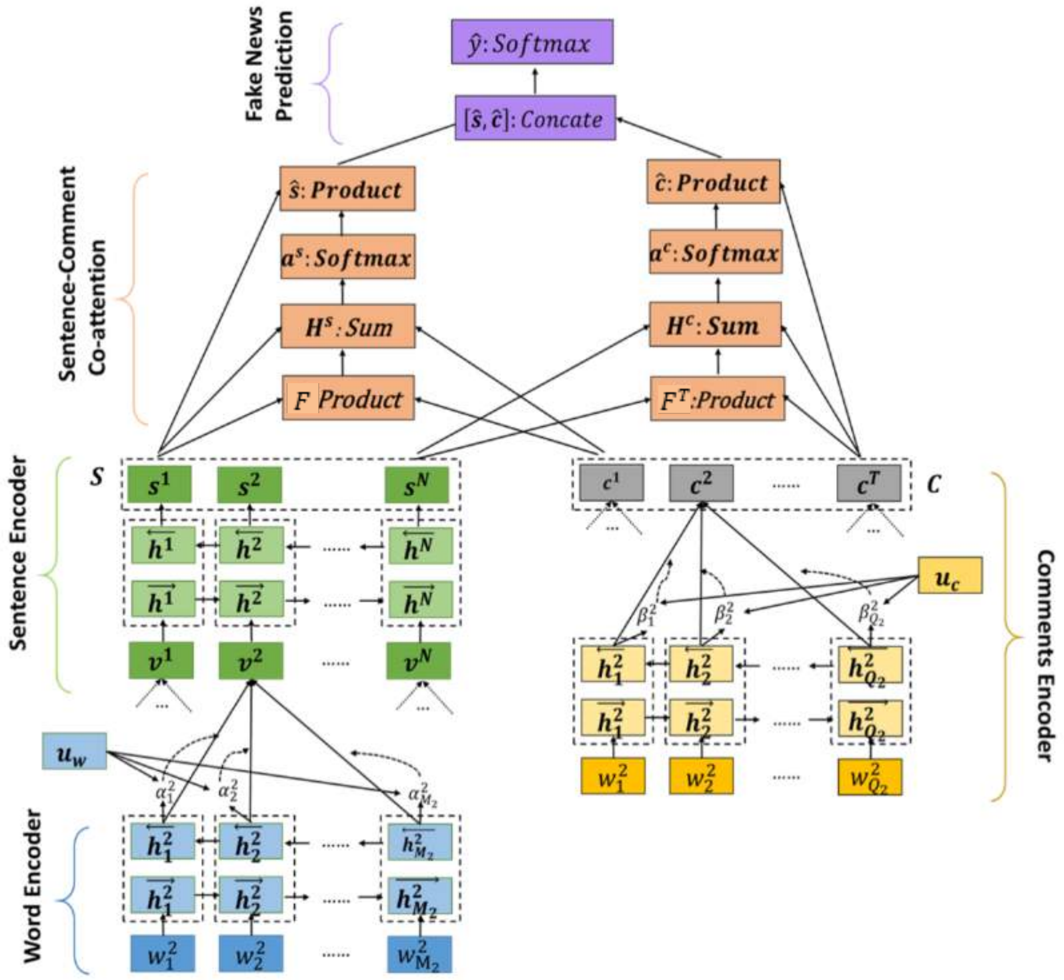
- Existing work focuses on ***detecting*** fake news, but cannot ***explain why*** it is detected as fake
- Explanation is important
  - Provide insights and knowledge to practitioners
  - Extracting explainable features can further improve the fake news detection performance

# Contents, Comments, and Their Relations

- News contents and user comments are inherently **related**
  - News contents contain false information
  - User comments have rich information from the crowd such as  opinions, stances, and sentiment

# dEFEND can explain why it is fake

- A hierarchical attention network to capture world-level and sentence-level structure
- An attention-based bidirectional GRU network to model word sequences in comments
- A co-attention network to model the relationship between contents and comments




X. Zhou, R. Zafarani, K. Shu, H. Liu

# Evaluation Setting

- Datasets: FakeNewsNet with information for news contents, user comments and ground truth labels from fact-checking websites

| Platform | PolitiFact | GossipCop |
|---|---|---|
| # Users | 68,523 | 156,467 |
| # Comments | 89,999 | 231,269 |
| # Candidate news | 415 | 5,816 |
| # True news | 145 | 3,586 |
| # Fake news | 270 | 2,230 |

- Compared baselines:
  - RST: rhetorical relations among the words in the text
  - LIWC: lexicons falling into psycholinguistic categories
  - HANL hierarchical attention networks
  - textCNN: features with convolutional neural network
  - HPA-BLSTM: temporal modeling of comments with attention network
  - CSI: deep network modeling news, source and comments
  - TCNN-URG: CNN for news and conditional VAE for comments

News Content

User Comments

News Content
+ User Comments

# Evaluation Results - Detection Performance

- User comment based methods are more effective than news content based methods
- dEFEND performs the best than other methods using both news content and user comments

User Comments

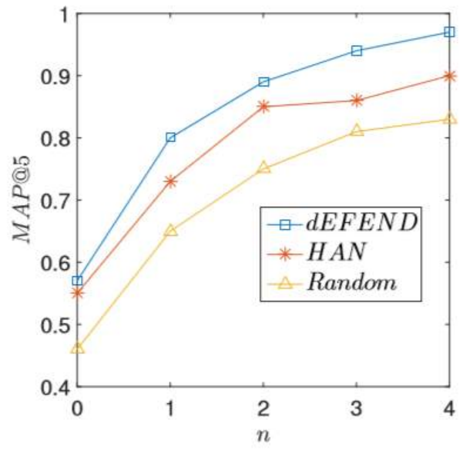
| Datasets | Metric | RST | LIWC | text-CNN | HAN | TCNN-URG | HPA-BLSTM | CSI | dEFEND |
|---|---|---|---|---|---|---|---|---|---|
| **PolitiFact** | Accuracy | 0.607 | 0.769 | 0.653 | 0.837 | 0.712 | 0.846 | 0.827 | **0.904** |
| | Precision | 0.625 | 0.843 | 0.678 | 0.824 | 0.711 | 0.894 | 0.847 | **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

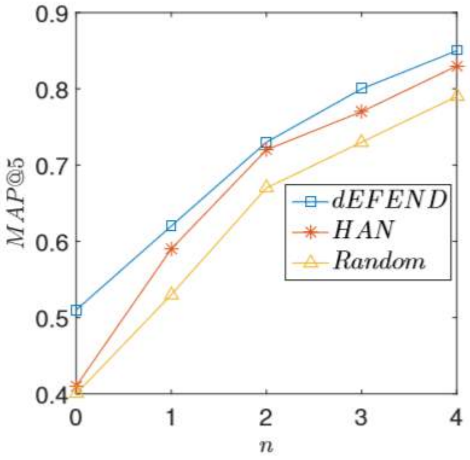
News Content + Social Context

# Evaluation Results - Explainability on news sentences

- News sentence explainability: the degree of check-worthy
- Ground truth: obtained with ClaimBuster[1]
- dEFEND can achieve better performance to capture more check-worthy sentences than HAN and random
- With the increase of window size n, the MAP performances increase




(a) MAP@5 on PolitiFact



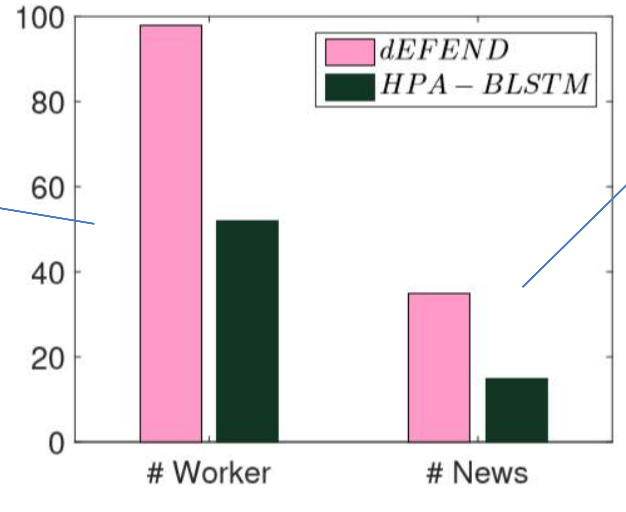

(c) MAP@5 on GossipCop

[1] Hassan, Naeemul, et al. "Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster." KDD 2017.

# Evaluation Results - Explainability on user comments

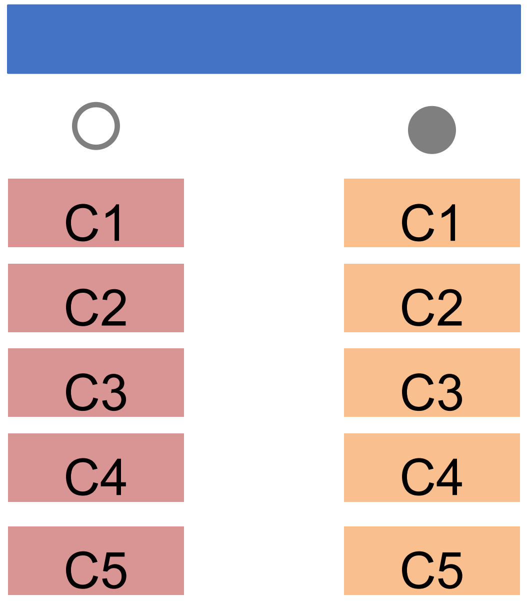
- HPA-BSLTM, attention modeling on temporal structure of comments
- Using Amazon Mechanical Turk to perform human evaluation tasks
- **Task 1**: selecting top-k ranking list **collectively** better between HPA-BSLTM and dEFEND

Worker-level:
WR 0.65




(a) Winning Count

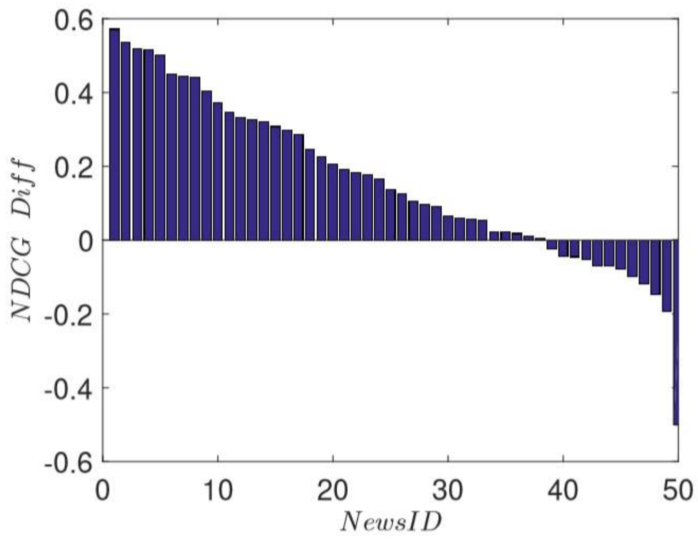
News-level:
WR 0.64

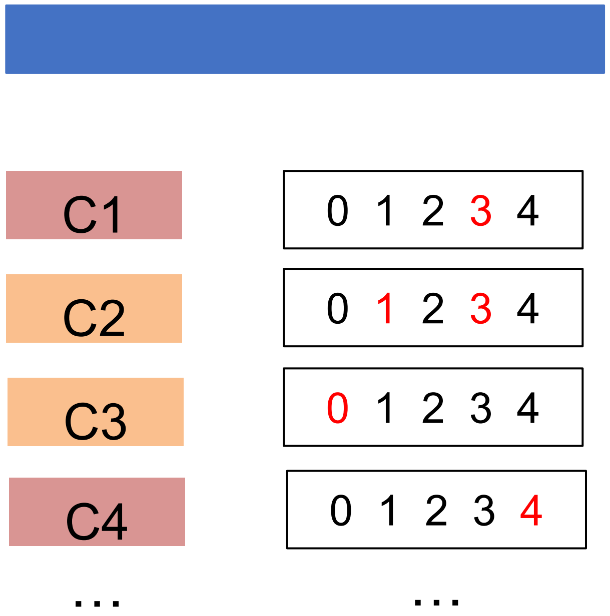



X. Zhou, R. Zafarani, K. Shu, H. Liu

# Evaluation Results - Explainability on user comments

- **Task 2:** assigning scores for each comments in a mixed list from HPA-BSLTM and dEFEND
- NDCG Diff = NDCG (dEFEND)-NDCG (HPA-BLSTM)



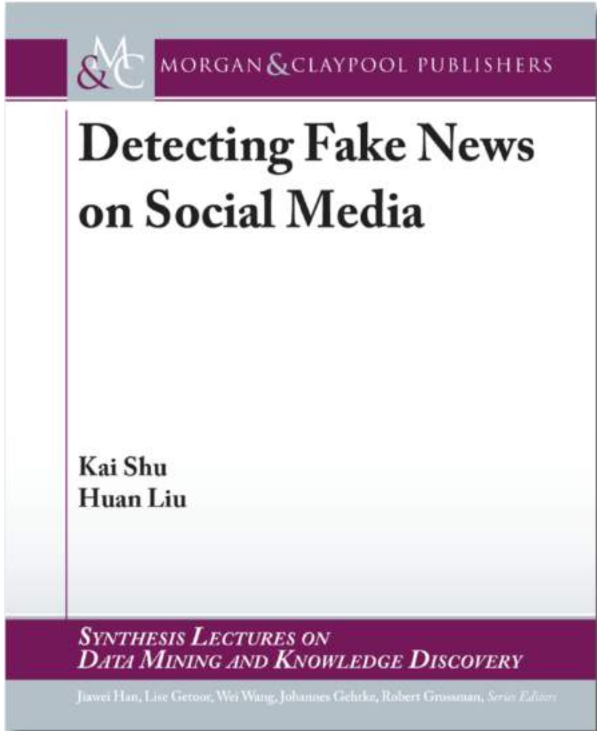



# Summary

- A new framework for the novel problem of explainable fake news detection
- Achieve higher accuracy than the state-of-the-art fake news detection methods
- Discover explainable news sentences and user comments to understand why news pieces are identified as fake

# Recent work at DMML on Fake News Detection

- Book: Detecting Fake News on Social Media
- Edited book: Misinformation, disinformation, and fake news. [**CFP**]: http://www.public.asu.edu/~skai2/fndm.html
- Survey: Fake News Detection on Social Media: A Data Mining Perspective
- Data repository: FakeNewsNet, [Github], [Kaggle], [Paper]
- Software: FakeNewsTracker
- Book chapter: Studying Fake News via Network Analysis: Detection and Mitigation
- Other Publications: related publications are updated at: http://www.public.asu.edu/~skai2/




http://dmml.asu.edu/dfn/

# Challenges and Highlights

- Fake News Early Detection
- Identify Check-worthy Content
- Cross-domain, -topic, -language Fake News Studies
- Weakly-supervised Fake News Detection

# Fake News Early Detection

*Why is Fake News Early Detection is important?*

- The more fake news spreads, the more likely for people to trust it
- Once people have trusted the fake news, it is difficult to correct users' perceptions

| | Term | Phenomenon |
|---|---|---|
| **Social influence** | *Attentional bias* | **Exposure frequency** - individuals tend to believe information is correct after repeated exposures. |
| | *Validity effect* | |
| | *Echo chamber effect* | |
| | *Bandwagon effect* | **Peer pressure** - individuals do something primarily because others are doing it and to conform to be liked and accepted by others. |
| | *Normative influence theory* | |
| | *Social identity theory* | |
| | *Availability cascade* | |

| Term | Phenomenon |
|---|---|
| *Backfire effect* | Given evidence against their beliefs, individuals can reject it even more strongly |
| *Conservatism bias* | The tendency to revise one's belief insufficiently when presented with new evidence. |
| *Semmelweis reflex* | Individuals tend to reject new evidence as it contradicts with established norms and beliefs. |

# Fake News Early Detection

*How to achieve Fake News Early Detection?*

I. **Verification Efficiency**, e.g., compare knowledge in the framework that
   - Knowledge graphs with timely ground truth
   - To-be-verified news content is check-worthy – *Check-worthy content identification*

II. **Feature Compatibility**, e.g., to extract features that can capture
   - The generality of deceptive content styles *across* domain, topic, and language
   - The evolution of deceptive content styles *within* domain, topic, and language

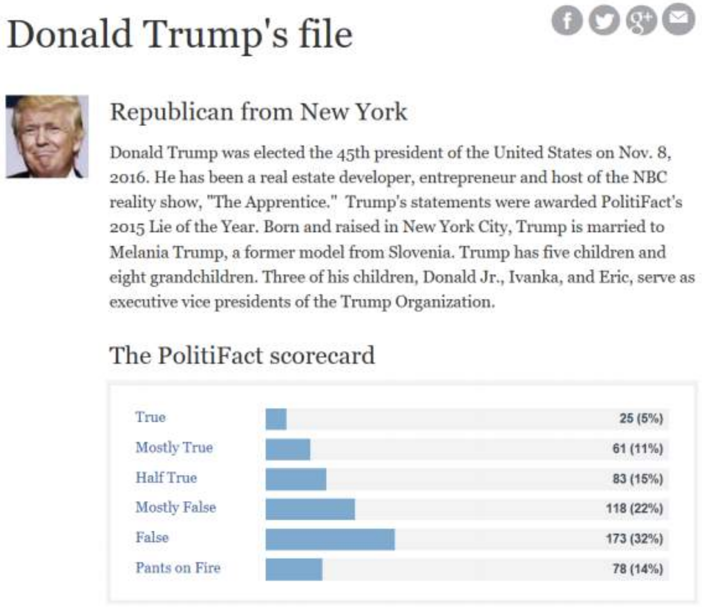
III. **Information Availability**, e.g., detect fake news with limited propagation information

# Check-worthy Content Identification

*How to measure Check-worthy Content?*

I. **News-worthiness or Potential Influence on the Society,** e.g., if it is related to national affairs

II. **Spammer Preference,**
i.e., news historical likelihood of being fake




(a) (Expert-based) PolitiFact: the PolitiFact scorecard

*Related Studies*:

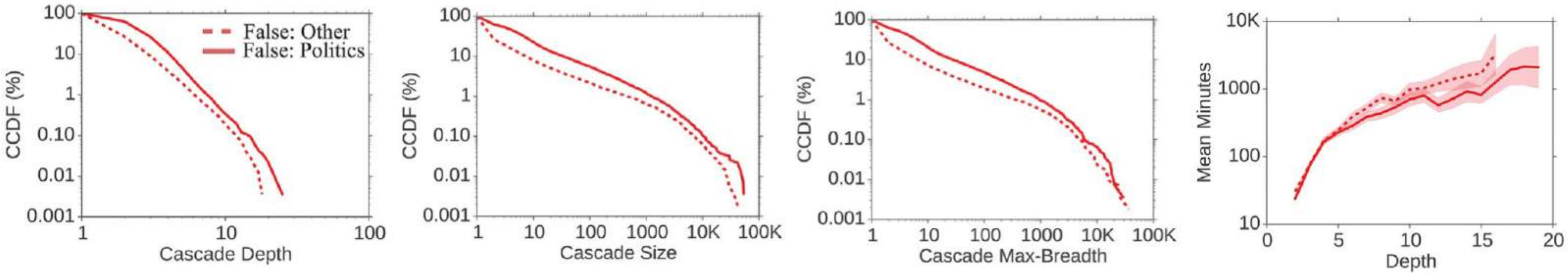
- N. Hassan, et al. Detecting Check-worthy Factual Claims in Presidential Debates, CIKM'15
- N. Hassan et al., Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster, KDD'17

# Cross-domain, -topic, -language

*How to facilitate Cross-domain, -topic, -language Fake News Studies?*

I. Develop **fake news datasets** containing cross-domain, -topic, -language data

II. Explore **patterns** among fake news within different domains, topics and languages




III. Develop **techniques** enables cross-domain, -topic, -language fake news detection

# Weakly-supervised Fake News Detection

- Annotating fake news is usually time-consuming and labor-intensive
- How to build semi-supervised, unsupervised models?
- How to learn weak supervision from rich social context information?