Fake News: Fundamental Theories, Detection Strategies and Challenges

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

Tutorial | 12th ACM International WSDM Conference

Syracuse University

Arizona State University

Meet our Team

Xinyi Zhou

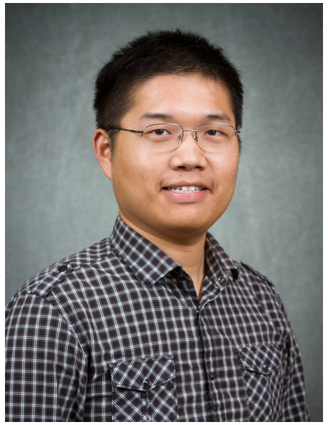
Syracuse University

Ph.D. Student
Data Lab,
EECS Department

Reza Zafarani

Syracuse University

Assistant Professor
Data Lab,
EECS Department



Kai Shu

Arizona State University

Ph.D. Student
Computer science
and Engineering

Huan Liu

Arizona State University

Professor
Computer Science
and Engineering

Introduction

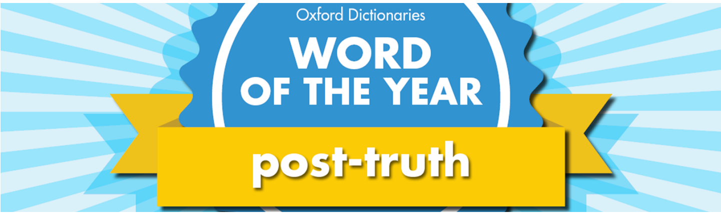
- Research Background
- What is Fake News?
- Related Concepts
- Fundamental Theories

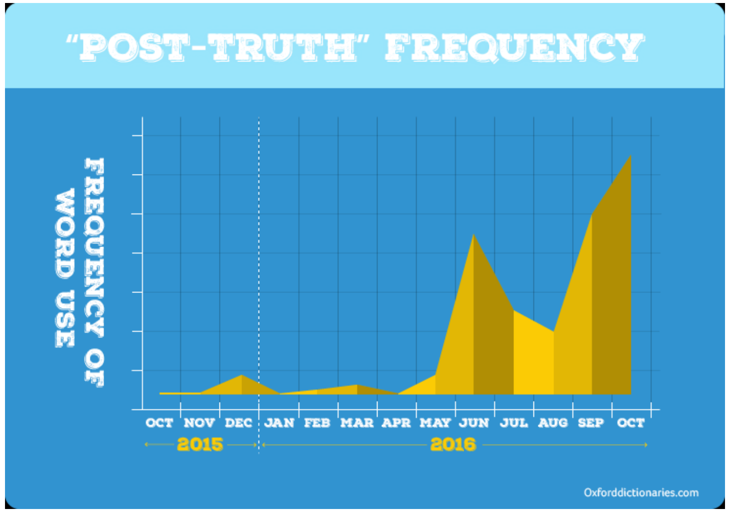
Research Background

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."



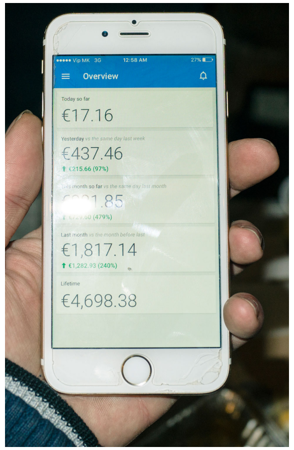


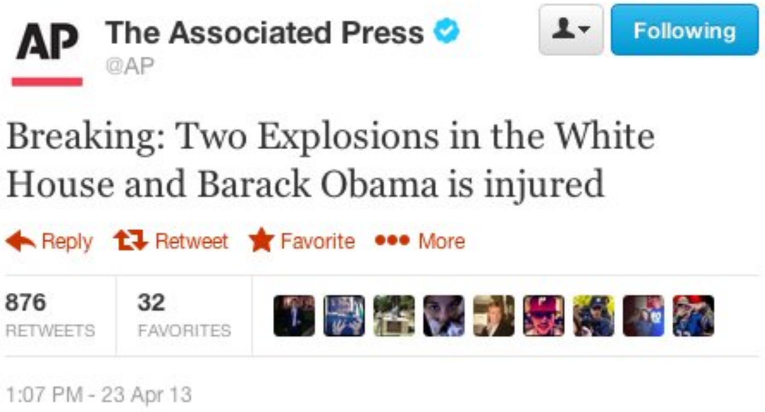
[1] C. Silverman. This analysis shows how viral fake election news stories outperformed real news on Facebook. BuzzFeed News, 2016.

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

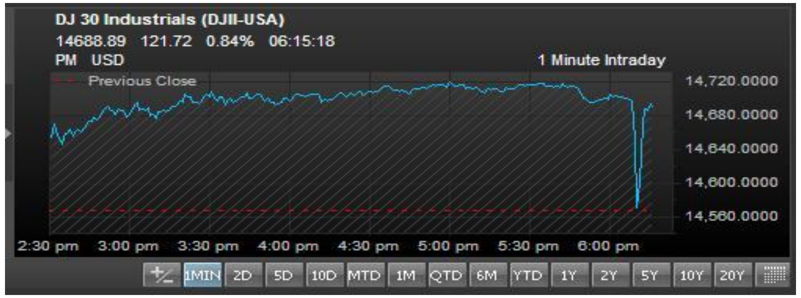



AP The Associated Press
@AP
Following

Breaking: Two Explosions in the White House and Barack Obama is injured

876 RETWEETS
32 FAVORITES

1:07 PM - 23 Apr 13



[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**

CHAINSAWSUIT.COM
i've heard the rhetoric from both sides... time to do my own research on the real truth
Googie
hotly debated topic
Found 80,000 results.
Literally the first link that agrees with what you already believe
Completely supports your viewpoint without challenging it in any way
Another link
Don't worry about this one
...jackpot

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 fake news *dissemination.*
  - 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 fake news *evolution.*
  - **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

ECHO CHAMBER
David H
Jonny opened the door to the one place he always heard the truth.

[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.

Fake News & Related Concepts

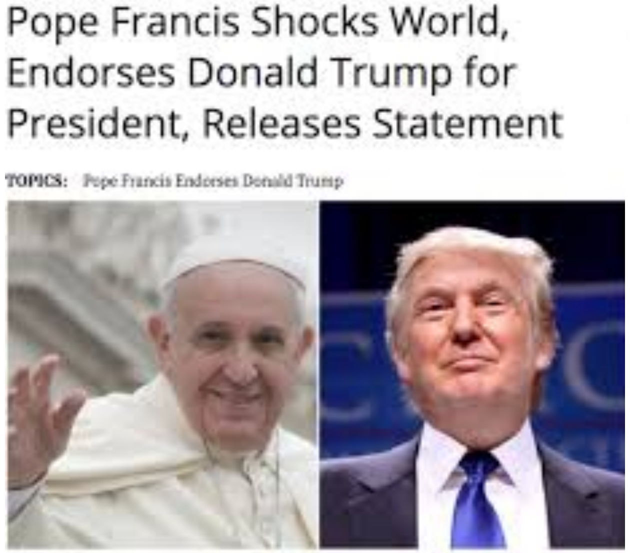
Definition of fake news

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

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

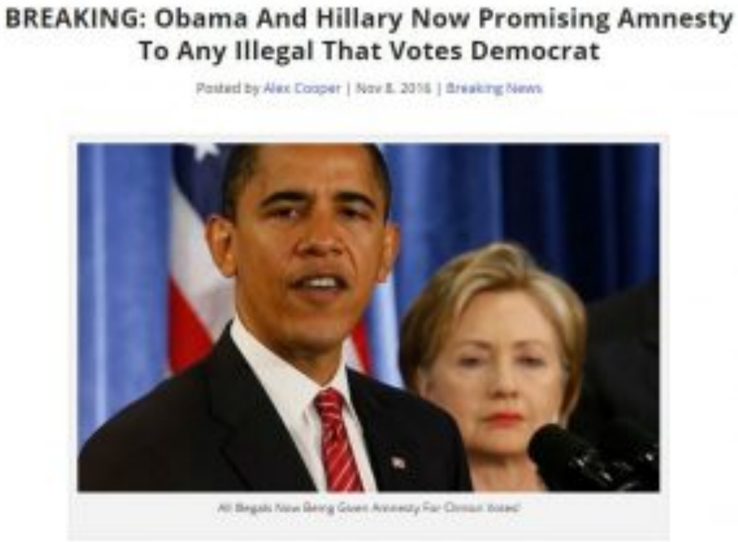
A more broad definition:

- *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

Posted by Alex Cooper | Nov 8, 2016 | Breaking News

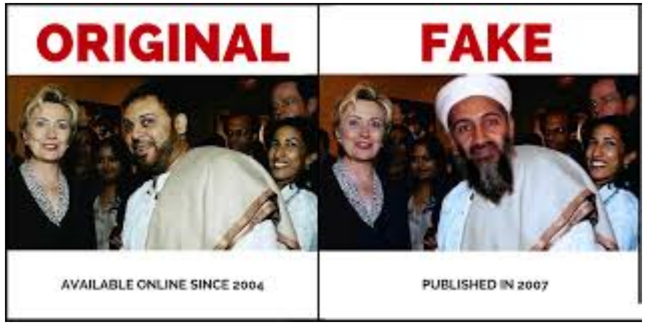
All Illegals Now Being Given Amnesty For Clinton Votes!

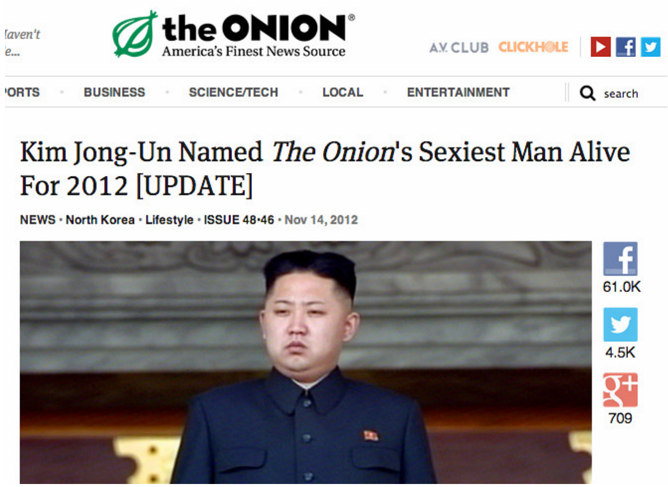
Fake News & Related Concepts

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.





Kim Jong-Un Named *The Onion*'s Sexiest Man Alive For 2012 [UPDATE]

NEWS · North Korea · Lifestyle · ISSUE 48•46 · Nov 14, 2012

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

Why is it necessary to study Fundamental Theories?

Fundamental human cognition and behavior theories developed across various discipline such as psychology, philosophy, social science, and economics provide invaluab insights for fake news studies.

1. Pro e opportunities for **qualitative and quantitative studies** of big fake news data;
2. Su rt to build **well-justified and explaina le models** for fake news detection and ention; and

elop dat *l tru* ch

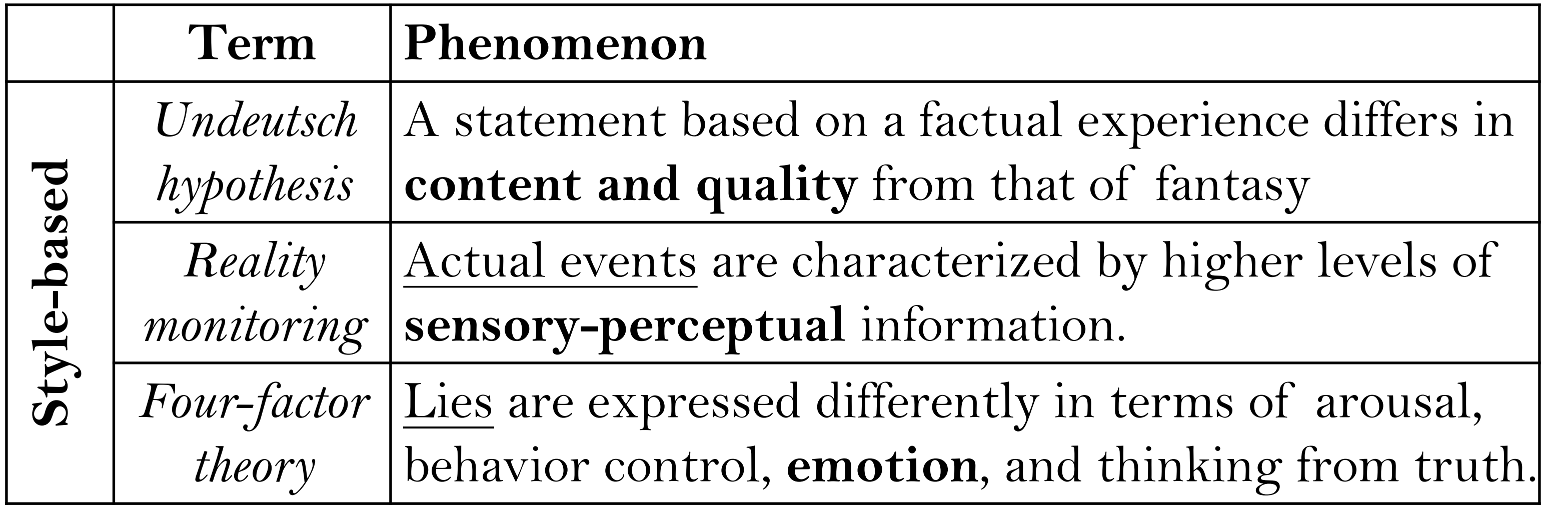
[Udo] Undeutsch hypothesis:
A **statement** based on a factual experience differs in **content and quality** from that of fantasy.

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**?

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.

		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	
		Contrast effect	

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

User-based Fundamental Theories

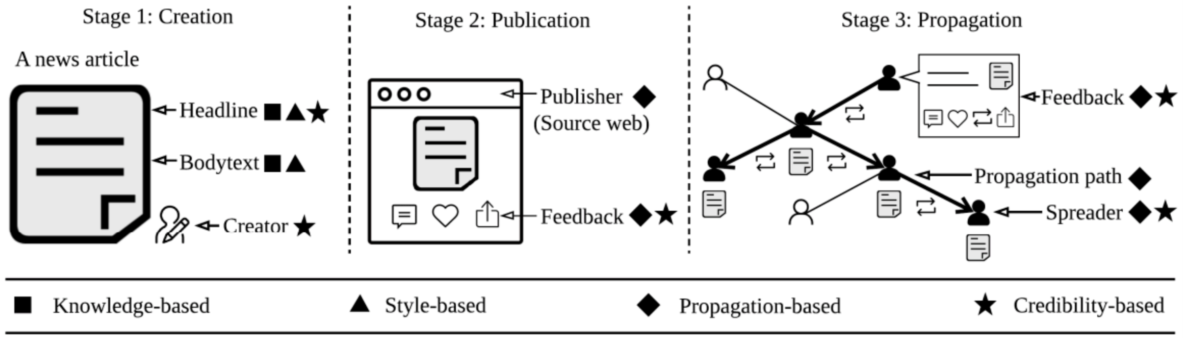
Studying fake news from a perspective of users:
How users engage with fake news and the role users play (or can play) in fake news creation, propagation, or intervention

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



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

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.

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

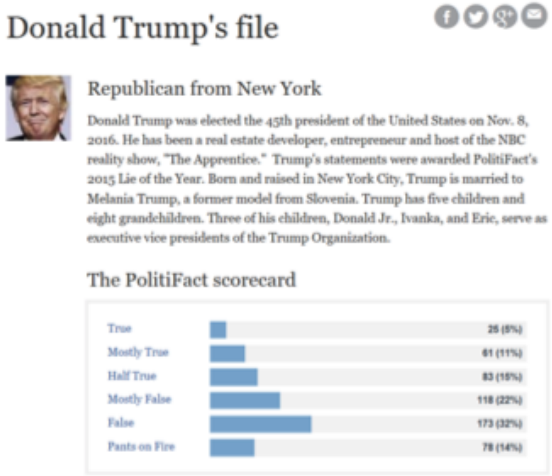
Multilabel classification

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


Donald Trump's file

Republican from New York

Donald Trump was elected the 45th president of the United States on Nov. 8, 2016. He has been a real estate developer, entrepreneur and host of the NBC reality show, "The Apprentice." Trump's statements were awarded PolitiFact's 2015 Lie of the Year. Born and raised in New York City, Trump is married to Melania Trump, a former model from Slovenia. Trump has five children and eight grandchildren. Three of his children, Donald Jr., Ivanka, and Eric, serve as executive vice presidents of the Trump Organization.

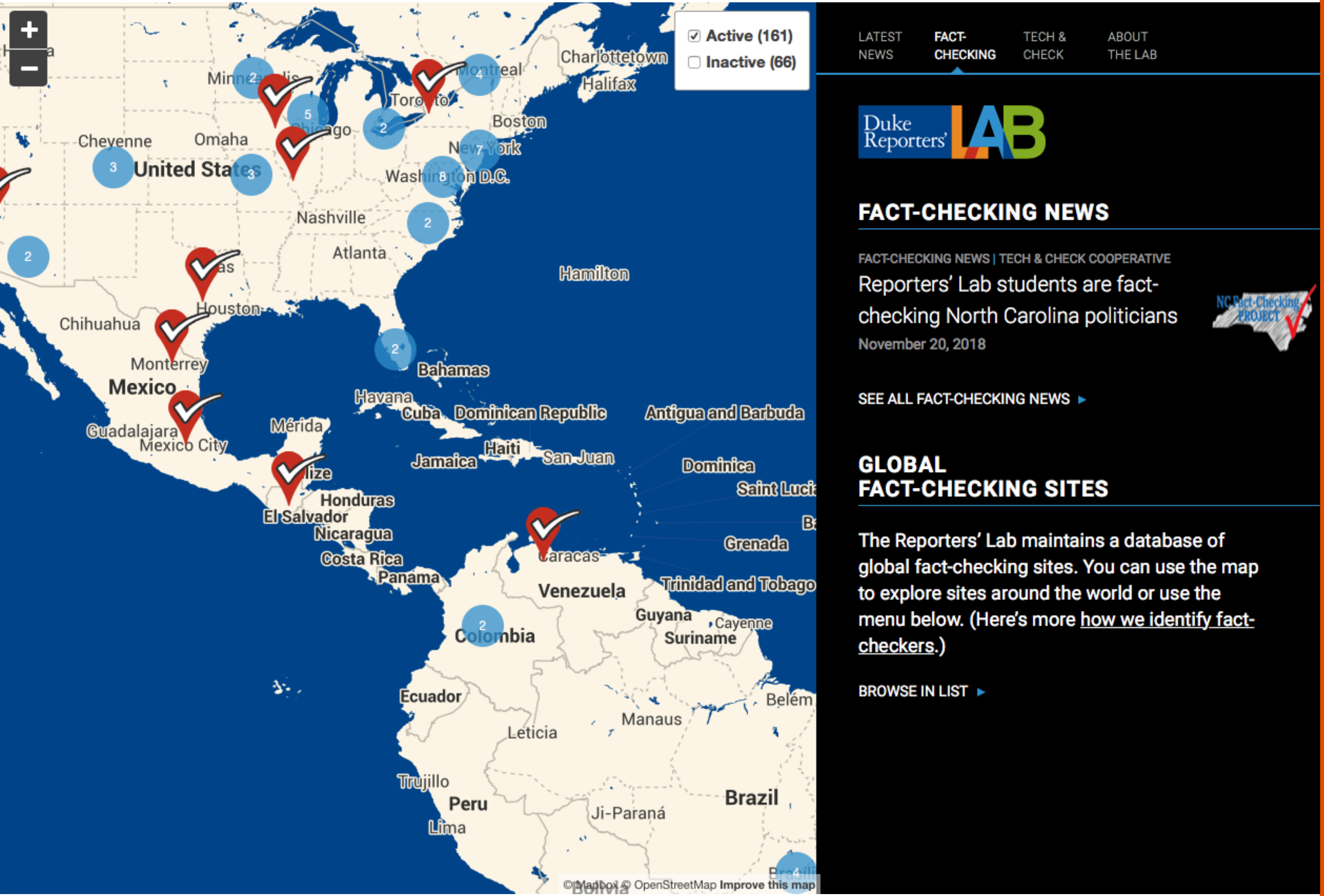
The PolitiFact scorecard

True	25 (5%)
Mostly True	61 (11%)
Half True	83 (15%)
Mostly False	118 (22%)
False	173 (32%)
Pants on Fire	78 (14%)

Expert-based Manual Fact-checking

Current resources

Reporters Lab – Duke University



☑ Active (161)

□ Inactive (66)

FACT-CHECKING NEWS

FACT-CHECKING NEWS | TECH & CHECK COOPERATIVE

Reporters' Lab students are fact-checking North Carolina politicians

November 20, 2018

SEE ALL FACT-CHECKING NEWS ▸

GLOBAL FACT-CHECKING SITES

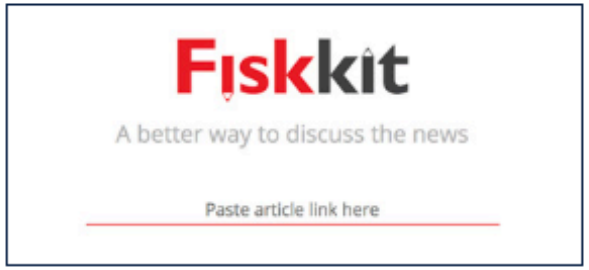
The Reporters' Lab maintains a database of global fact-checking sites. You can use the map to explore sites around the world or use the menu below. (Here's more how we identify fact-checkers.)

BROWSE IN LIST ▸

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

Crowd-sourced Manual Fact-checking

Current resources



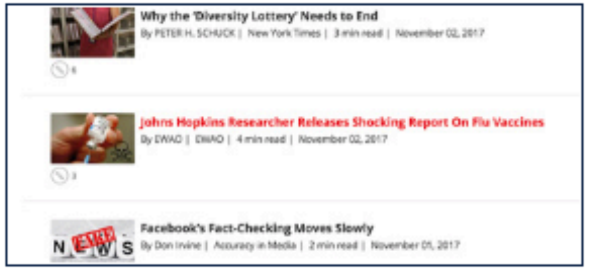
Fiskkit

A better way to discuss the news

Paste article link here

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

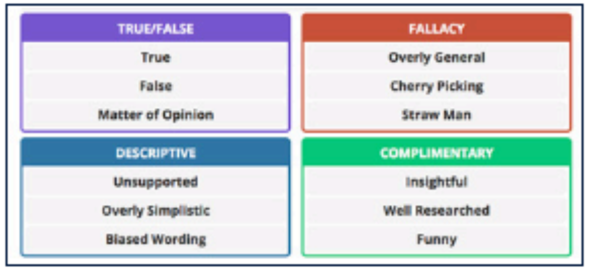


Why the 'Diversity Lottery' Needs to End
By PETER H. SCHUCK | New York Times | 3 min read | November 02, 2017
6

Johns Hopkins Researcher Releases Shocking Report On Flu Vaccines
By EWAO | EWAO | 4 min read | November 02, 2017
3

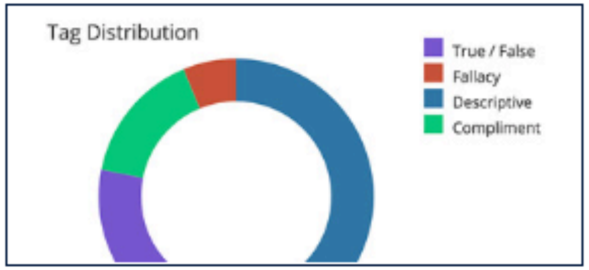
Facebook's Fact-Checking Moves Slowly
By Don Irvine | Accuracy in Media | 2 min read | November 01, 2017

Click on an article you find interesting.



TRUE/FALSE	FALLACY
True	Overly General
False	Cherry Picking
Matter of Opinion	Straw Man
DESCRIPTIVE	**COMPLIMENTARY**
Unsupported	Insightful
Overly Simplistic	Well Researched
Biased Wording	Funny

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

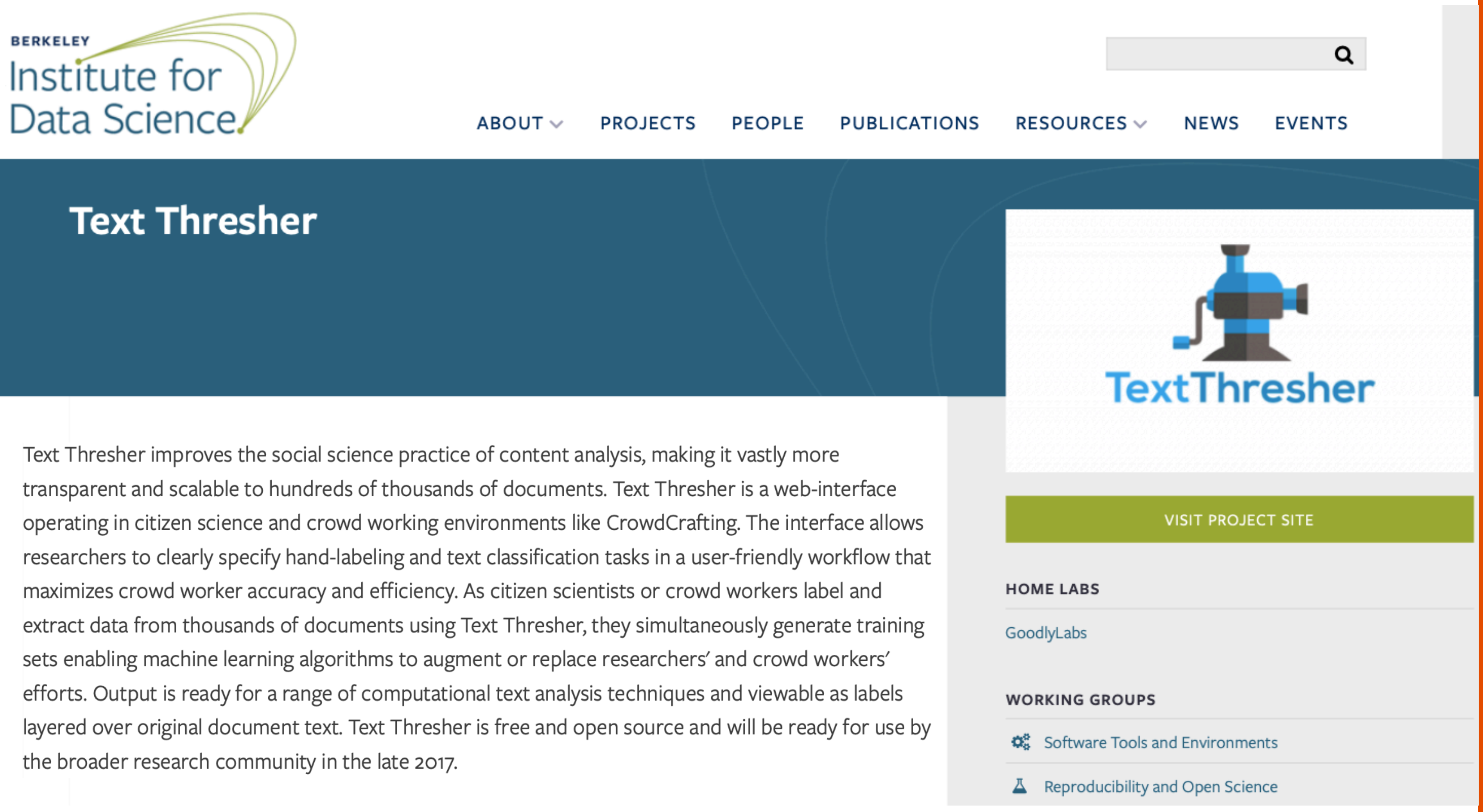


3 See how the article has been rated by other people through our insights page. Share the article so that your friends can come comment too.

http://www.fiskkit.com/

Crowd-sourced Manual Fact-checking

Current resources



Text Thresher

Text Thresher improves the social science practice of content analysis, making it vastly more transparent and scalable to hundreds of thousands of documents. Text Thresher is a web-interface operating in citizen science and crowd working environments like CrowdCrafting. The interface allows researchers to clearly specify hand-labeling and text classification tasks in a user-friendly workflow that maximizes crowd worker accuracy and efficiency. As citizen scientists or crowd workers label and extract data from thousands of documents using Text Thresher, they simultaneously generate training sets enabling machine learning algorithms to augment or replace researchers' and crowd workers' efforts. Output is ready for a range of computational text analysis techniques and viewable as labels layered over original document text. Text Thresher is free and open source and will be ready for use by the broader research community in the late 2017.

HOME LABS

GoodlyLabs

WORKING GROUPS

- Software Tools and Environments
- Reproducibility and Open Science

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**.

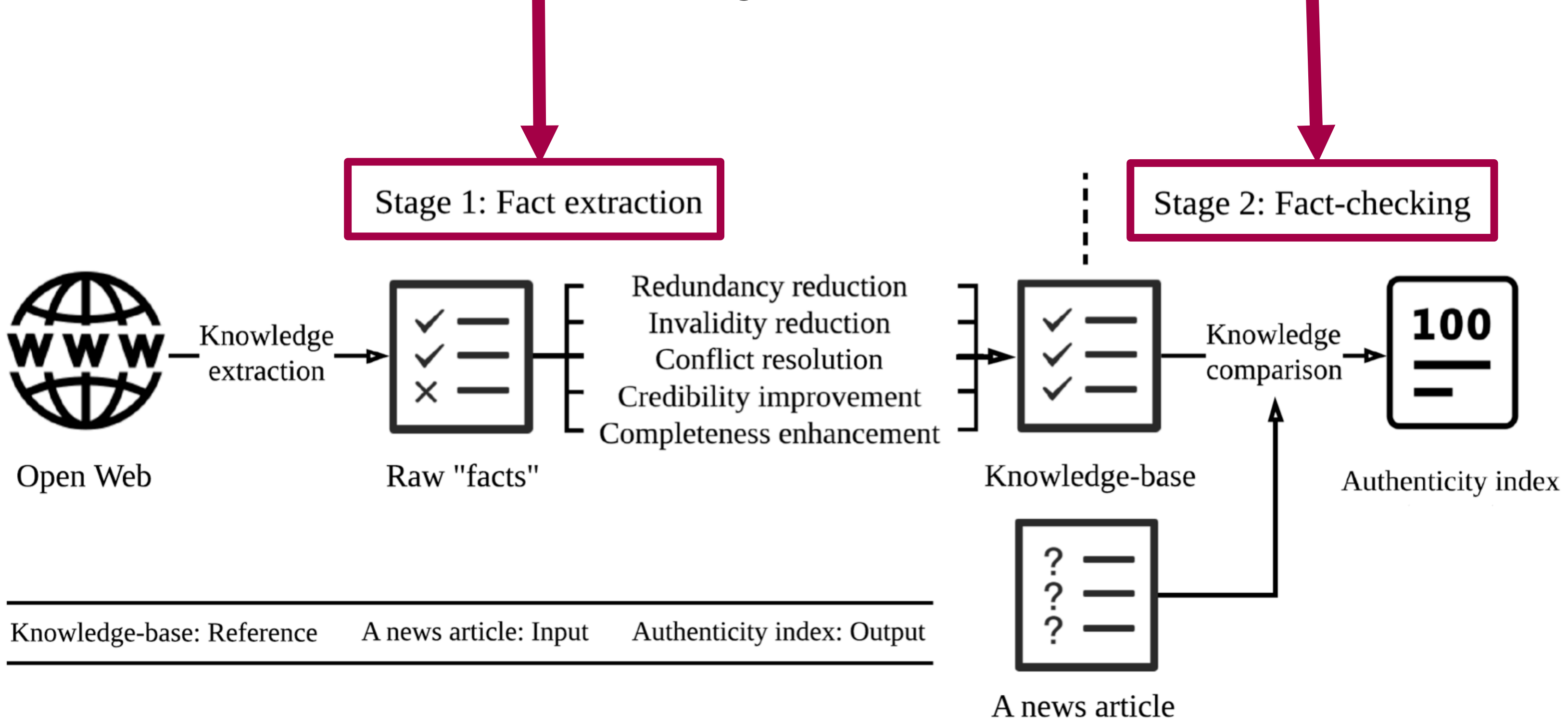
- *Manual Fact-checking* – providing ground truth.
- *Automatic Fact-checking* – a better choice for scalability.

Automatic Fact-checking

Overview

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?

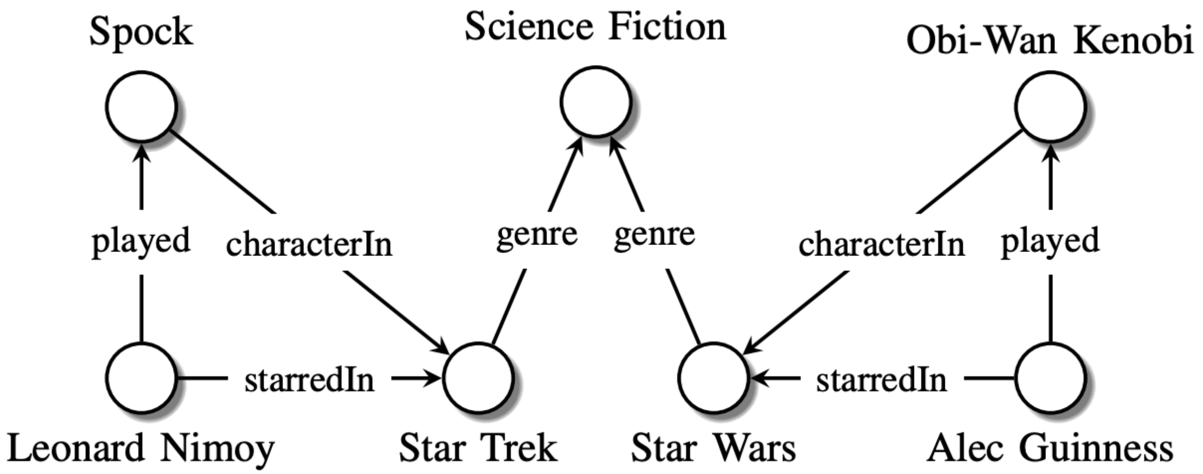


Knowledge Representation

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

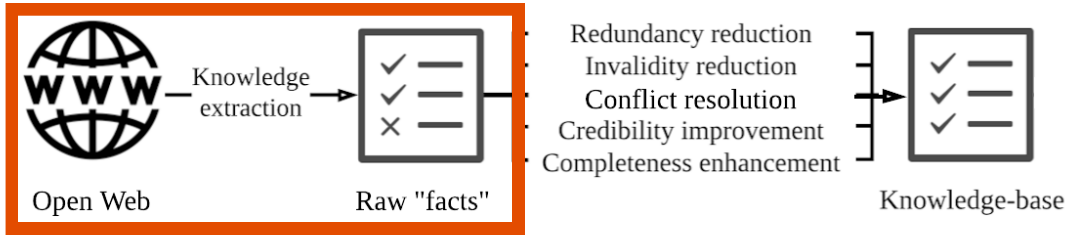
Stage 1. Fact Extraction

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

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

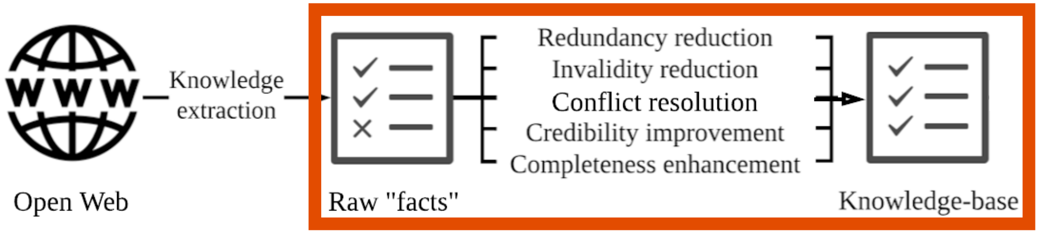
- E.g., *(Britain, joinIn, EuropeanUnion)* has been outdated.
- Use Compound Value Type (CVT): facts having beginning and end dates
- Timeliness studies are limited

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).



Open Web

Knowledge extraction

Raw "facts"

Redundancy reduction

Invalidity reduction

Conflict resolution

Credibility improvement

Completeness enhancement

Knowledge-base

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

Stage 1. Fact Extraction

Constructing knowledge graph to obtain the known facts

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**

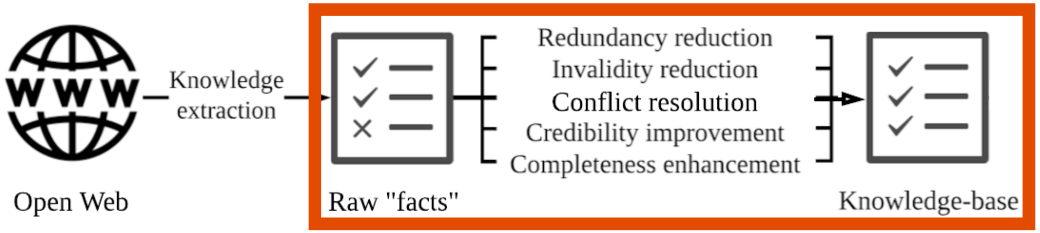
Assume the existence of knowledge-base triples is conditionally independent given latent features and parameters

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



Open Web

Knowledge extraction

Raw "facts"

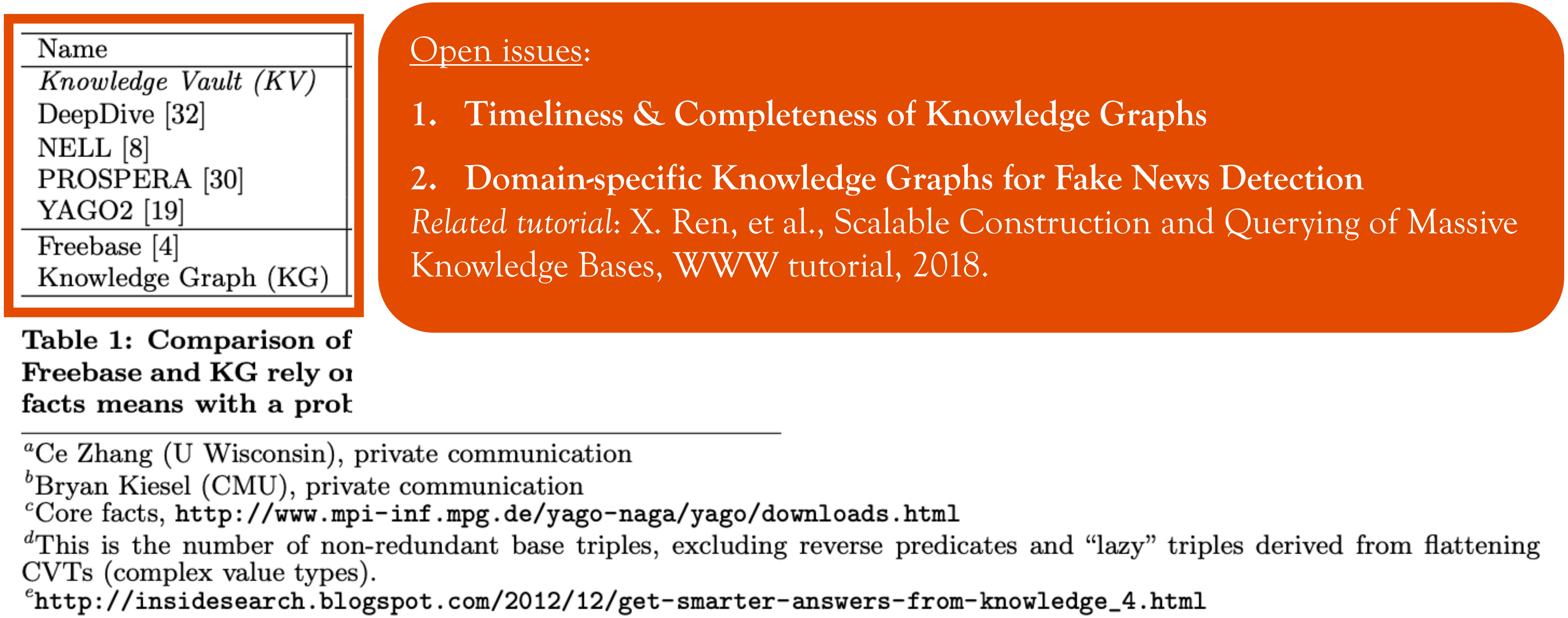
Redundancy reduction
Invalidity reduction
Conflict resolution
Credibility improvement
Completeness enhancement

Knowledge-base

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

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 on facts means with a prob

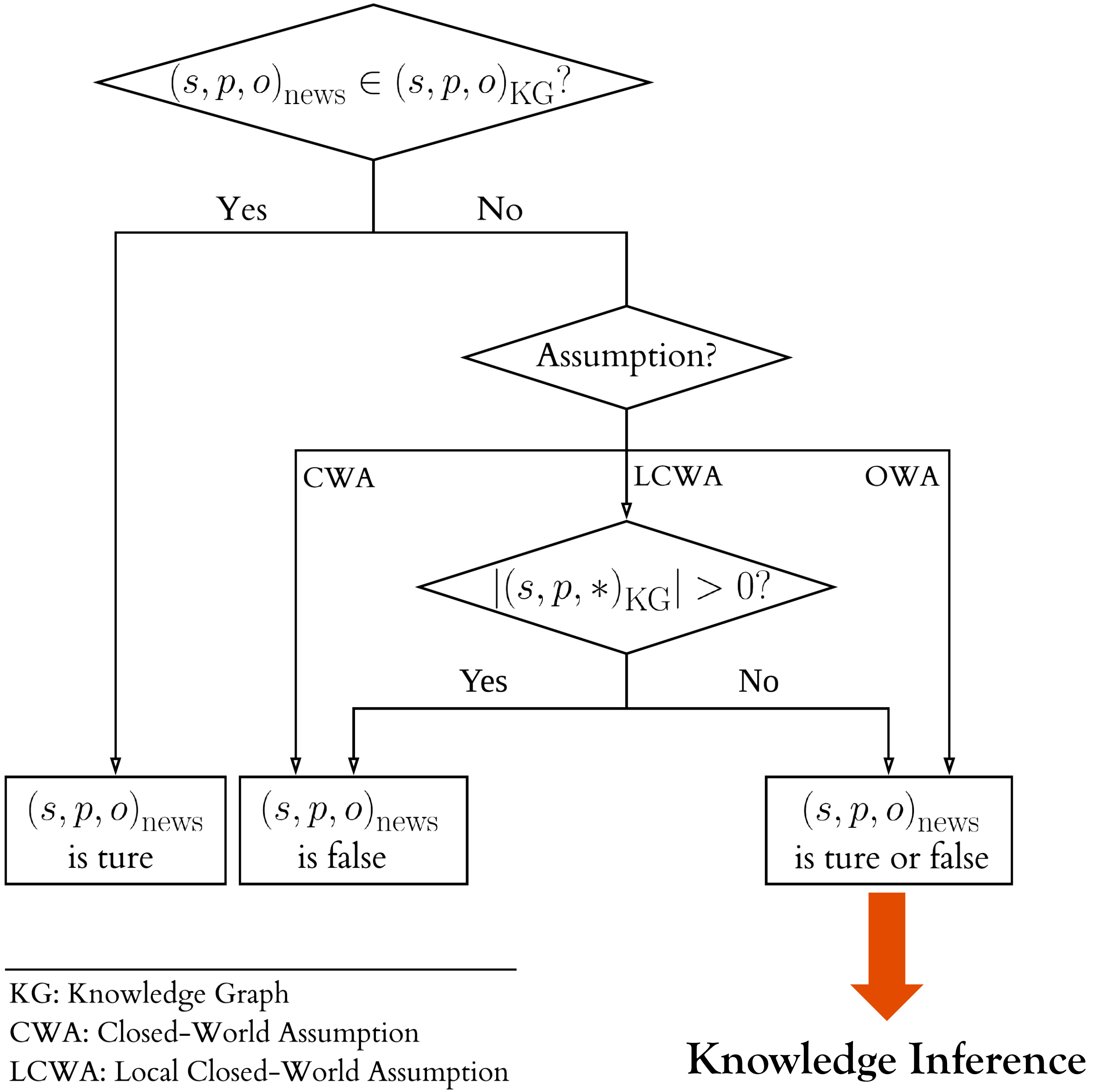
[a] Ce Zhang (U Wisconsin), private communication
[b] Bryan Kiesel (CMU), private communication
[c] Core facts, `http://www.mpi-inf.mpg.de/yago-naga/yago/downloads.html`
[d] This is the number of non-redundant base triples, excluding reverse predicates and "lazy" triples derived from flattening CVTs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

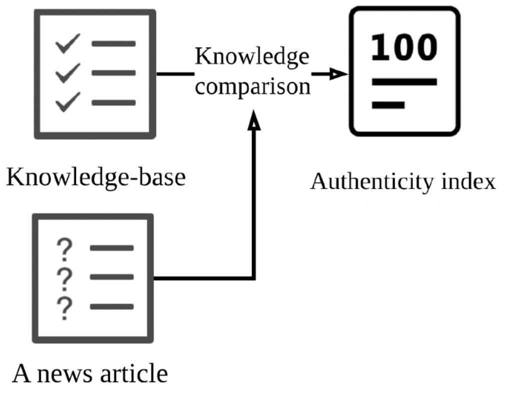

$(s,p,o)_{\text{news}} \in (s,p,o)_{\text{KG}}$?
Yes
No
Assumption?
CWA
LCWA
OWA
$|(s,p,*)_{\text{KG}}| > 0$?
Yes
No
$(s,p,o)_{\text{news}}$ is ture
$(s,p,o)_{\text{news}}$ is false
$(s,p,o)_{\text{news}}$ is ture or false
Knowledge Inference

KG: Knowledge Graph
CWA: Closed-World Assumption
LCWA: Local Closed-World Assumption
OWA: Open-World Assumption

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

Stage 2. Fact-checking

Comparing knowledge between news articles and knowledge graphs



Stage 2. Fact-checking

Knowledge Inference for unknown SPO triples: Illustrated studies

Shortest path-based method:

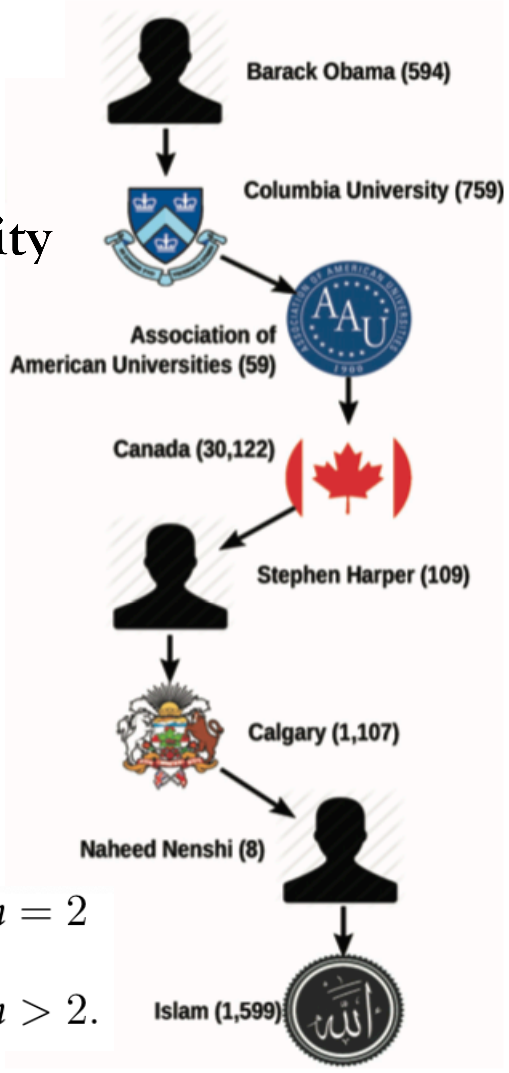
By finding the **shortest path** between concept nodes under properly defined **semantic proximity** metrics on knowledge graphs

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

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

An alternative formulation (widest bottleneck)

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

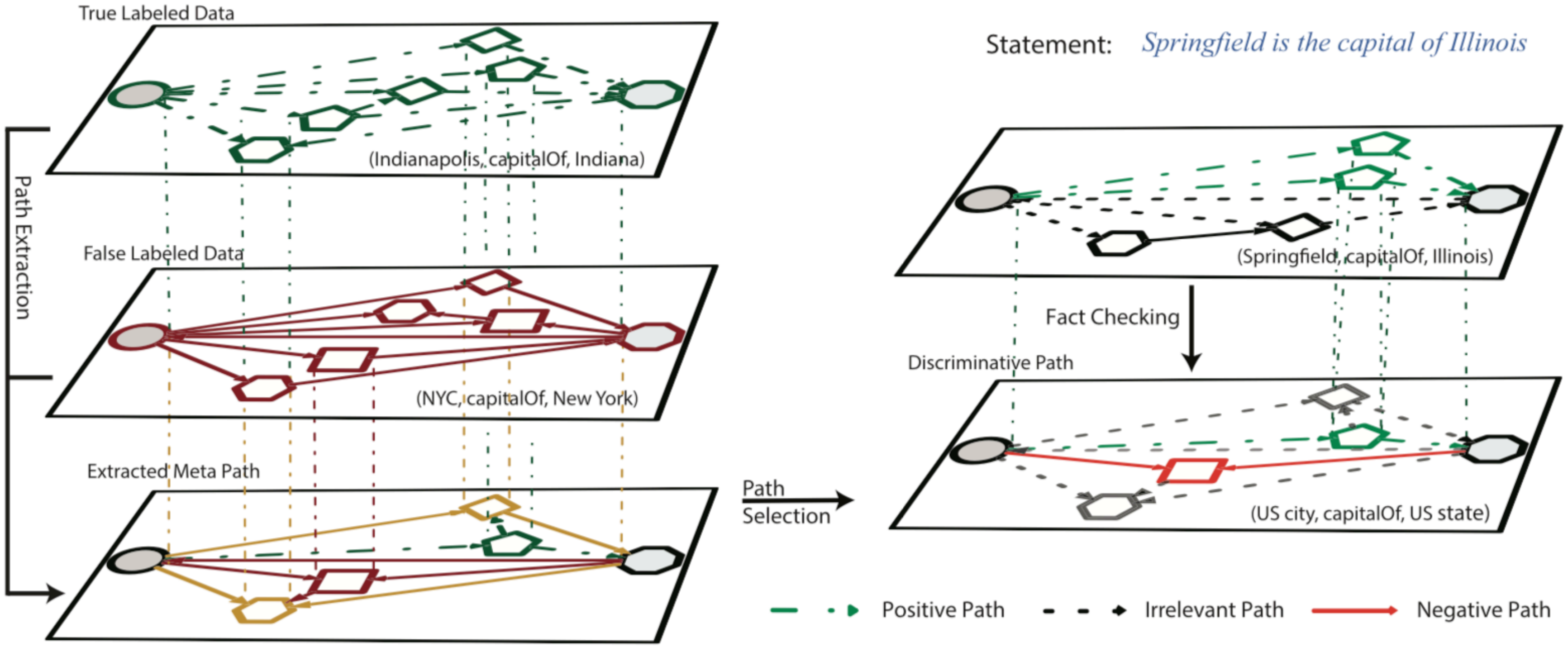


G. Ciampaglia, et al. Computational Fact Checking from Knowledge Networks, 2016

Stage 2. Fact-checking

Knowledge Inference for unknown SPO triples: Illustrated studies

Discriminative path-based method:


True Labeled Data
(Indianapolis, capitalOf, Indiana)
False Labeled Data
(NYC, capitalOf, New York)
Extracted Meta Path
Path Extraction
Path Selection
Statement: Springfield is the capital of Illinois
(Springfield, capitalOf, Illinois)
Fact Checking
Discriminative Path
(US city, capitalOf, US state)
Positive Path
Irrelevant Path
Negative Path

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

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

Style-based Fake News Detection

Overview

Style-based Fake News Detection is able to assess news intention by comparing the *writing style* extracted from to-be-verified *news content* with fake news style.

Fake News Style is a set of machine learning features that can well represent fake news and differentiate fake news from truth.

- *Textual* (*linguistic*) style features
- *Visual* style features

- *Manually* select features → Often within a *supervised* machine learning framework
- *Automatically* select features → Often within a *deep* machine learning framework

Style-based Fake News Detection

Overview

Style-based Fake News Detection is able to assess news intention by comparing the *writing style* extracted from to-be-verified *news content* with fake news style.

Fake News Style is a set of machine learning features that can well represent fake news and differentiate fake news from truth.

- *Textual (linguistic)* style features
- *Visual* style features

“More people watched President Trump’s 2019 State of the Union address on television than watched Super Bowl Super Bowl LlII”

Textual (Linguistic) Style of Fake News

Structure-based language features

Level	Feature(s)	Technique(s) and Tool(s)	Reference(s)
Lexicon	Words	Bag of words	Perez-Rosas et al., 2017
		+ n-gram to capture the word sequence	
		+ TF-IDF to unify the content length	
Syntax	Part-Of-Speech (POS) Tags	POS Taggers	Feng et al., 2012 Petrov and Klein, 2007
	Context-Free Grammars (CFGs)	Probabilistic Context Free Grammars (PCFGs) Parsers	
Semantic	Psycholinguistic Words	Linguistic Inquiry and Word Count (LIWC)	Perez-Rosas et al., 2017
Discourse	Rhetorical Relationships	Rhetorical Structure Theory (RST) Parser	Rubin and Lukoianova, 2015 Ji and Eisenstein, 2014

Textual (Linguistic) Style of Fake News

Structure-based language features

“The rat ate the cheese”

Level	Feature(s)	Technique(s) and Tool(s)	Reference(s)
Lexicon	Words	Bag of words + n-gram to capture the word sequence TF-IDF to unify the content length	Perez-Rosas et al., 2017
Syntax	Part-Of-Speech (P ags	Taggers	Feng et al., 2012 Petrov and Klein, 2007
	Co G	Free Grammars (PCFGs) Parsers	
Semantic		d Word Count (LIWC)	Perez-Rosas et al., 2017
Discourse	Rhetorical Relationships	Rhetorical Structure Theory (RST) Parser	Rubin and Lukoianova, 2015 Ji and Eisenstein, 2014

“the”: 2 “rat”: 1
“ate”: 1 “cheese”: 1

P(“ate” | “rat”) = ?
P(“cheese” | “ate”) = ?

Textual (Linguistic) Style of Fake News

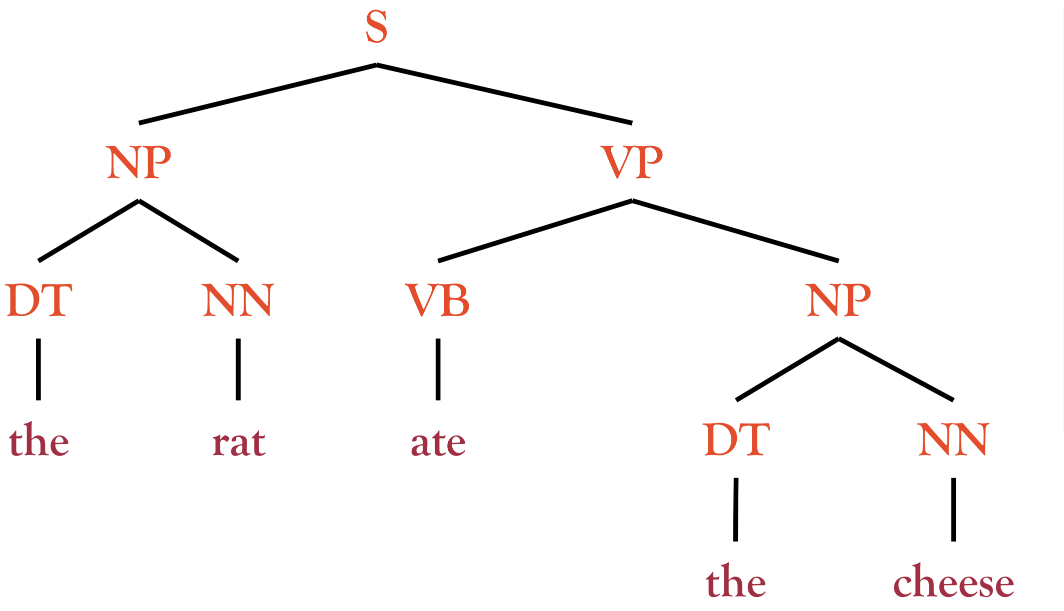
Structure-based language features

“The rat ate the cheese”

Level	Feature(s)	Technique(s) and To
Lexicon	Words	Bag of words
		+ n-gram to capture t
		+ TF-IDF to unify the
Syntax	Part-Of-Speech (POS) Tags	POS Taggers
	Context-Free Grammars (CFGs)	Probabilistic Context
Semantic	Ps linguisti Words	Linguistic Inquiry an
	al I	ure

NN: 2 (“rat”, “cheese”)
DT: 1 (“the”)
VB: 1 (“ate”)

S → NP VP
VP → VB NP
NN → rat
NN → cheese
NP → DT NN
DT → the
VB → ate

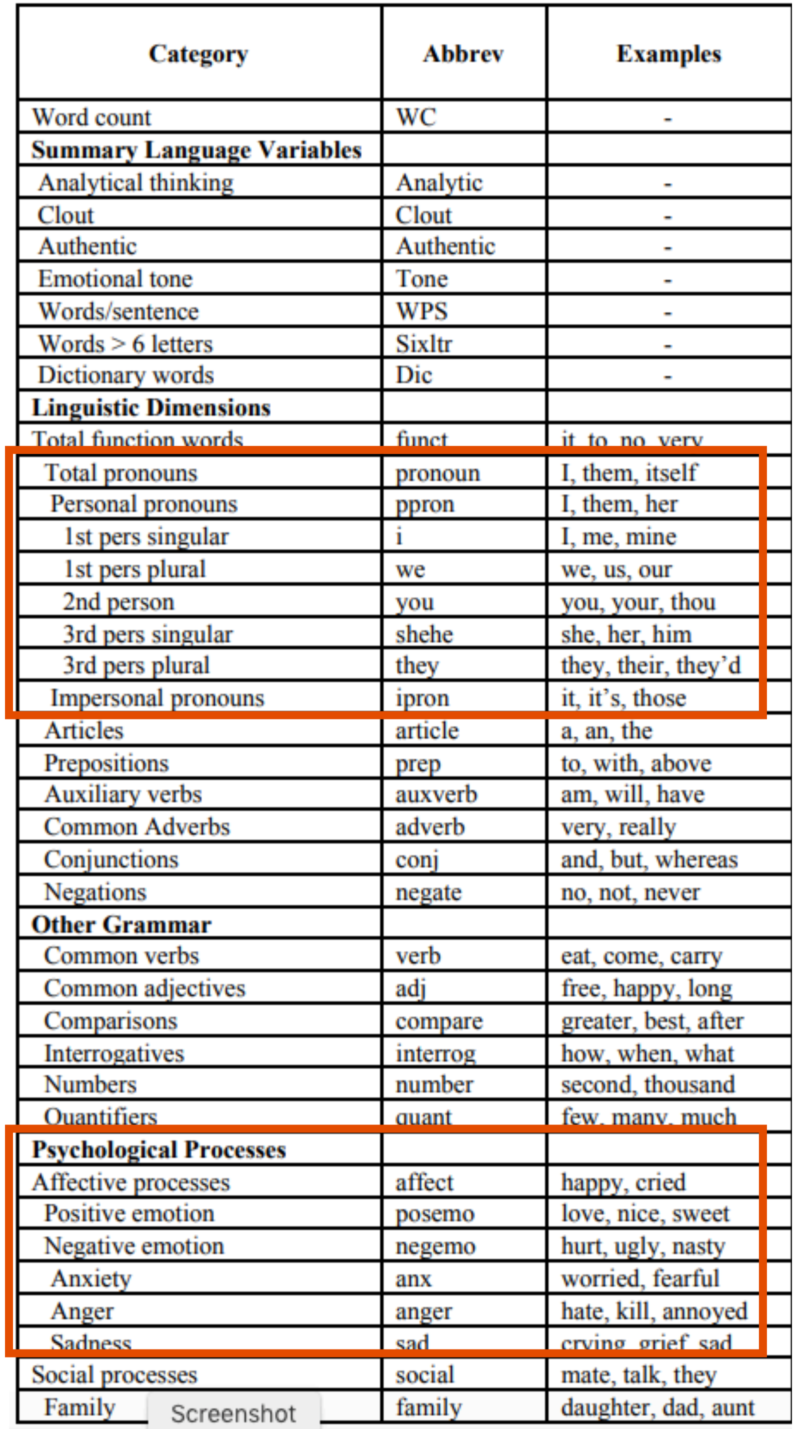


Textual (Linguistic) S

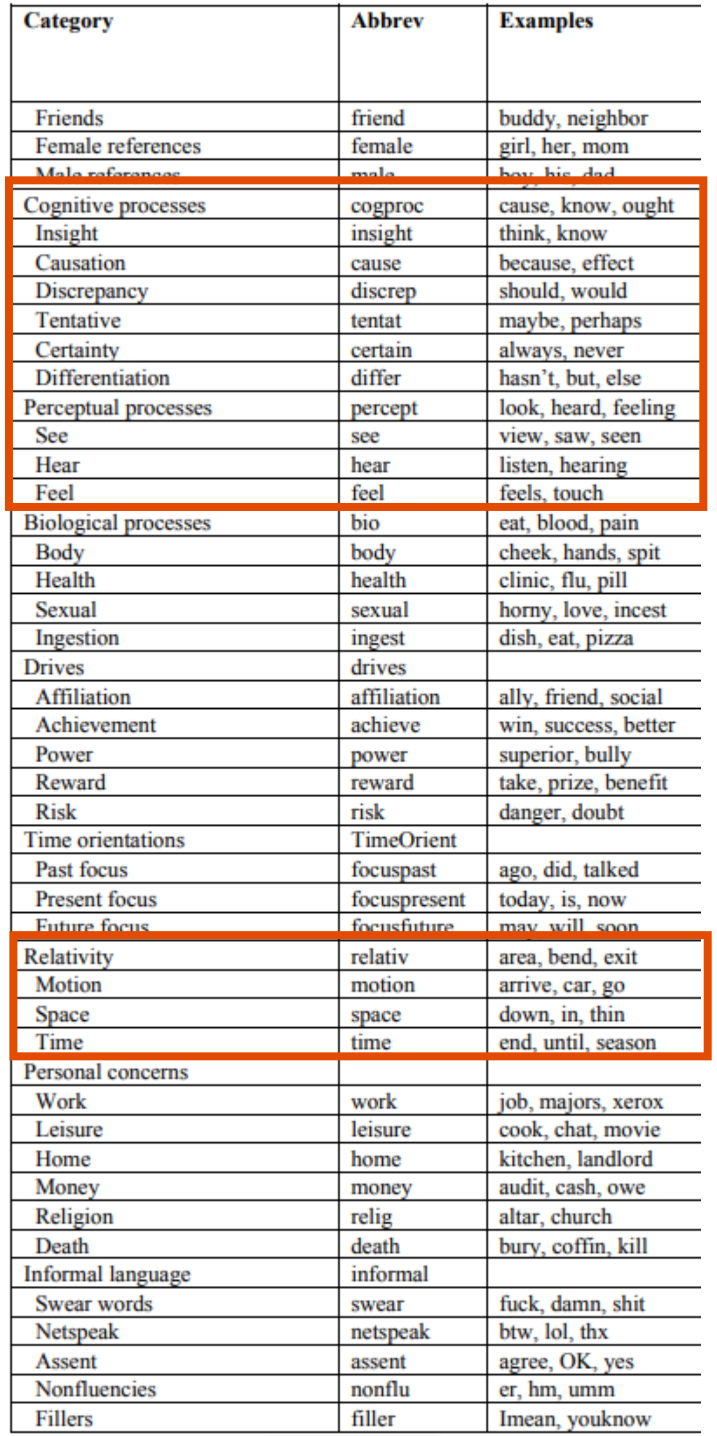
Structure-based language features

Level	Feature(s)
Lexicon	Words
Syntax	Part-Of-Speech (POS) Tags
	Context-Free Grammars (CFGs)
Semantic	Psycholinguistic Words
Discourse	Rhetorical Relationships

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


Screenshot

Category	Abbrev	Examples
Word count	WC	-
Summary Language Variables		
Analytical thinking	Analytic	-
Clout	Clout	-
Authentic	Authentic	-
Emotional tone	Tone	-
Words/sentence	WPS	-
Words > 6 letters	Sixltr	-
Dictionary words	Dic	-
Linguistic Dimensions		
Total function words	funct	it, to, no, very
Total pronouns	pronoun	I, them, itself
Personal pronouns	ppron	I, them, her
1st pers singular	i	I, me, mine
1st pers plural	we	we, us, our
2nd person	you	you, your, thou
3rd pers singular	shehe	she, her, him
3rd pers plural	they	they, their, they'd
Impersonal pronouns	ipron	it, it's, those
Articles	article	a, an, the
Prepositions	prep	to, with, above
Auxiliary verbs	auxverb	am, will, have
Common Adverbs	adverb	very, really
Conjunctions	conj	and, but, whereas
Negations	negate	no, not, never
Other Grammar		
Common verbs	verb	eat, come, carry
Common adjectives	adj	free, happy, long
Comparisons	compare	greater, best, after
Interrogatives	interrog	how, when, what
Numbers	number	second, thousand
Quantifiers	quant	few, many, much
Psychological Processes		
Affective processes	affect	happy, cried
Positive emotion	posemo	love, nice, sweet
Negative emotion	negemo	hurt, ugly, nasty
Anxiety	anx	worried, fearful
Anger	anger	hate, kill, annoyed
Sadness	sad	crying, grief, sad
Social processes	social	mate, talk, they
Family	family	daughter, dad, aunt

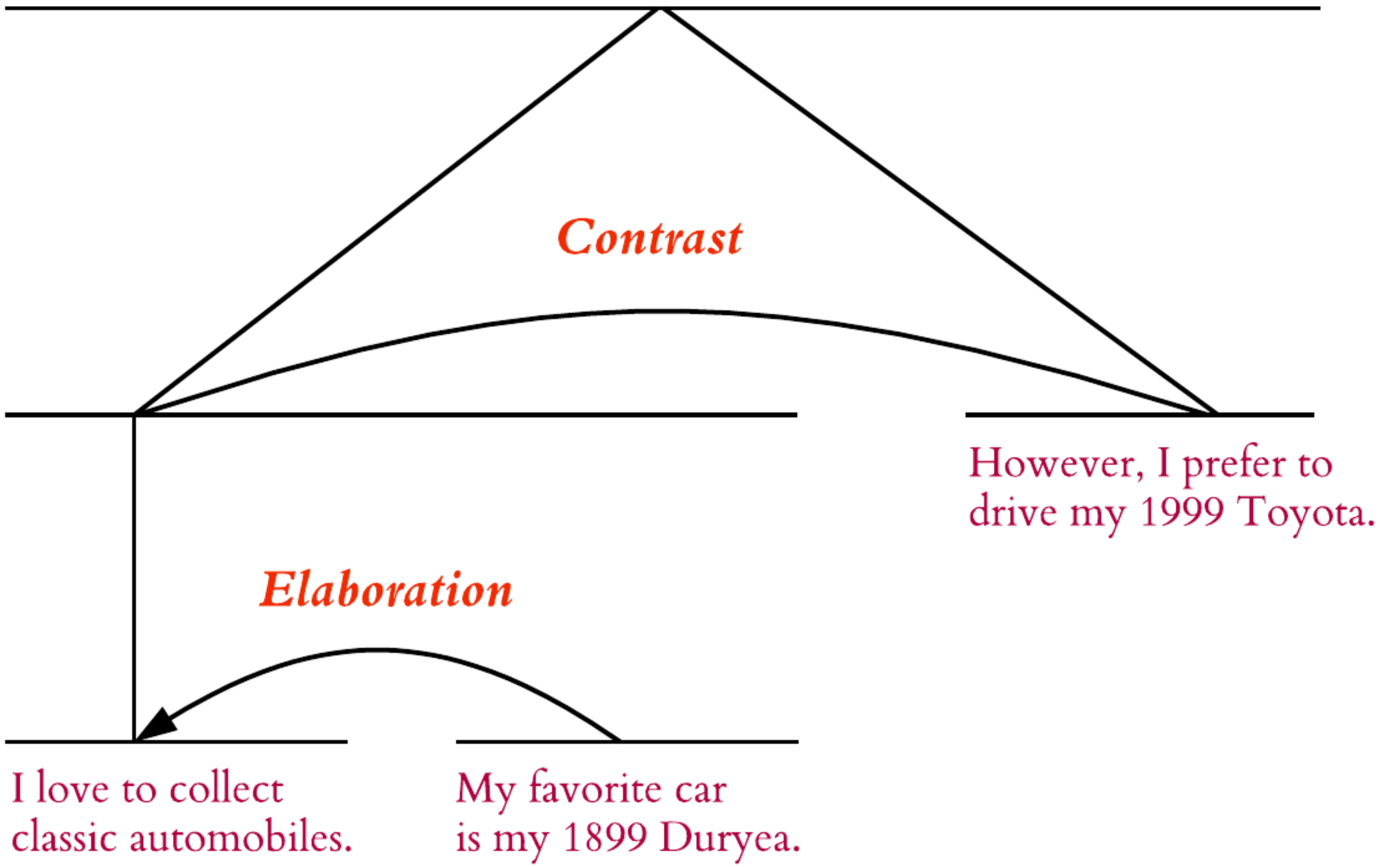


Category	Abbrev	Examples
Friends	friend	buddy, neighbor
Female references	female	girl, her, mom
Male references	male	boy, his, dad
Cognitive processes	cogproc	cause, know, ought
Insight	insight	think, know
Causation	cause	because, effect
Discrepancy	discrep	should, would
Tentative	tentat	maybe, perhaps
Certainty	certain	always, never
Differentiation	differ	hasn't, but, else
Perceptual processes	percept	look, heard, feeling
See	see	view, saw, seen
Hear	hear	listen, hearing
Feel	feel	feels, touch
Biological processes	bio	eat, blood, pain
Body	body	cheek, hands, spit
Health	health	clinic, flu, pill
Sexual	sexual	horny, love, incest
Ingestion	ingest	dish, eat, pizza
Drives	drives	
Affiliation	affiliation	ally, friend, social
Achievement	achieve	win, success, better
Power	power	superior, bully
Reward	reward	take, prize, benefit
Risk	risk	danger, doubt
Time orientations	TimeOrient	
Past focus	focuspast	ago, did, talked
Present focus	focuspresent	today, is, now
Future focus	focusfuture	may, will, soon
Relativity	relativ	area, bend, exit
Motion	motion	arrive, car, go
Space	space	down, in, thin
Time	time	end, until, season
Personal concerns		
Work	work	job, majors, xerox
Leisure	leisure	cook, chat, movie
Home	home	kitchen, landlord
Money	money	audit, cash, owe
Religion	relig	altar, church
Death	death	bury, coffin, kill
Informal language	informal	
Swear words	swear	fuck, damn, shit
Netspeak	netspeak	btw, lol, thx
Assent	assent	agree, OK, yes
Nonfluencies	nonflu	er, hm, umm
Fillers	filler	Imean, youknow

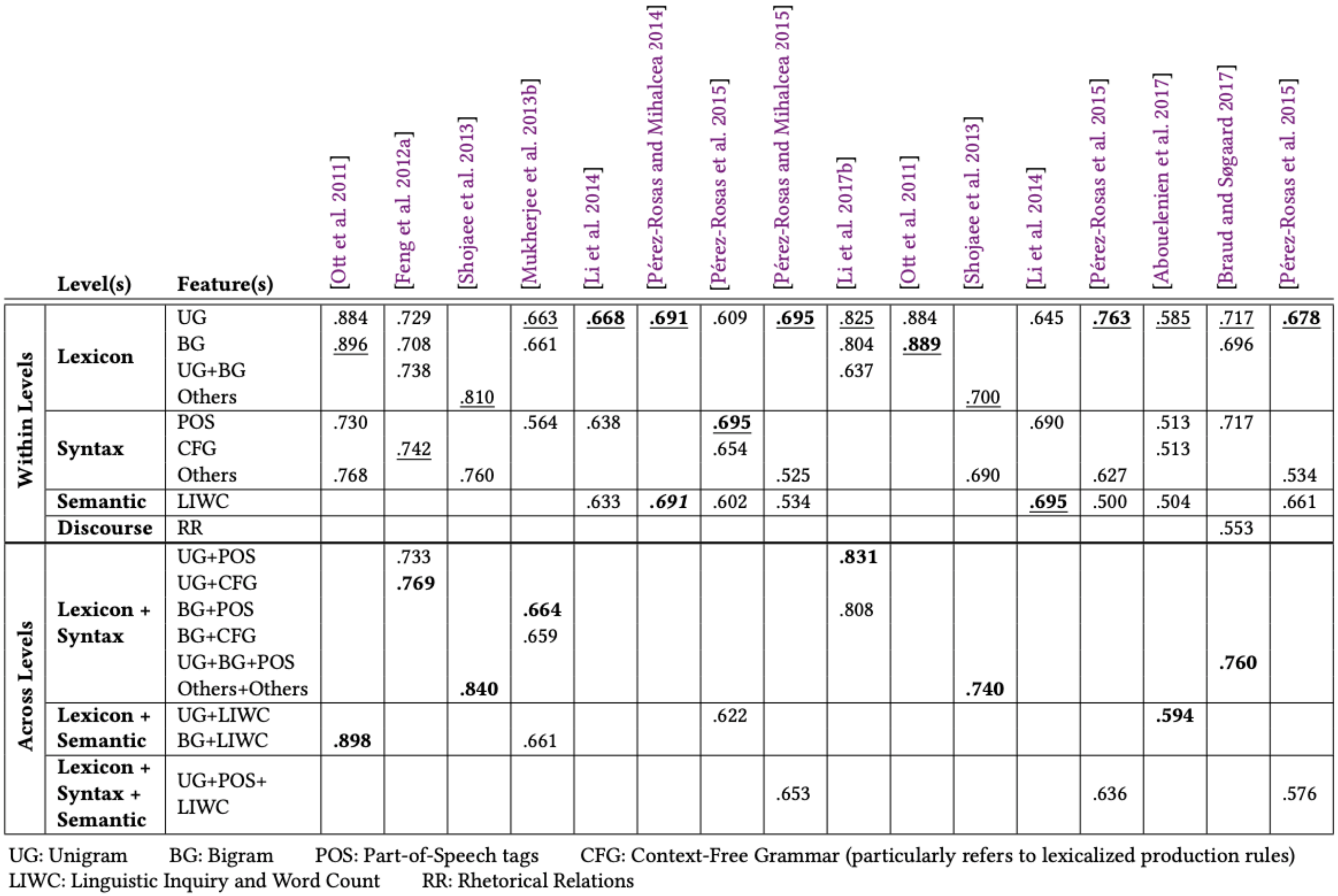
Textual (Linguistic) Style of Fake News

Structure-based language features

Level	Feature(s)
Lexicon	Words
Syntax	Part-Of-Speech (POS) Tags
	Context-Free Grammars (CFGs)
Semantic	Psycholinguistic Words
Discourse	Rhetorical Relationships


Contrast
Elaboration
However, I prefer to drive my 1999 Toyota.
I love to collect classic automobiles.
My favorite car is my 1899 Duryea.

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



	Level(s)	Feature(s)	[Ott et al. 2011]	[Feng et al. 2012a]	[Shojaee et al. 2013]	[Mukherjee et al. 2013b]	[Li et al. 2014]	[Pérez-Rosas and Mihalcea 2014]	[Pérez-Rosas et al. 2015]	[Pérez-Rosas and Mihalcea 2015]	[Li et al. 2017b]	[Ott et al. 2011]	[Shojaee et al. 2013]	[Li et al. 2014]	[Pérez-Rosas et al. 2015]	[Abouelenien et al. 2017]	[Braud and Søgaard 2017]	[Pérez-Rosas et al. 2015]
Within Levels	**Lexicon**	UG	.884	.729		.663	**.668**	**.691**	.609	**.695**	.825	.884		.645	**.763**	.585	.717	**.678**
		BG	.896	.708		.661					.804	**.889**					.696	
		UG+BG		.738							.637							
		Others			.810								.700					
	Syntax	POS	.730			.564	.638		**.695**					.690		.513	.717	
		CFG		.742					.654							.513		
		Others	.768		.760					.525			.690		.627			.534
	Semantic	LIWC					.633	***.691***	.602	.534				**.695**	.500	.504		.661
	Discourse	RR															.553	
Across Levels	**Lexicon + Syntax**	UG+POS		.733							**.831**							
		UG+CFG		**.769**														
		BG+POS				**.664**					.808							
		BG+CFG				.659												
		UG+BG+POS															**.760**	
		Others+Others			**.840**								**.740**					
	Lexicon + Semantic	UG+LIWC							.622							**.594**		
		BG+LIWC	**.898**			.661												
	Lexicon + Syntax + Semantic	UG+POS+ LIWC								.653					.636			.576

UG: Unigram BG: Bigram POS: Part-of-Speech tags CFG: Context-Free Grammar (particularly refers to lexicalized production rules)
LIWC: Linguistic Inquiry and Word Count RR: Rhetorical Relations

Textual (Linguistic) Style of Fake News

Performance of structure-based language features

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

Textual (Linguistic) Style of Fake News

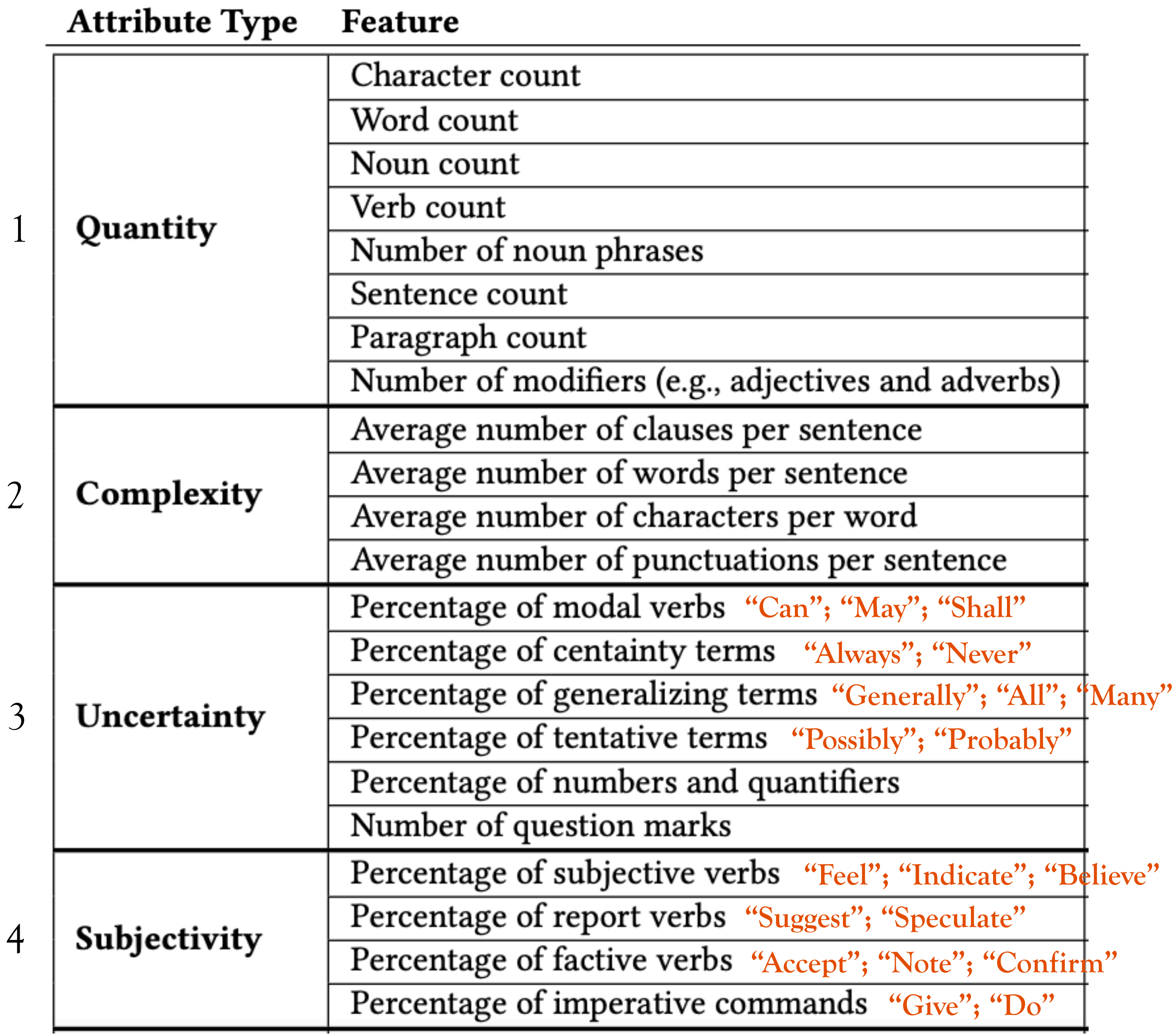
Attribute-based language features

- Most related studies belong to the general area of **Deception Detection**.
- Deception is **disinformation**, including **fake statements, fake reviews, fake news**, etc.
- Attributes are generally inspired from **forensic psychological theories**, e.g.,

Term	**Phenomenon**
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.

Textual (Linguistic) Style of Fake News

Attribute-based language features

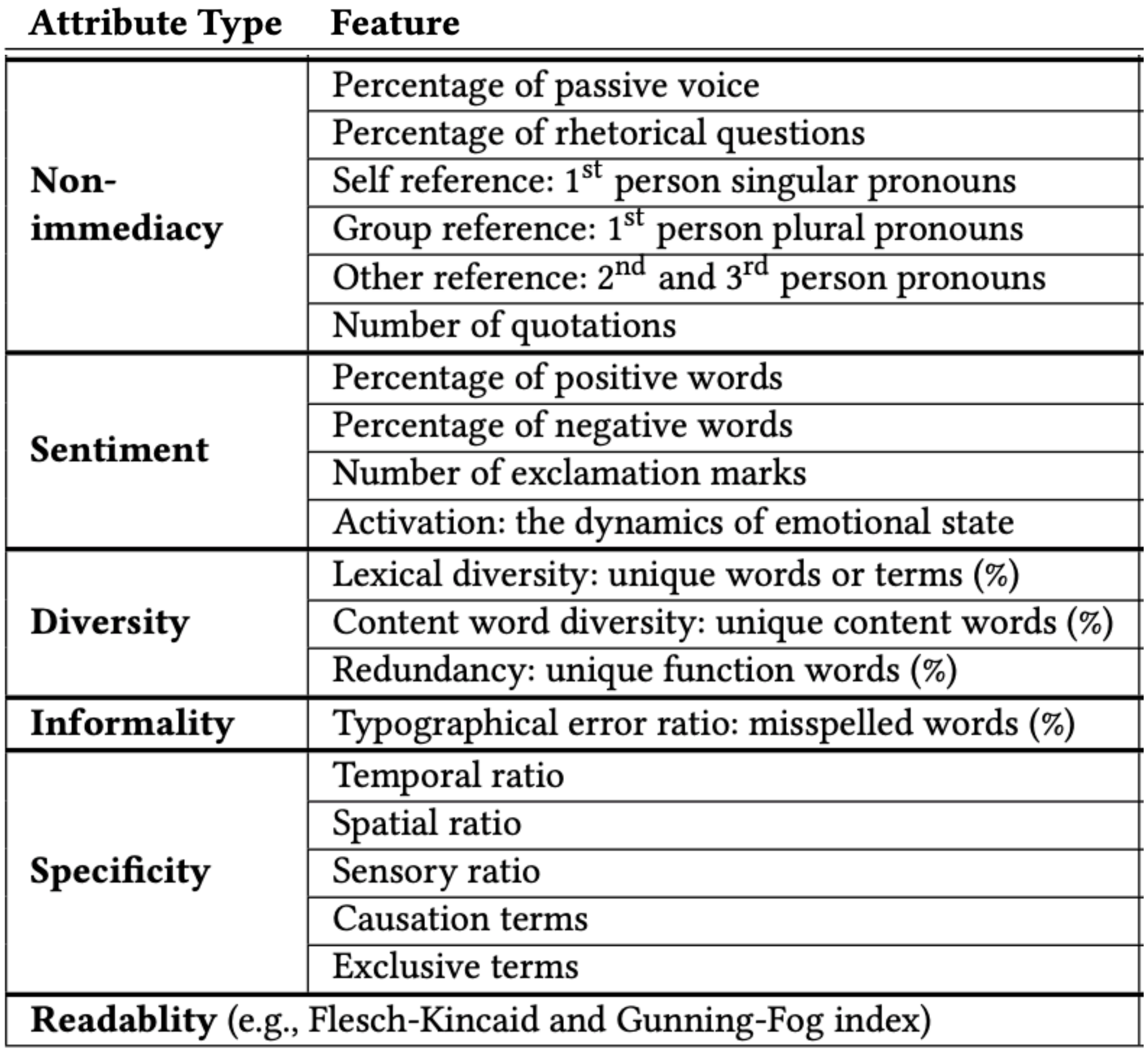


	Attribute Type	Feature
1	**Quantity**	Character count
		Word count
		Noun count
		Verb count
		Number of noun phrases
		Sentence count
		Paragraph count
		Number of modifiers (e.g., adjectives and adverbs)
2	**Complexity**	Average number of clauses per sentence
		Average number of words per sentence
		Average number of characters per word
		Average number of punctuations per sentence
3	**Uncertainty**	Percentage of modal verbs “Can”; “May”; “Shall”
		Percentage of centainty terms “Always”; “Never”
		Percentage of generalizing terms “Generally”; “All”; “Many”
		Percentage of tentative terms “Possibly”; “Probably”
		Percentage of numbers and quantifiers
		Number of question marks
4	**Subjectivity**	Percentage of subjective verbs “Feel”; “Indicate”; “Believe”
		Percentage of report verbs “Suggest”; “Speculate”
		Percentage of factive verbs “Accept”; “Note”; “Confirm”
		Percentage of imperative commands “Give”; “Do”

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

Textual (Linguistic) Style of Fake News

Attribute-based language features



	Attribute Type	Feature
5	**Non-immediacy**	Percentage of passive voice
		Percentage of rhetorical questions
		Self reference: 1st person singular pronouns
		Group reference: 1st person plural pronouns
		Other reference: 2nd and 3rd person pronouns
		Number of quotations
6	**Sentiment**	Percentage of positive words
		Percentage of negative words
		Number of exclamation marks
		Activation: the dynamics of emotional state
7	**Diversity**	Lexical diversity: unique words or terms (%)
		Content word diversity: unique content words (%)
		Redundancy: unique function words (%)
8	**Informality**	Typographical error ratio: misspelled words (%)
9	**Specificity**	Temporal ratio
		Spatial ratio
		Sensory ratio
		Causation terms
		Exclusive terms
10	**Readablity** (e.g., Flesch-Kincaid and Gunning-Fog index)	

"Car"; "Red"

"Are"; "An"

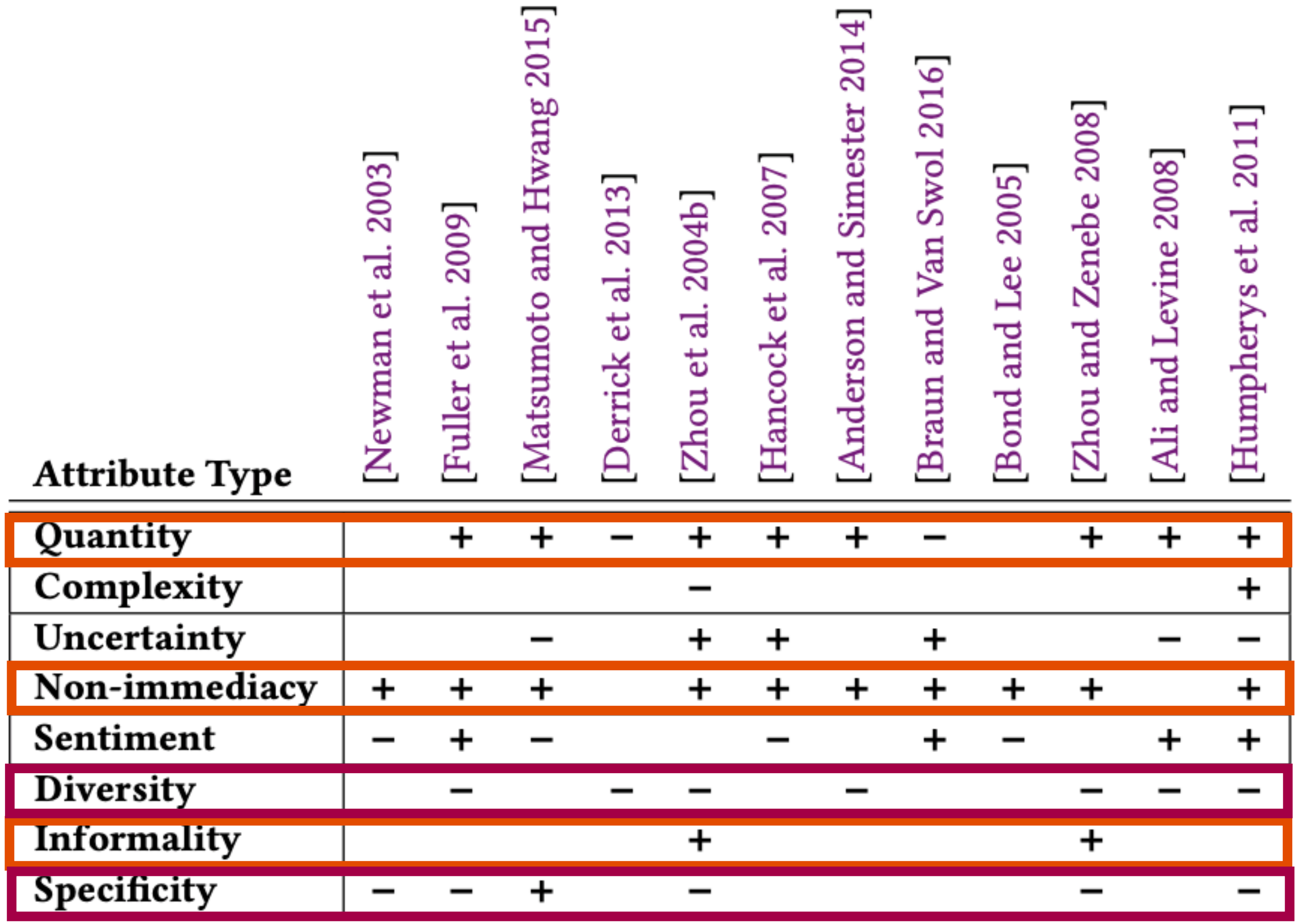
0.4[(#words/#sentences)+(#long_words/#words)

The general construct of immediacy and nonimmediacy refers to (non-)verbal behaviors that create **a psychological sense of closeness or distance**.

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

Textual (Linguistic) Style of Fake News

Performance of attribute-based language features



Attribute Type	[Newman et al. 2003]	[Fuller et al. 2009]	[Matsumoto and Hwang 2015]	[Derrick et al. 2013]	[Zhou et al. 2004b]	[Hancock et al. 2007]	[Anderson and Simester 2014]	[Braun and Van Swol 2016]	[Bond and Lee 2005]	[Zhou and Zenebe 2008]	[Ali and Levine 2008]	[Humpherys et al. 2011]
Quantity		+	+	−	+	+	+	−		+	+	+
Complexity					−							+
Uncertainty			−		+	+		+			−	−
Non-immediacy	+	+	+		+	+	+	+	+	+		+
Sentiment	−	+	−			−		+	−		+	+
Diversity		−		−	−		−			−	−	−
Informality					+					+		
Specificity	−	−	+		−					−		−

+: The attribute is positively related to the existence of deception;

−: The attribute is negatively related to the existence of deception.

- Quantity ↑
- Non-immediacy ↑
- Informality ↑
- Diversity ↓
- Specificity ↓

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

Style-based Fake News Detection

Overview

Style-based Fake News Detection is able to assess news intention by comparing the *writing style* extracted from to-be-verified *news content* with fake news style.

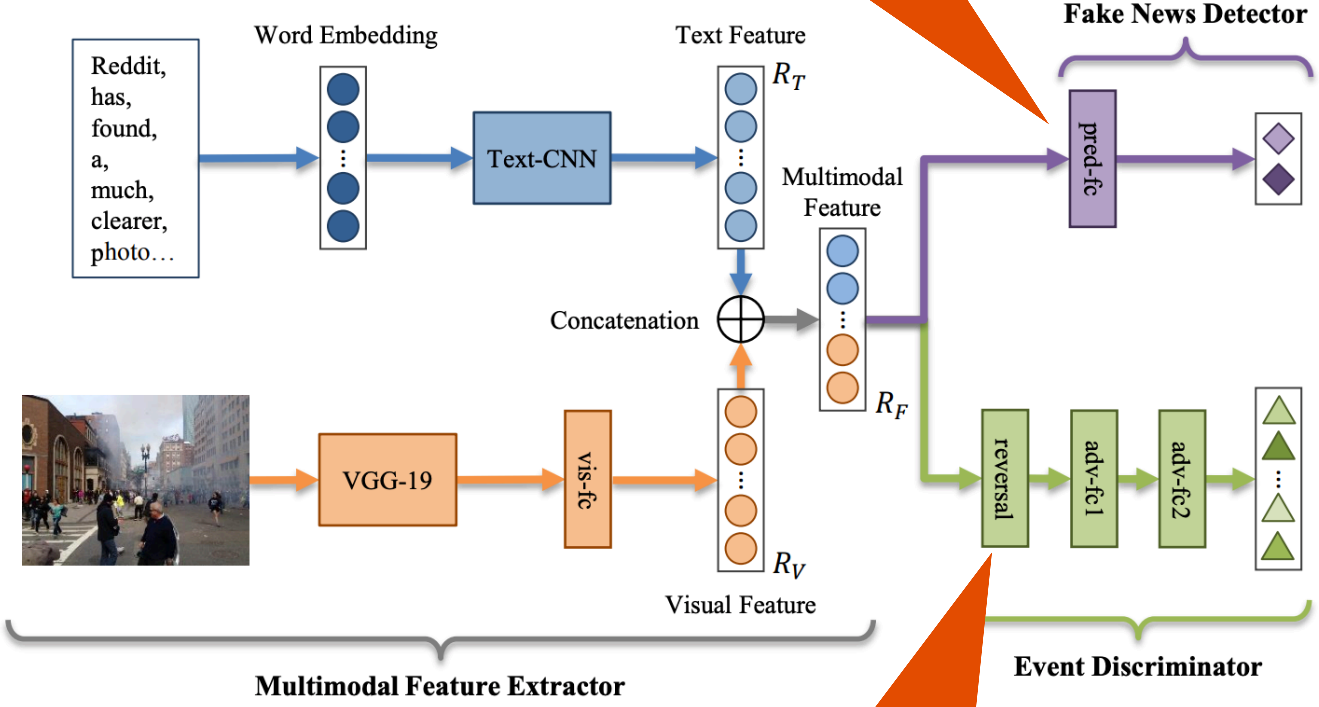
Fake News Style is a set of machine learning features that can well represent fake news and differentiate fake news from truth.

- *Textual (linguistic)* style features
- *Visual* style features

Visual Style of Fake News

An illustration: EANN

A fully connected layer with softmax



Two fully connected layers with activation functions

<u>EANN</u>:
multi-modal;
adversarial network inspired;
fake news early detection

<u>Fake News Early Detection:</u>
extract a set of **generalizable** and **discriminable** features to represent news content and detect fake news

W. Yaqing, et al., EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. *KDD'18*

Knowledge- & Style-based Fake News Detection

Summary

How to involve *social context information* of fake news, e.g., its propagation patterns on social networks?

	Knowledge-based fake news detection	**Style-based fake news detection**
Information utilized	News content	News content
Modality involved	Single: only text	Single or multi: text, visual, etc.
Objective(s) evaluated	News authenticity	News authenticity and intention
Framework for solving the problem	Link prediction	Machine learning
Related topic	Fact-checking	Deception detection
Open issues	Timeliness and completeness of knowledge graphs	Cross-domain, language, topic fake news studies

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

Propagation-based Fake News Detection

Overview

Propagation-based Fake News Detection utilizes social context information to explore the relationships among entities in news propagation.

- Entities, e.g., spreaders (users) of news, publishers of news, posts of users
- Relationships among the same or different entities

Basis of propagation-based fake news detection approaches

- **News cascades (propagation trees)** – a *direct* way to present news propagation
- **Self-defined graphs (networks)** – an *indirect* way to present news propagation

Propagation-based Fake News Detection

Overview

Propagation-based Fake News Detection utilizes social context information to explore the relationships among entities in news propagation.

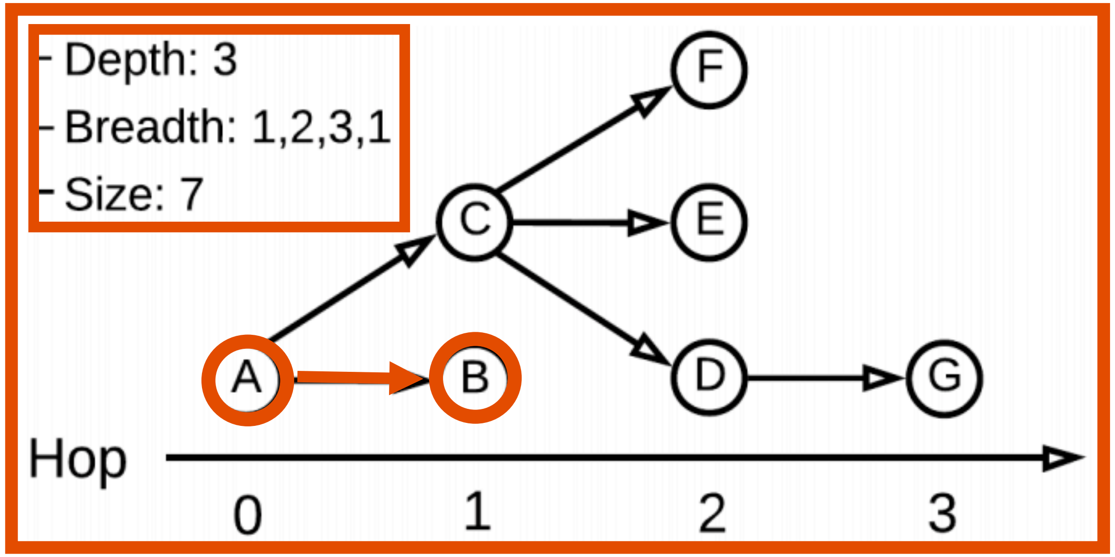
- Entities, e.g., spreaders (users) of news, publishers of news, posts of users
- Relationships among the same or different entities

Basis of propagation-based fake news detection approaches

- **News cascades (propagation trees)** – a *direct* way to present news propagation
- **Self-defined graphs (networks)** – an *indirect* way to present news propagation

News Cascade

Definition



A news cascade: One propagation path of a news article

Root node: The original post of user related to the news article

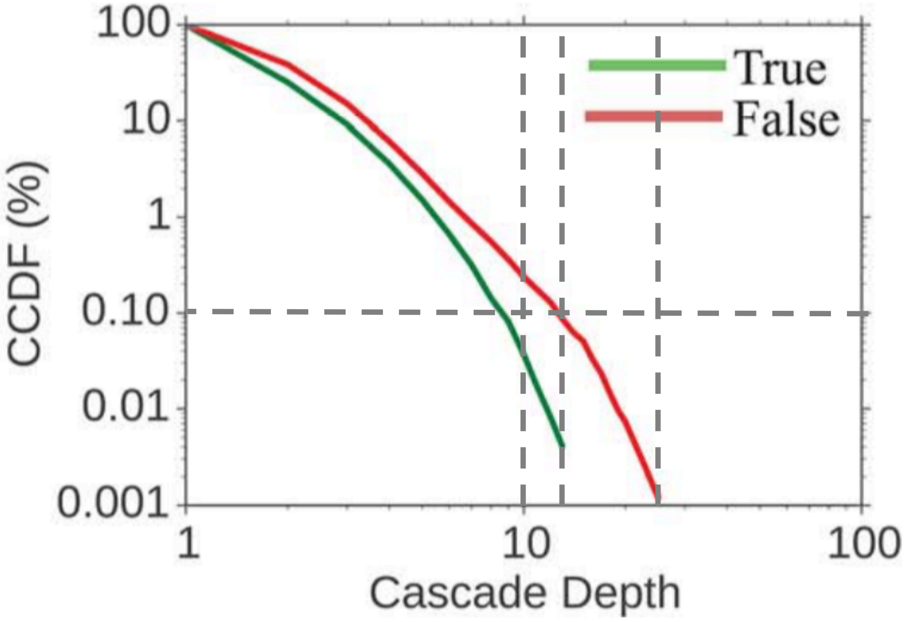
Other node: The re-post of the post of parent node

Directed Edge: Post → repost relationships

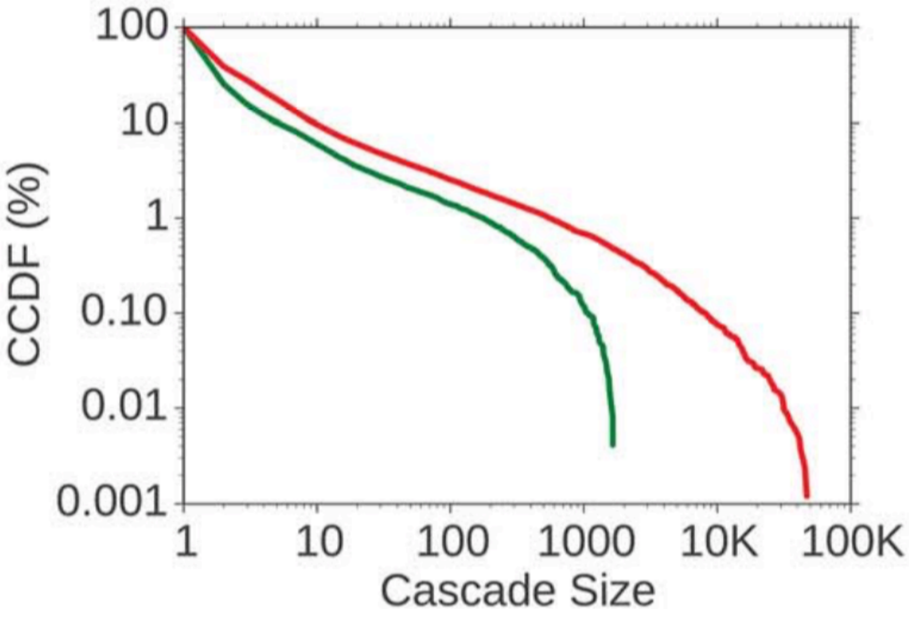
News Cascade

Illustrated studies –

A. Cascade-based pattern discovering

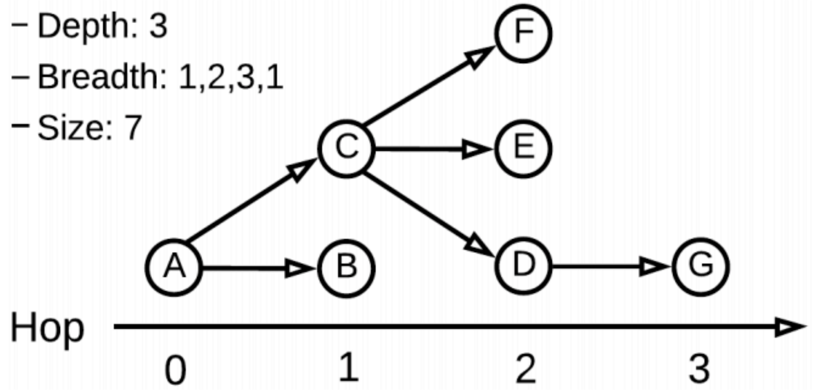


Fake news spreads deeper than the truth



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

Fake news spreads farther than the truth



- Depth: 3
- Breadth: 1,2,3,1
- Size: 7

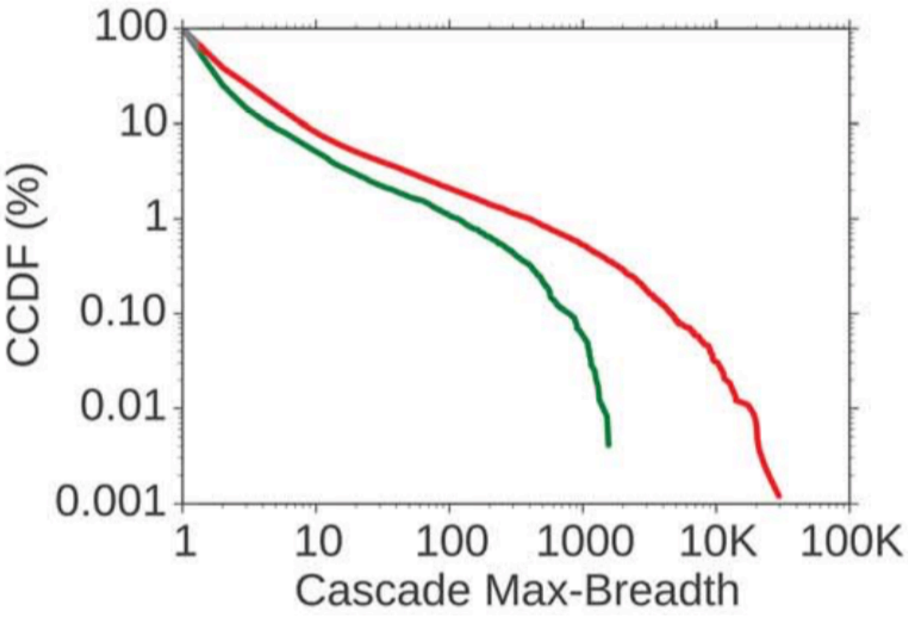
S. Vosoughi, et al. The spread of true and false news online. Science, 2018

News Cascade

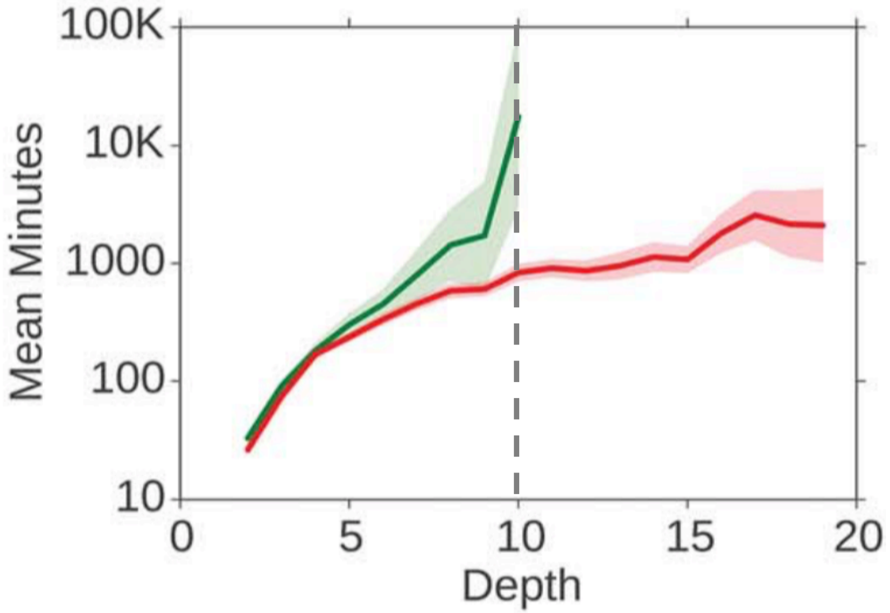
Illustrated studies –

*A. Cascade-based **pattern***

discovering of fake news

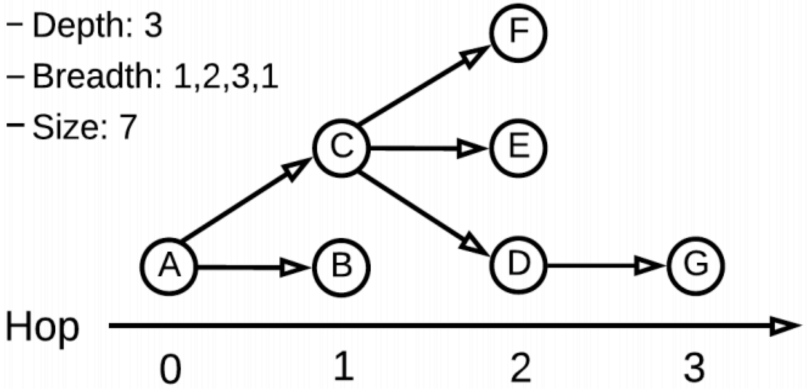


Fake news spreads more broadly than the truth



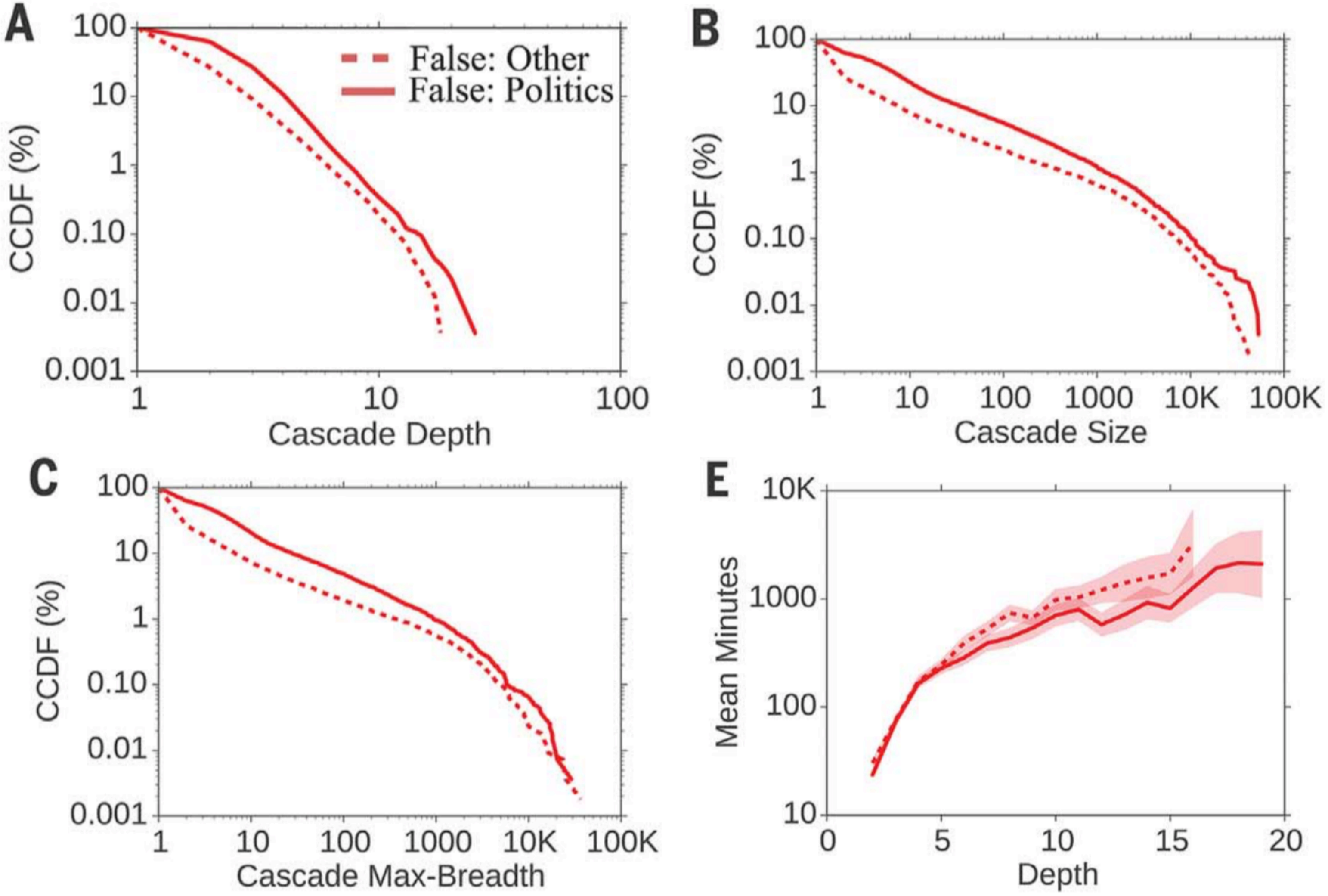
Fake news spreads faster than the truth

S. Vosoughi, et al. The spread of true and false news online. Science, 2018



- Depth: 3
- Breadth: 1,2,3,1
- Size: 7

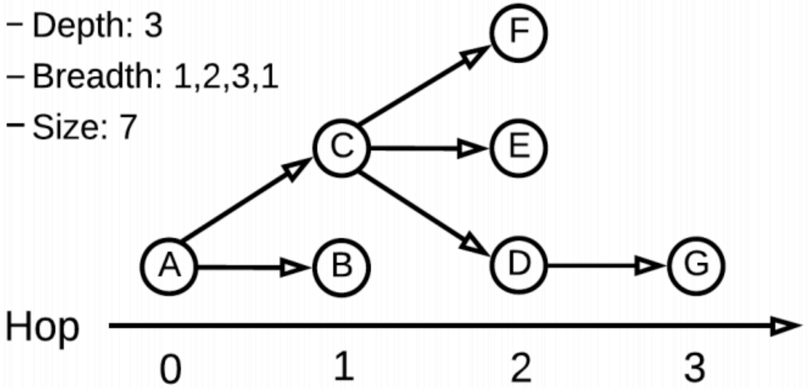
X. Zhou, R. Zafarani, K. Shu, H. Liu



News Cascade

Illustrated studies –

*A. Cascade-based **pattern** discovering of fake news*



- Depth: 3
- Breadth: 1,2,3,1
- Size: 7

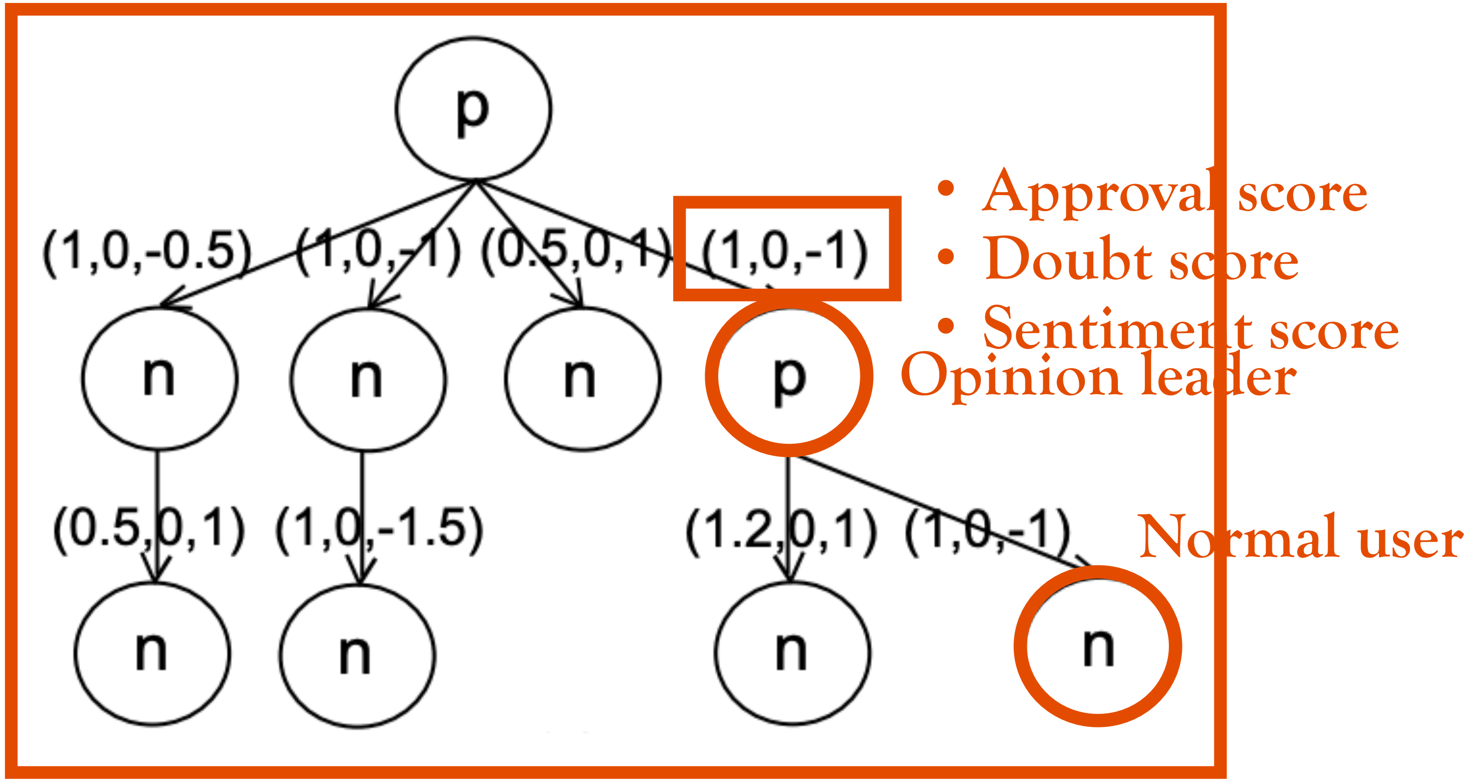
Political fake news spreads deeper, farther, more broadly and faster than fake news in other domains

S. Vosoughi, et al. The spread of true and false news online. Science, 2018

News Cascade

Illustrated studies –

*B. Fake news detection based on cascade **similarity***



Random walk graph kernel

Challenges:

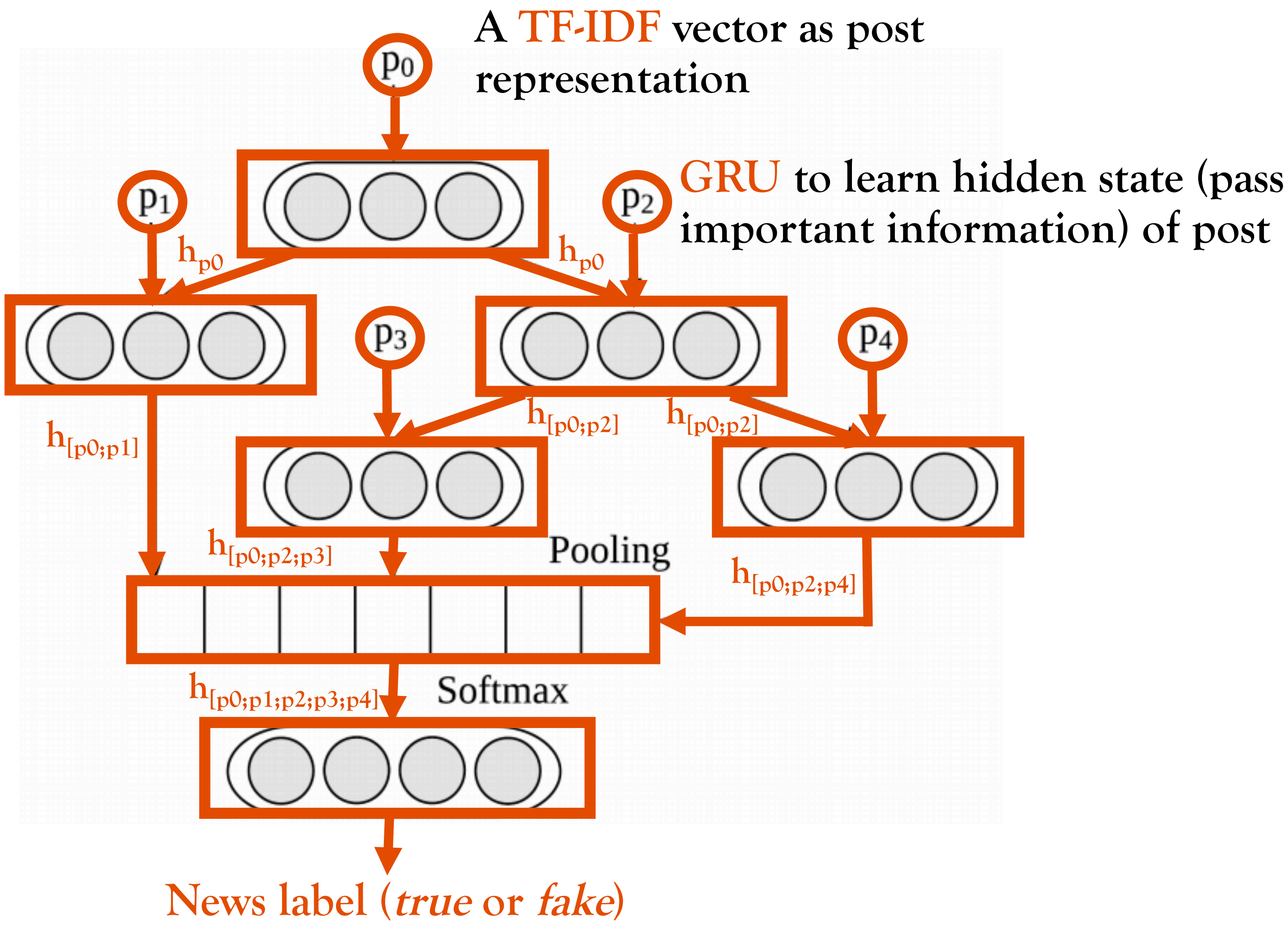
Computational expense, as similarity will be computed between pairwise cascades.

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

News Cascade

Illustrated studies –

C. Fake news detection based on cascade ***representation***


p0
A TF-IDF vector as post representation
p1
p2
GRU to learn hidden state (pass important information) of post
hp0
hp0
p3
p4
h[p0;p1]
h[p0;p2]
h[p0;p2]
h[p0;p2;p3]
Pooling
h[p0;p2;p4]
h[p0;p1;p2;p3;p4]
Softmax
News label (true or fake)

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

Challenges:

Cascade depth sensitivity, as the depth of cascade is equivalent to that of neural network.

Propagation-based Fake News Detection

Overview

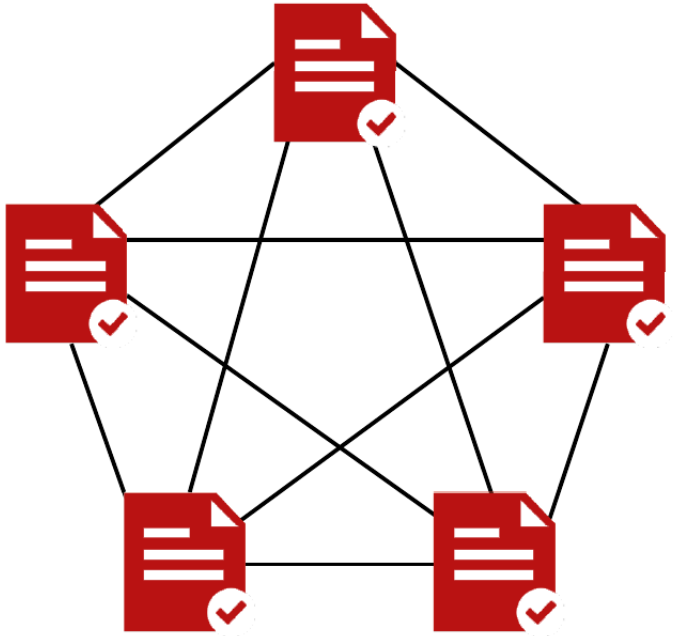
- **Homogeneous Networks** contain a single type of nodes and edge.
- **Heterogeneous Networks** contain multiple types of nodes or edges.
- **Hierarchical Networks**, whose various nodes and edges form set-subset relationships.

Basis of propagation-based fake news detection approaches

- **News cascades (propagation trees)** – a *direct* way to present news propagation
- **Self-defined graphs (networks)** – an *indirect* way to present news propagation

Homogeneous Network

Illustrations of homogeneous networks



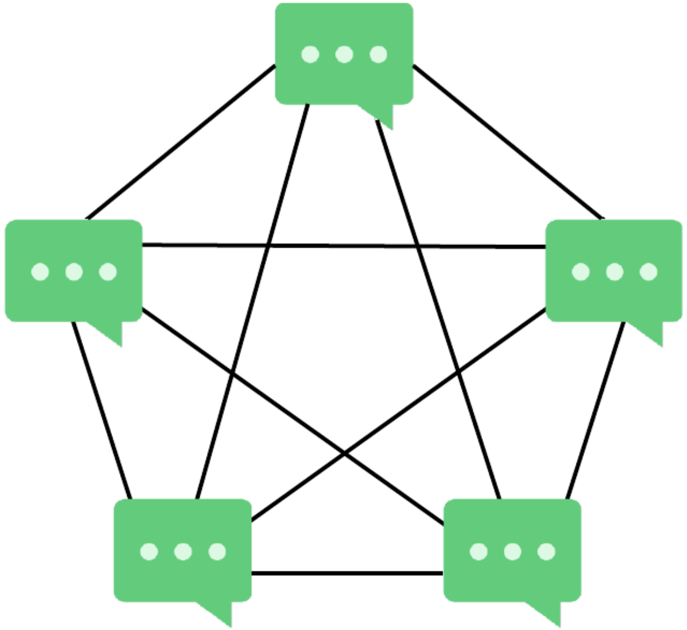
Stance Network

News article

Similarity of text, stance, topic, etc.

Homogeneous Network

Illustrations of homogeneous networks



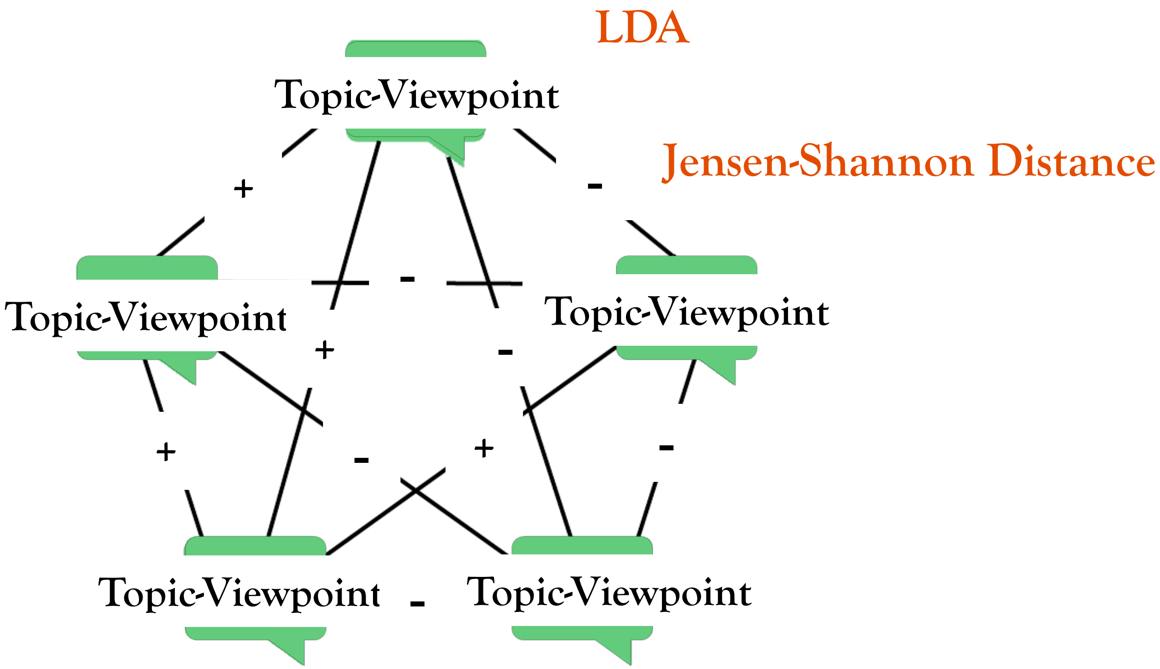
Stance Network

💬 User post

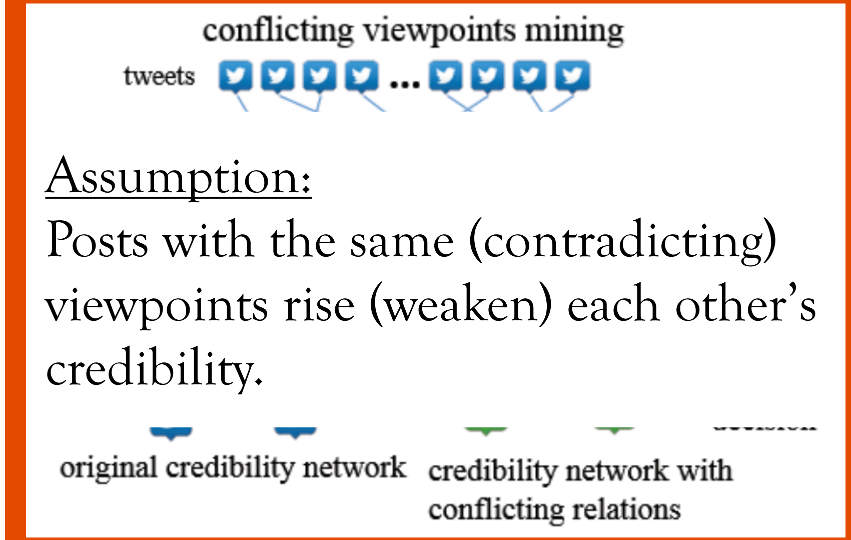
— Similarity of text, stance, topic, etc.

Homogeneous Network

Illustrations of related studies



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


conflicting viewpoints mining
tweets
original credibility network
credibility network with conflicting relations

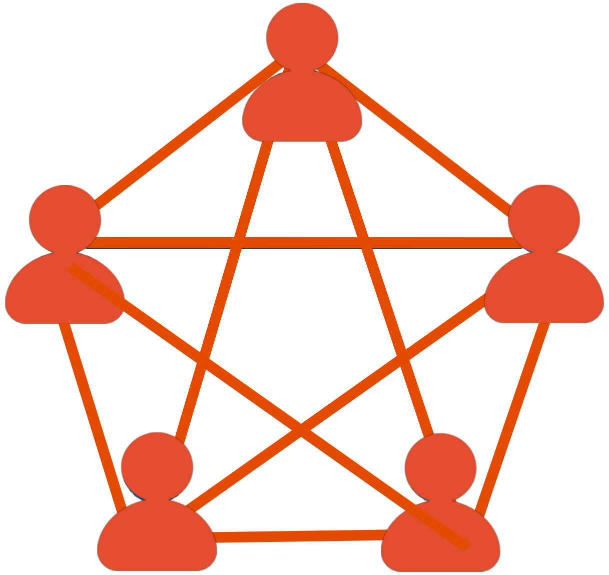
Assumption:

Posts with the same (contradicting) viewpoints rise (weaken) each other's credibility.

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

Homogeneous Network

Illustrations of homogeneous networks



Friendship Network

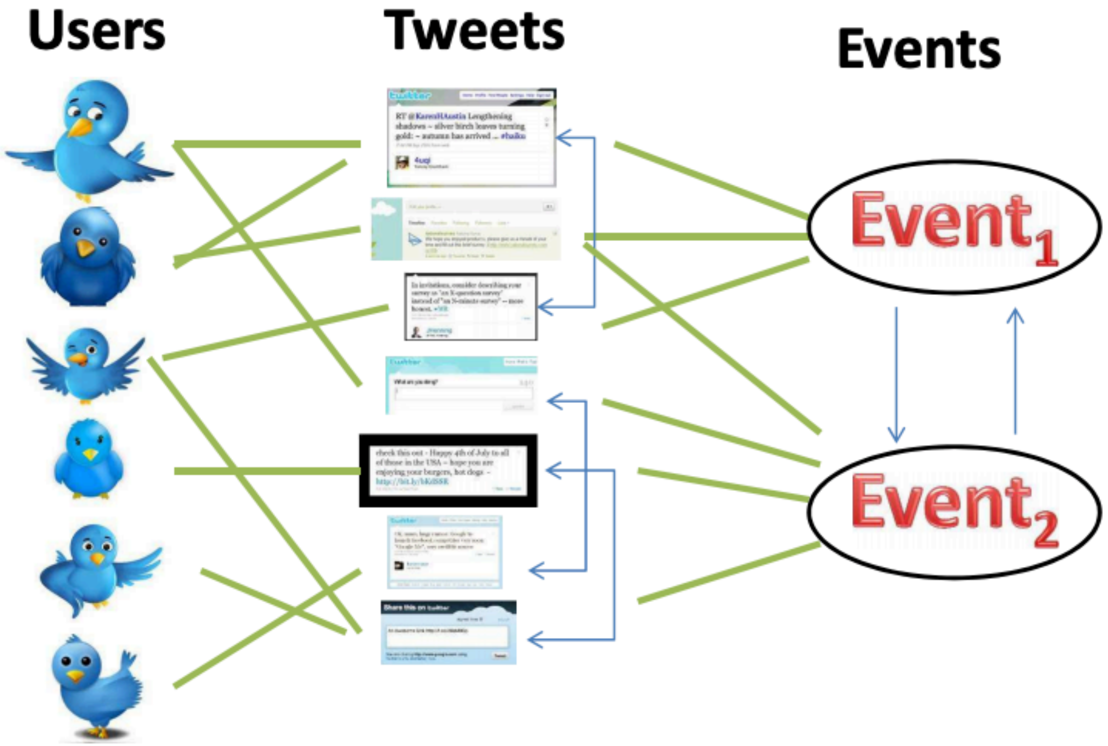
👤 User/Spreader

— Friend relationship

X. Zhou and R. Zafarani, Fake News in Networks: Patterns, Representation and Detection.

Heterogeneous Network

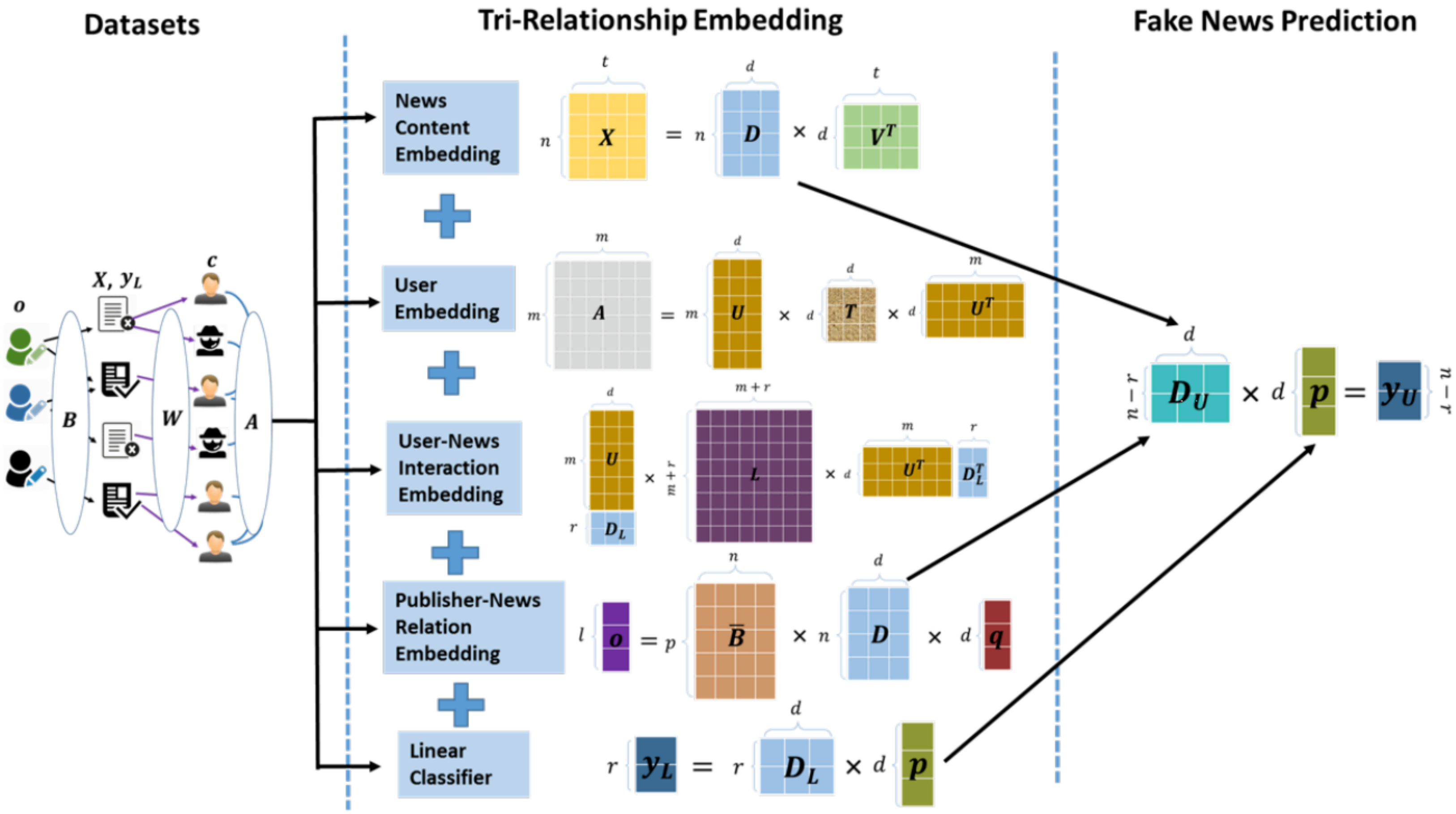
Illustrations of homogeneous networks and related studies



Assumption:

- Credible user → Credible tweets
- Average credibility of tweets: Credible events > Incredible events

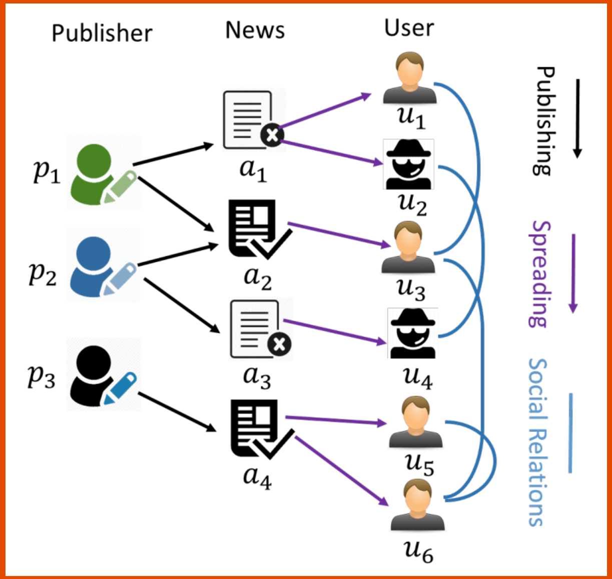
M. Gupta, et al. Evaluating Event Credibility on Twitter, SDM'12



Datasets
Tri-Relationship Embedding
Fake News Prediction
o
X, y_L
c
B
W
A
News Content Embedding
User Embedding
User-News Interaction Embedding
Publisher-News Relation Embedding
Linear Classifier
$X = D \times V^T$
$A = U \times T \times U^T$
$U, D_L \times L \times U^T, D_L^T$
$o = \bar{B} \times D \times q$
$y_L = D_L \times p$
$D_U \times p = y_U$
n, t, d, m, r, l, p, $m+r$, $n-r$

Heterogeneous Network

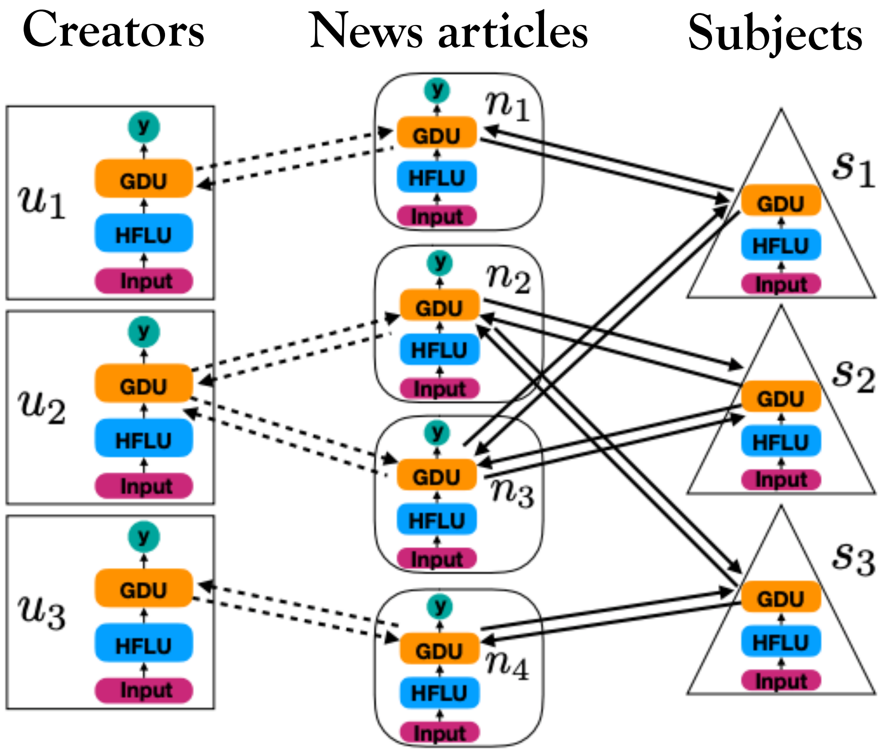
Illustrations of homogeneous networks and related studies



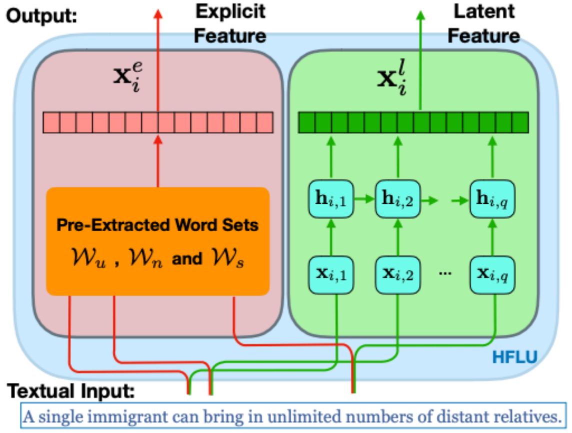
K. Shu, et al. Beyond News Contents: The Role of Social Context for Fake News Detection, WSDM'19.

Heterogeneous Network

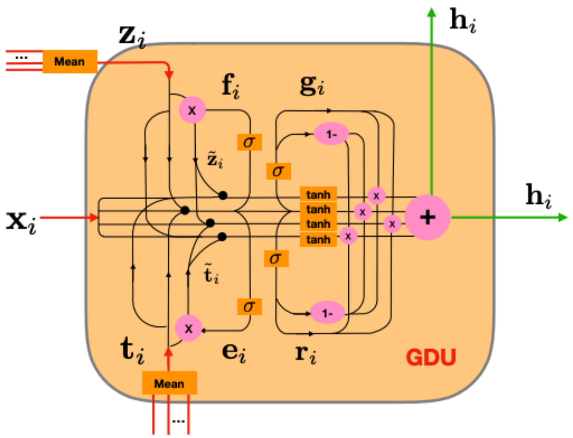
Illustrations of homogeneous networks and related studies



(c) Framework Architecture.

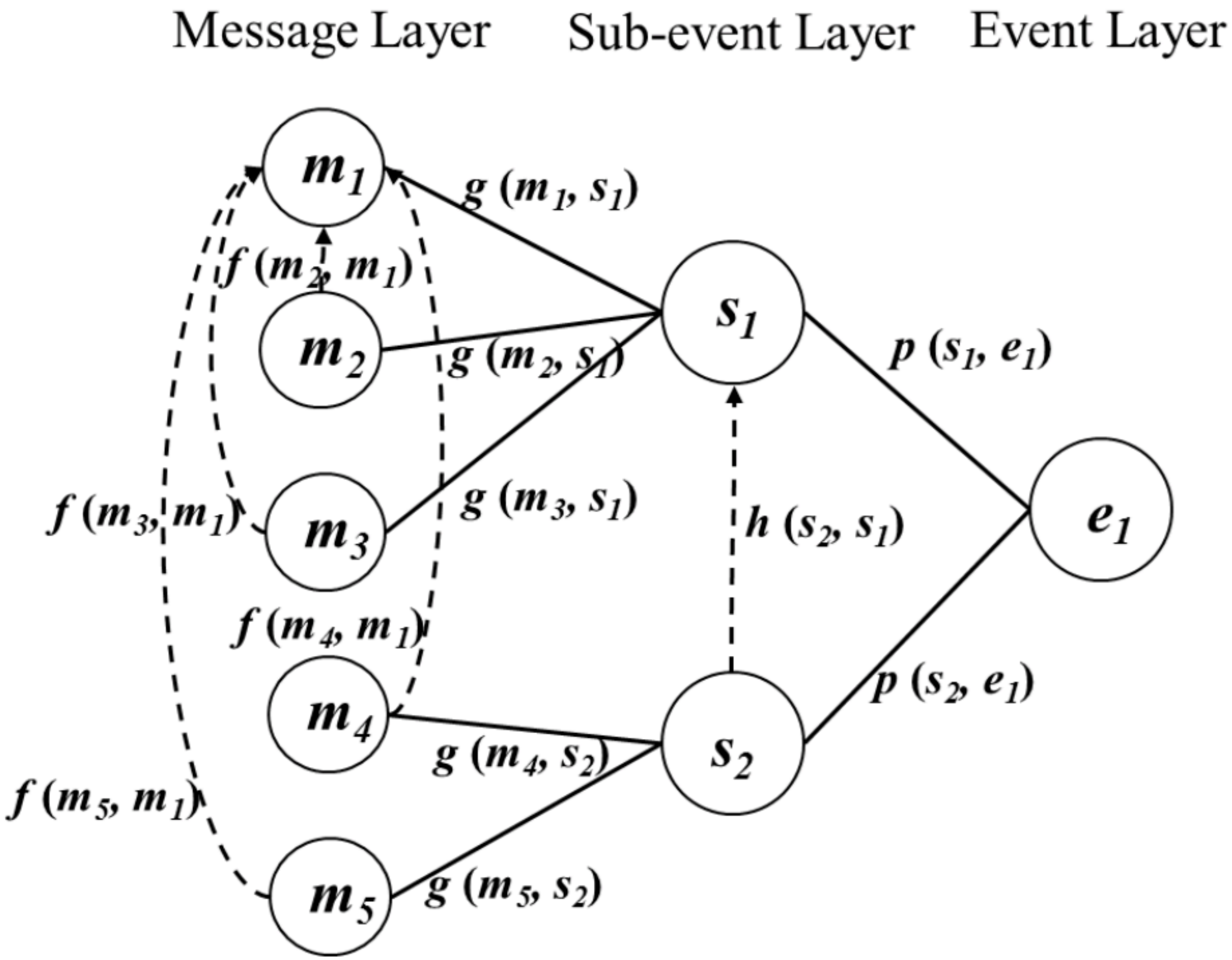


(a) Hybrid Feature Learning Unit (HFLU).



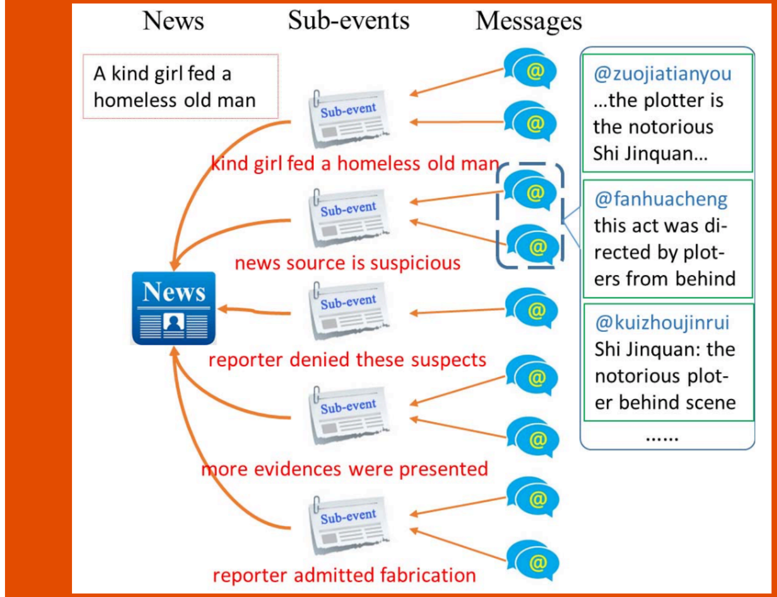
(b) Gated Diffusive Unit (GDU).

J. Zhang, et al. Fake News Detection with Deep Diffusive Network Model, arXiv: 1805.08751, 2018



Hierarchical Network

Illustrations of hierarchical networks and related studies


News
Sub-events
Messages
A kind girl fed a homeless old man
Sub-event
kind girl fed a homeless old man
news source is suspicious
News
reporter denied these suspects
more evidences were presented
reporter admitted fabrication
@

@zuojiatianyou …the plotter is the notorious Shi Jinquan…

@fanhuacheng this act was directed by plotters from behind

@kuizhoujinrui Shi Jinquan: the notorious plotter behind scene

……

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

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

Credibility-based Fake News Detection

Overview

Credibility-based Fake News Detection also involve social context information

- Credibility of entities, e.g., **news headlines**, **comments** and **spreaders**
- Relationships among the credibility of the same or different entities

Overlaps with propagation-based fake news detection

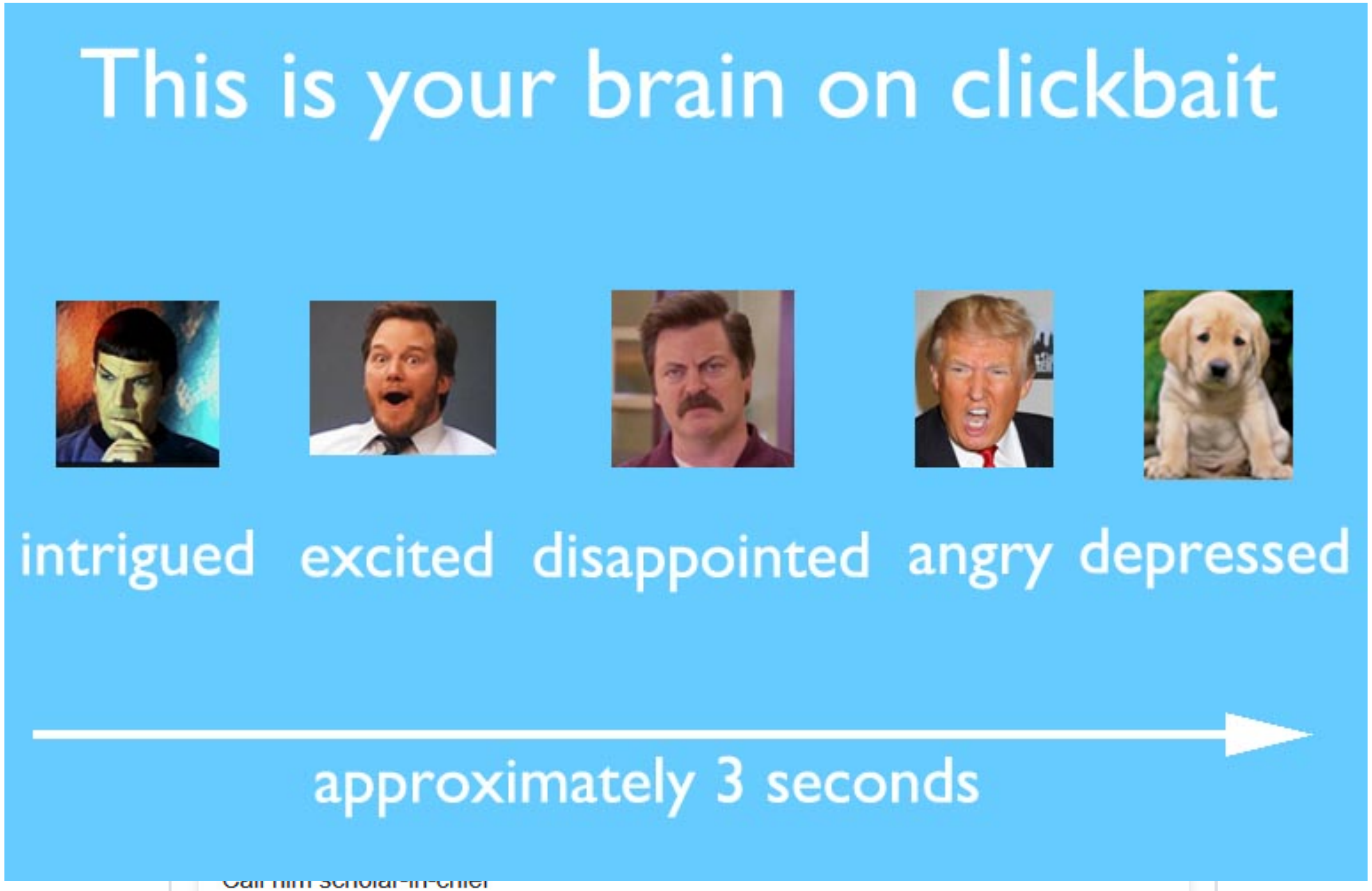
Clickbait detection

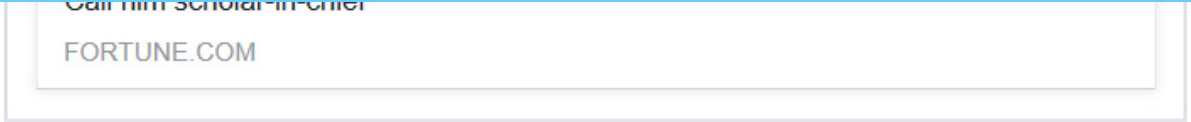
Review spam(mer) detection

Bot detection;

News Headline Credibility

Clickbait




Call him scholar-in-chief

FORTUNE.COM

Clickbait is headlines whose main purpose is to attract the attention of visitors and encourage them to click on a link to a particular web page.

News Headline Credibility

Clickbait & Fake News

When news articles meet clickbait:

- Attract eyeballs but are rarely newsworthy
- Increase click rate and **further gain the public trust**

Term	Phenomenon
Attentional bias	**Exposure frequency** - individuals tend to believe information is correct after repeated exposures.
Validity effect	
Echo chamber effect	

News Headline Credibility

By detecting clickbait

Feature engineering within a supervised machine learning framework[8]

- N-gram and POS tags → Structure-based style features
- Informality, readability and immediacy → Attribute-based style features
- **Similarity between news headline and body-text**

News with clickbait < News without clickbait

Deep clickbait detection

[8] P. Biyani, et al., "8 Amazing Secrets for Getting More Clicks": Detecting Clickbaits in News Streams Using Article Informality . AAAI'16

News Comment Credibility

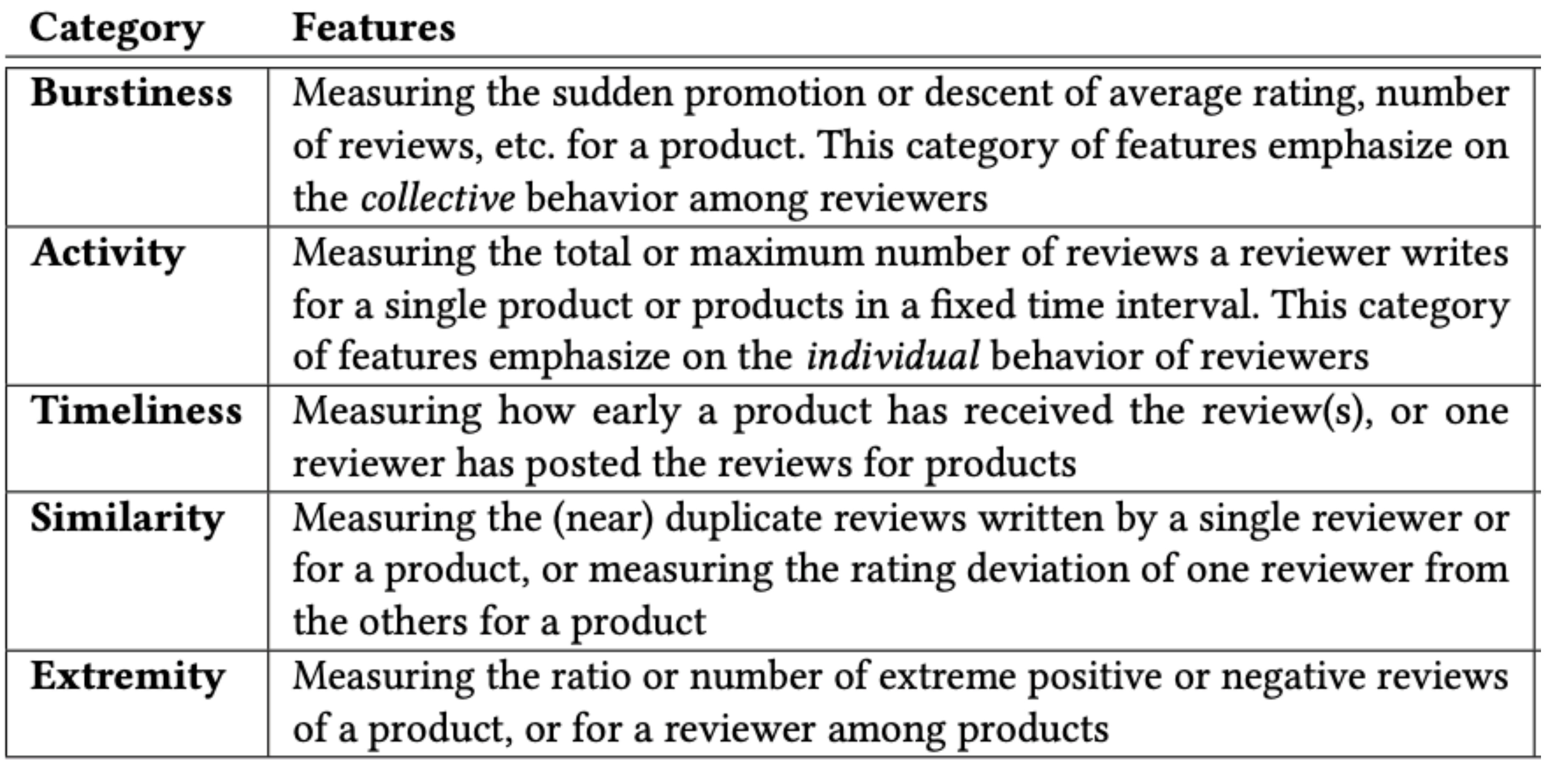
Review Spam Detection

- **Content-based / Style-based** models
- **Behavior-based** models
- **Graph-based** models

News Comment Credibility

Review Spam Detection

- **Content-based** / **Style-based**
- **Behavior-based** models
- **Graph-based** models

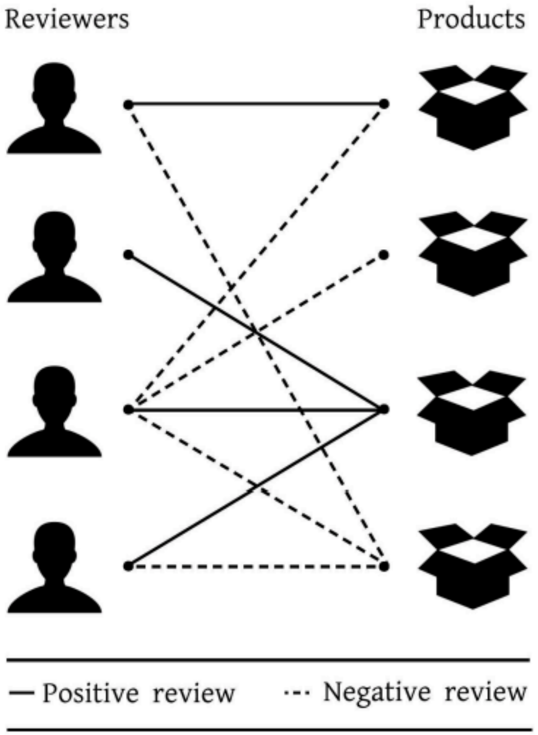


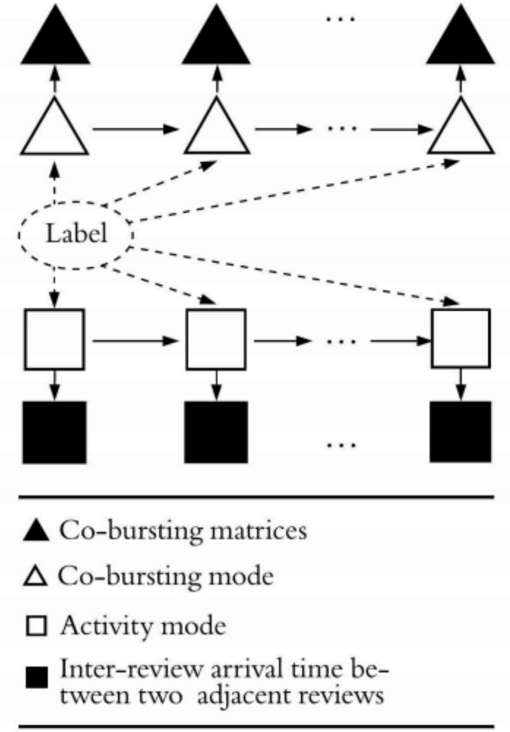
Category	Features
Burstiness	Measuring the sudden promotion or descent of average rating, number of reviews, etc. for a product. This category of features emphasize on the *collective* behavior among reviewers
Activity	Measuring the total or maximum number of reviews a reviewer writes for a single product or products in a fixed time interval. This category of features emphasize on the *individual* behavior of reviewers
Timeliness	Measuring how early a product has received the review(s), or one reviewer has posted the reviews for products
Similarity	Measuring the (near) duplicate reviews written by a single reviewer or for a product, or measuring the rating deviation of one reviewer from the others for a product
Extremity	Measuring the ratio or number of extreme positive or negative reviews of a product, or for a reviewer among products

News Comment Credibility

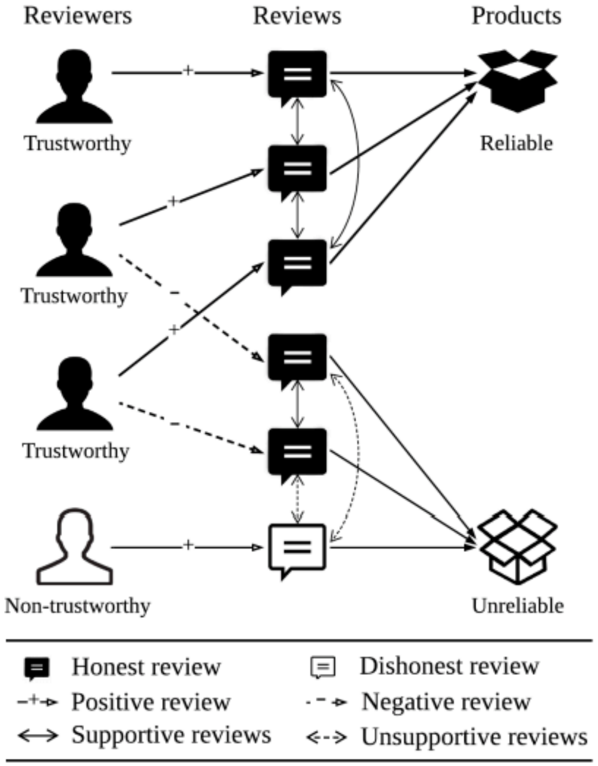
Review Spam Detection

- **Content-based** / **Style-ba**
- **Behavior-based** models
- **Graph-based** models





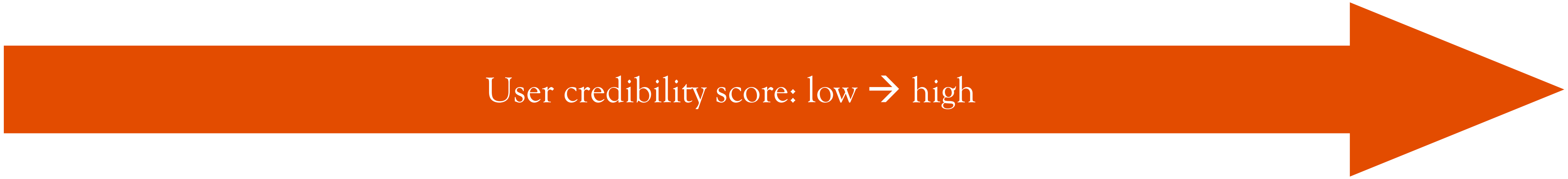
Probabilistic Graphical Models



Web ranking algorithm

News Spreader Credibility

User Classification



User credibility score: low → high

Malicious users

- **Intentionally** engage in fake news activities

Susceptible users

- **Unintentionally** engage in fake news activities

Insusceptible users

- Immune to fake news

News Spreader Credibility

Why normal users can unintentionally engage in spreading fake news?

	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	
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	

Social Influence → How widely the news article has been spread?

Self-influence → What preexisting knowledge a user has?

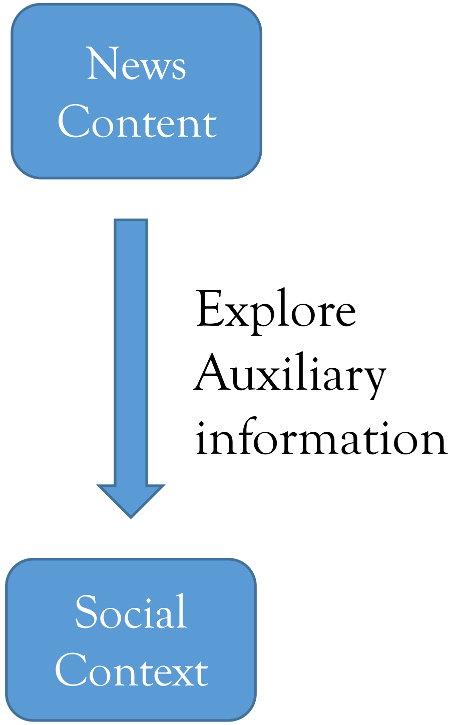
Beyond News Contents:
The Role of Social Context for Fake News Detection

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

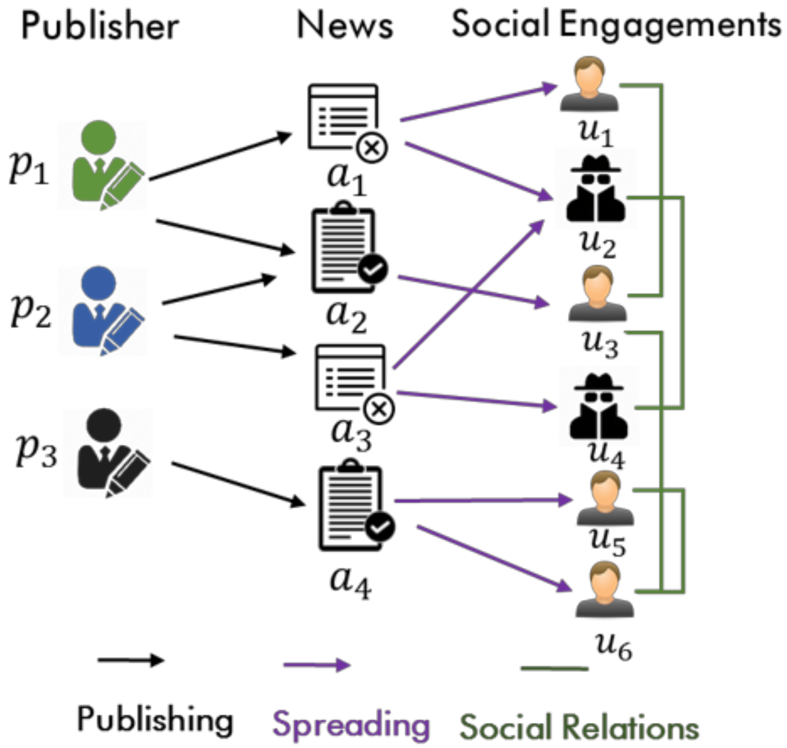


Fake News Detection – Multi-Source

- 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., $p_1 \rightarrow a_1 \quad p_2 \rightarrow a_3$
$p_3 \rightarrow a_4$



➢ *spreading*

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

➢ *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

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

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

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

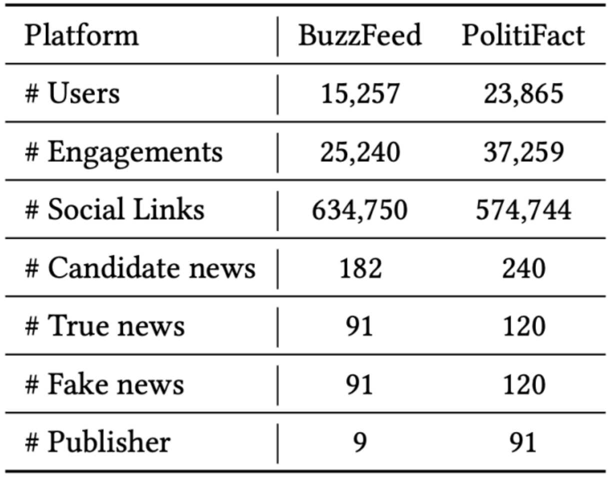
$$\min \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{r} \mathbf{W}_{ij}\mathbf{c}_i\Big(1 - \frac{1+\mathbf{y}_{Lj}}{2}\Big)\|\mathbf{U}_i - \mathbf{D}_{L_j}\|_2^2}_{\text{True news}}$$
$$+ \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{r} \mathbf{W}_{ij}(1 - \mathbf{c}_i)\Big(\frac{1+\mathbf{y}_{Lj}}{2}\Big)\|\mathbf{U}_i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

Evaluation Setting

- Datasets: FakeNewsNet with information for news content, social context and ground truth labels from fact-checking 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

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

News Content

Social Context

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

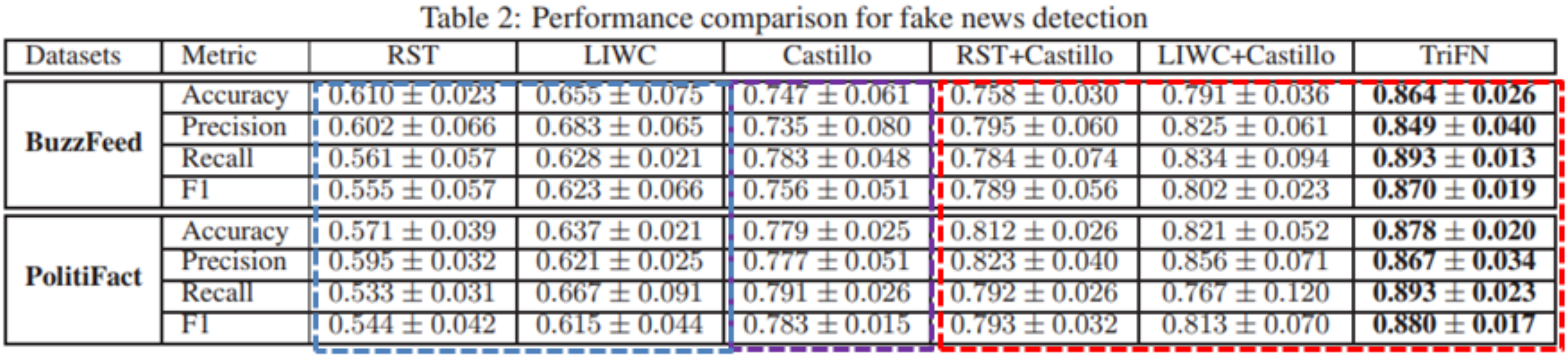


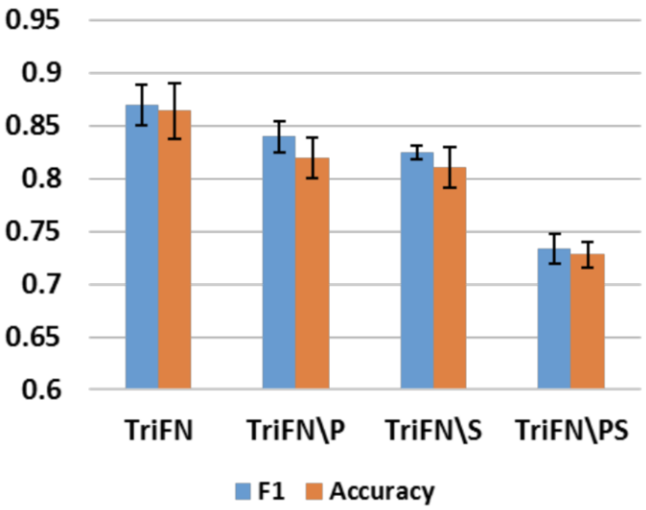
Table 2: Performance comparison for fake news detection

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

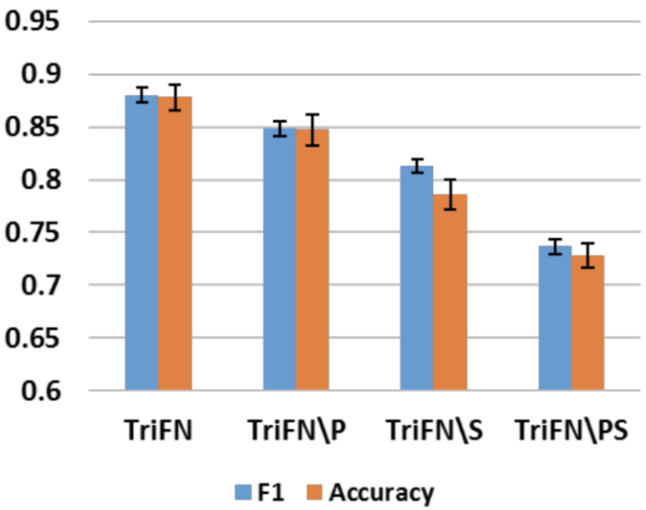
News Content Social Context News Content + Social Context

Evaluation Results - Component Analysis and Early Detection

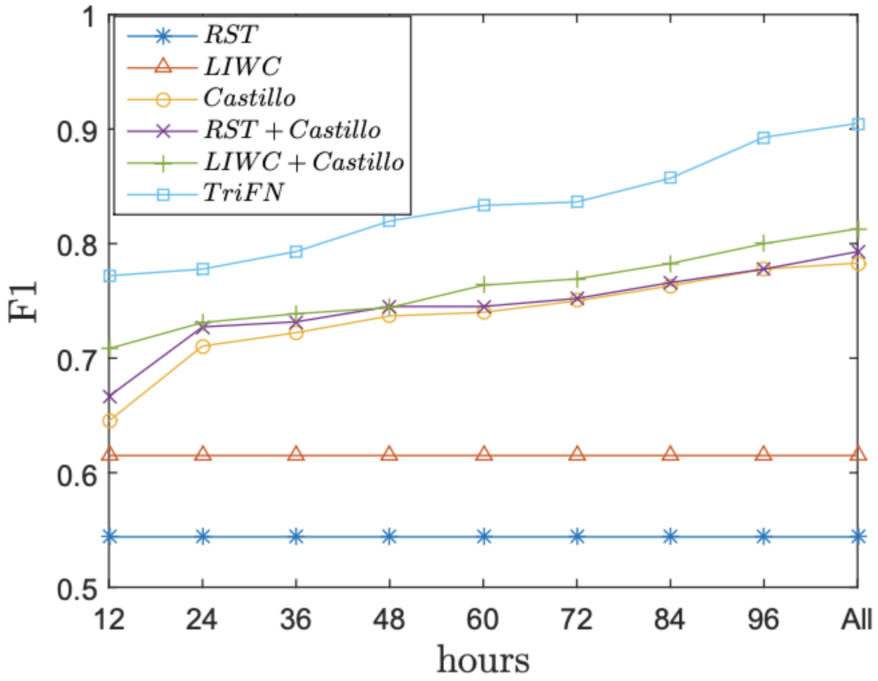
- Both publisher-news and news-user relations can contribute to the performance improvement of TriFN
- TriFN consistently achieves best performances in the early stage of news dissemination



(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

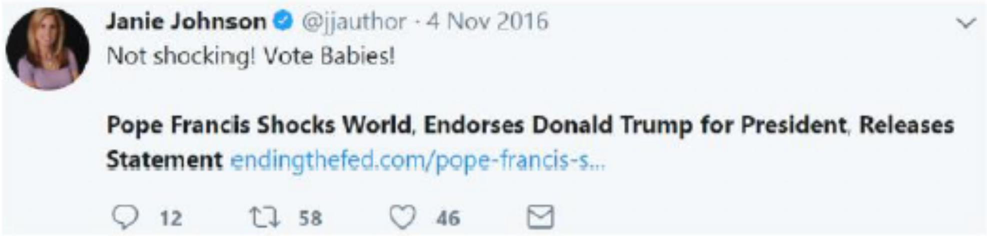
Unsupervised Fake News Detection: A Generative Approach

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

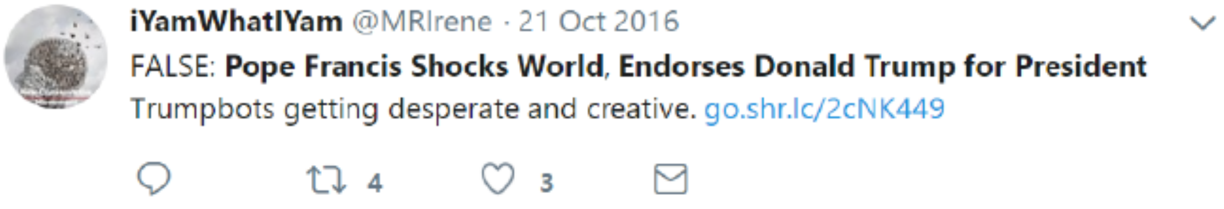


Janie Johnson @jjauthor · 4 Nov 2016

Not shocking! Vote Babies!

Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement endingthefed.com/pope-francis-s...

Agreeing the authenticity of the news



iYamWhatIYam @MRIrene · 21 Oct 2016

FALSE: **Pope Francis Shocks World, Endorses Donald Trump for President**

Trumpbots getting desperate and creative. go.shr.lc/2cNK449

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

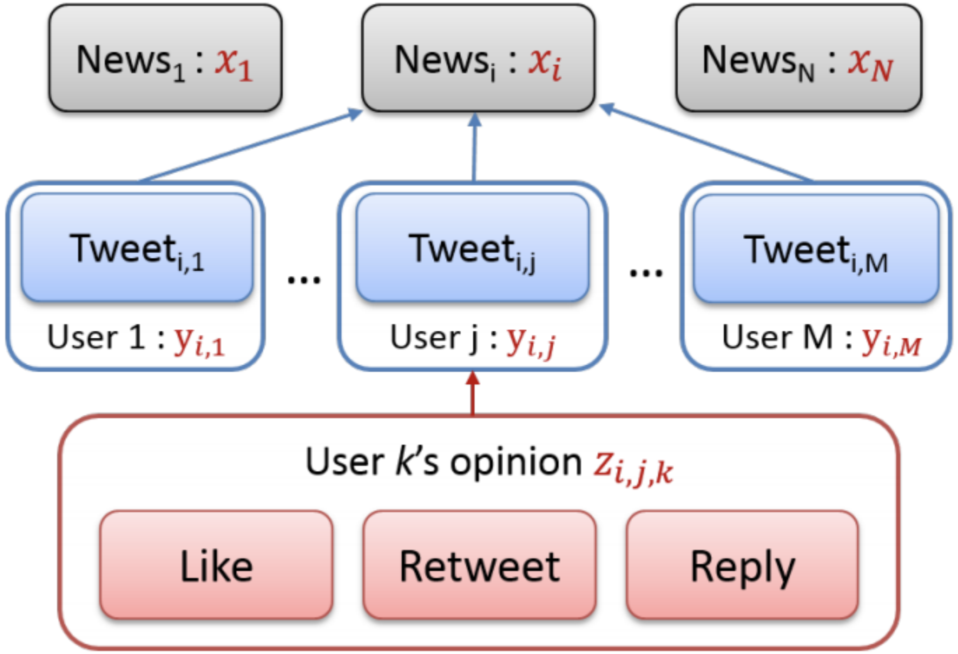
- The relationships among news, tweets, and users on social media form more complicated topologies
- Existing truth discovery methods mainly focus on “source-item” paths, and cannot be directly applied

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 to j $news_i$
  - ○ $z_{i,j,k}$ 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



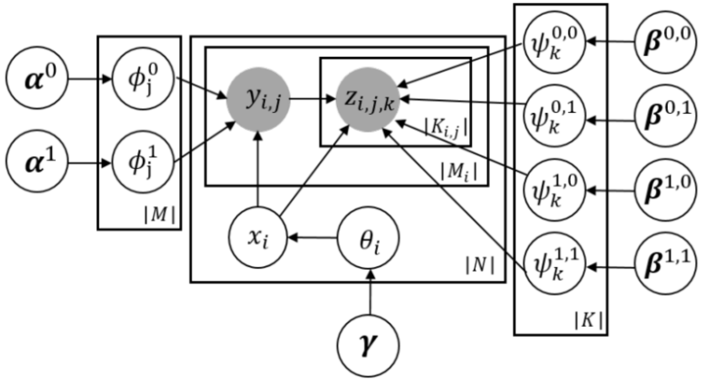
The Proposed Probabilistic Model (UFD)

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

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

- For verified user j $y_{i,j} \sim \text{Bernoulli}(\phi_j^{x_i})$
  - ϕ_j^1 (ϕ_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, $z_{i,j,k} \sim \text{Bernoulli}(\psi_k^{x_i,y_{i,j}})$
  - 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

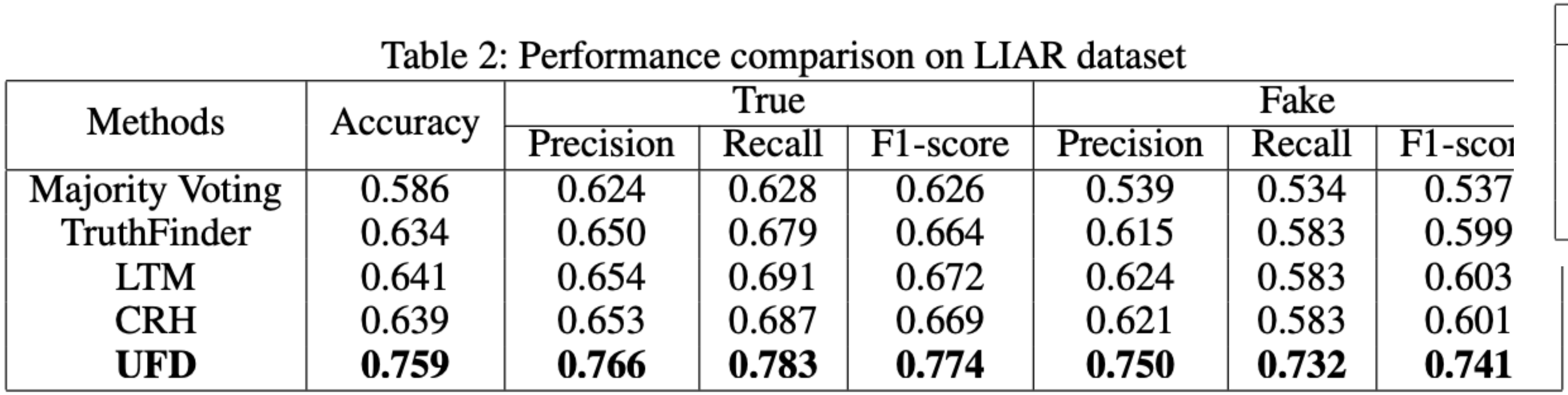


Table 2: Performance comparison on LIAR dataset

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**

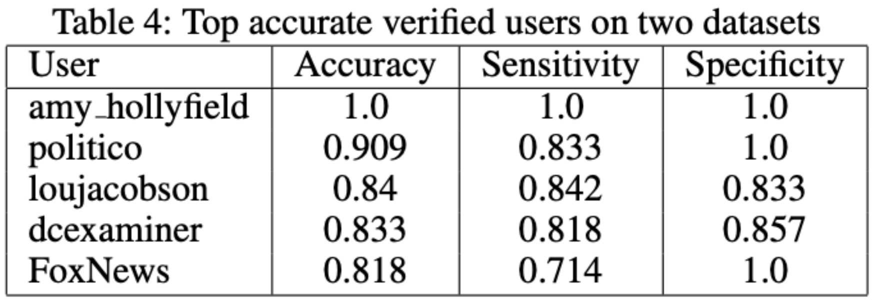


Table 4: Top accurate verified users on two datasets

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

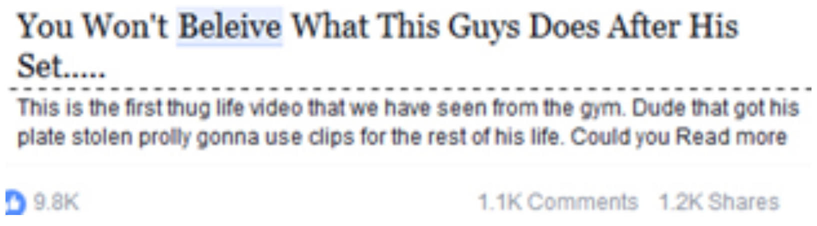
Deep Headline Generation for Clickbait Detection

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

ICDM 2018

Clickbaits

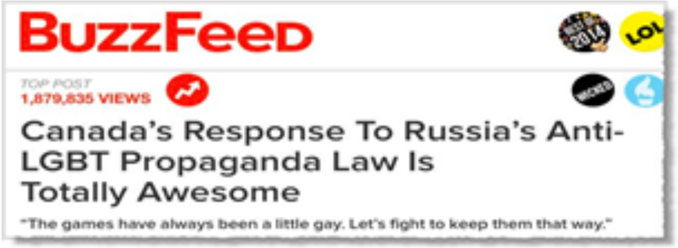
- Clickbaits are catchy social media posts or sensational headlines that attempt to lure the readers to click



You Won't Beleive What This Guys Does After His Set.....

This is the first thug life video that we have seen from the gym. Dude that got his plate stolen prolly gonna use clips for the rest of his life. Could you Read more

9.8K 1.1K Comments 1.2K Shares



BuzzFeed

TOP POST
1,879,835 VIEWS

Canada's Response To Russia's Anti-LGBT Propaganda Law Is Totally Awesome

"The games have always been a little gay. Let's fight to keep them that way."

- 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

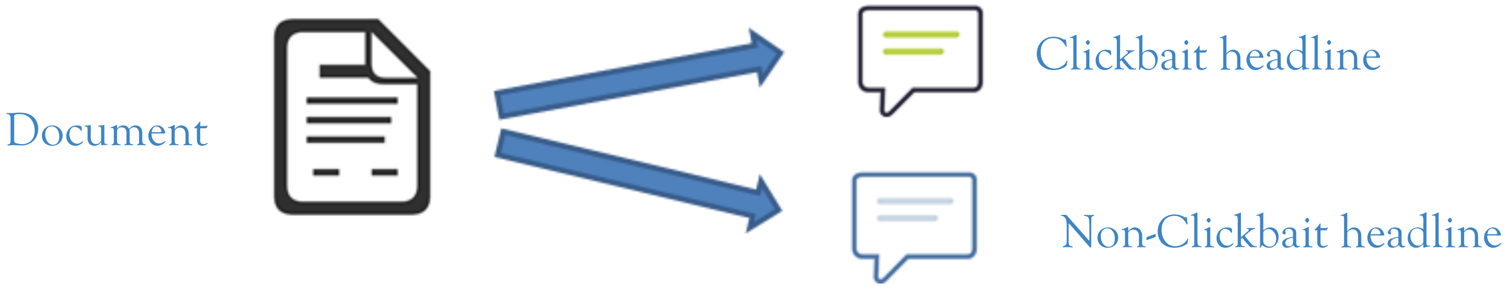
Clickbait Detection

- 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

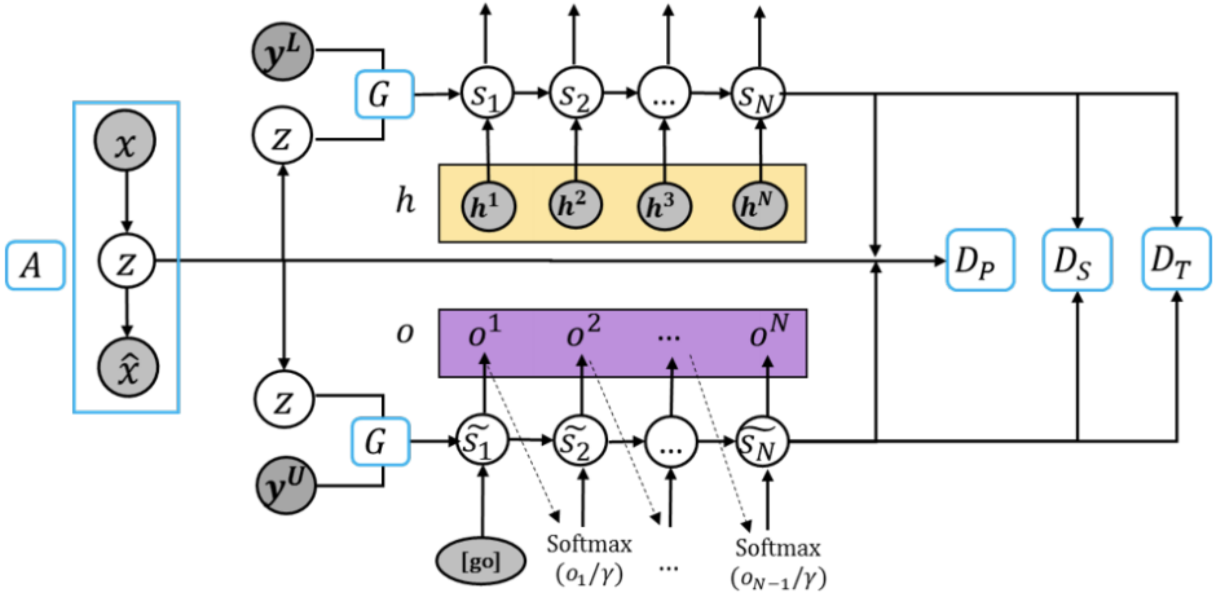
Problem Definition

- 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)

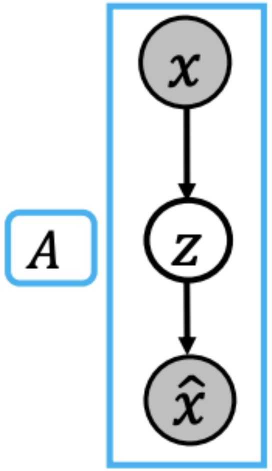
- We propose a deep learning model to generate both click-baits and non-clickbaits with style transfer
  - Generator Learning: a document autoencoder A , a headline generator G
  - Discriminator Learning: a transfer discriminator D_T , a style discriminator D_S , a pair discriminator D_P



Generator Learning

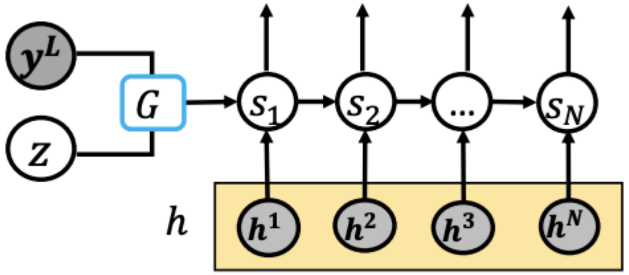
- Document autoencoder $\mathcal{A}$ extract document representation by minimizing the reconstruction error

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

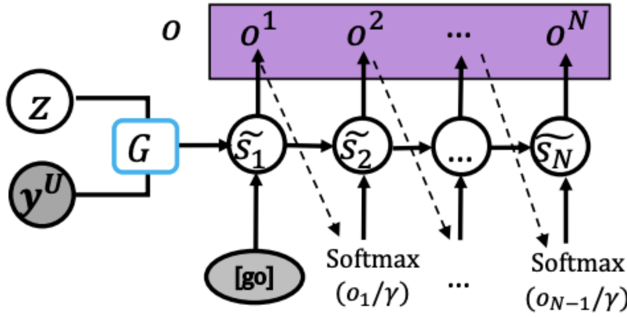


- Headline generator G
  - Generate stylized headline by minimizing the reconstructio error of original headline

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



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



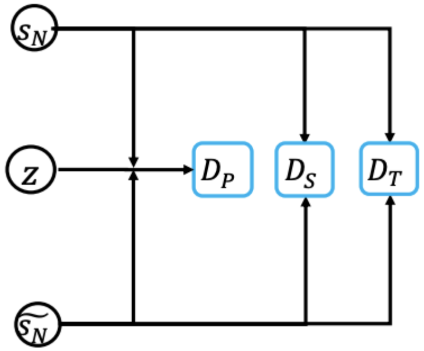
Discriminator Learning

- 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

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

Discriminator Learning

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

Proximity function

$$p(h_i, x_j) = \frac{1}{1 + \exp(-\mathbf{s}^{(i)}\mathbf{Q}\mathbf{z}^{(j)})}$$

Document representation

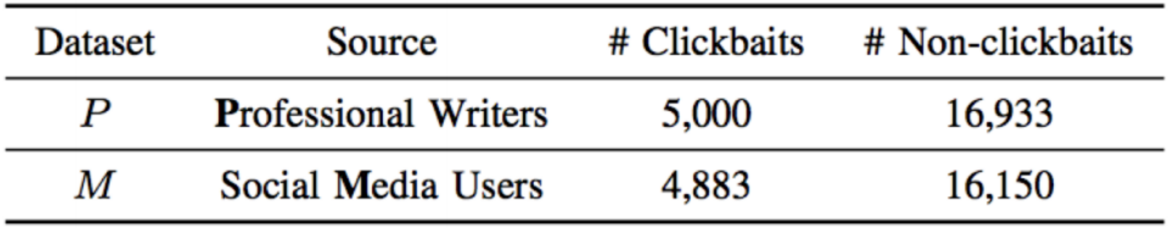
Headline representation

- 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

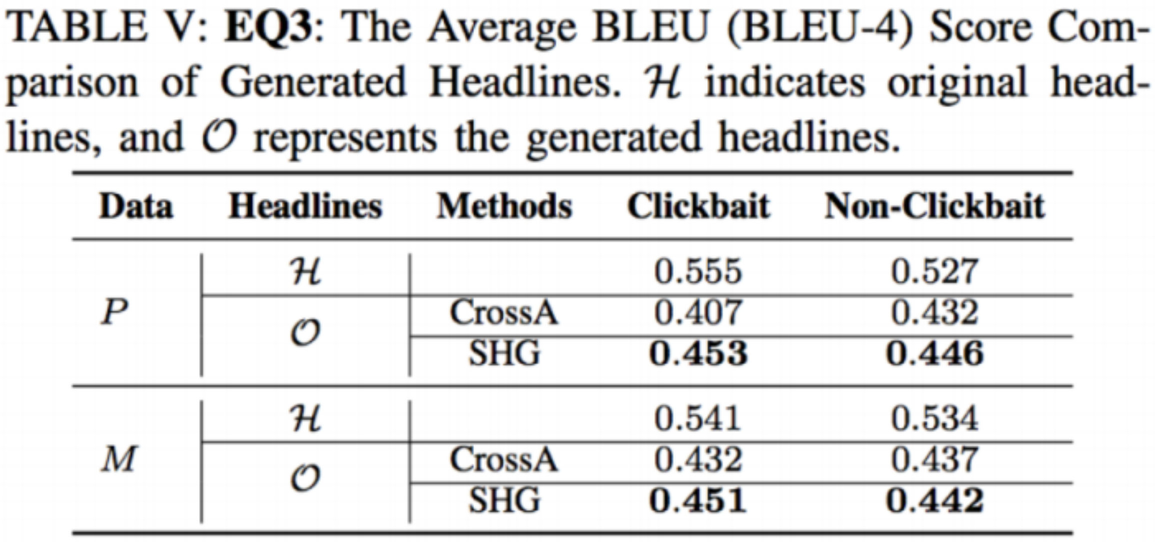


TABLE V: **EQ3**: The Average BLEU (BLEU-4) Score Comparison of Generated Headlines. $\mathcal{H}$ indicates original headlines, and $\mathcal{O}$ represents the generated headlines.

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**

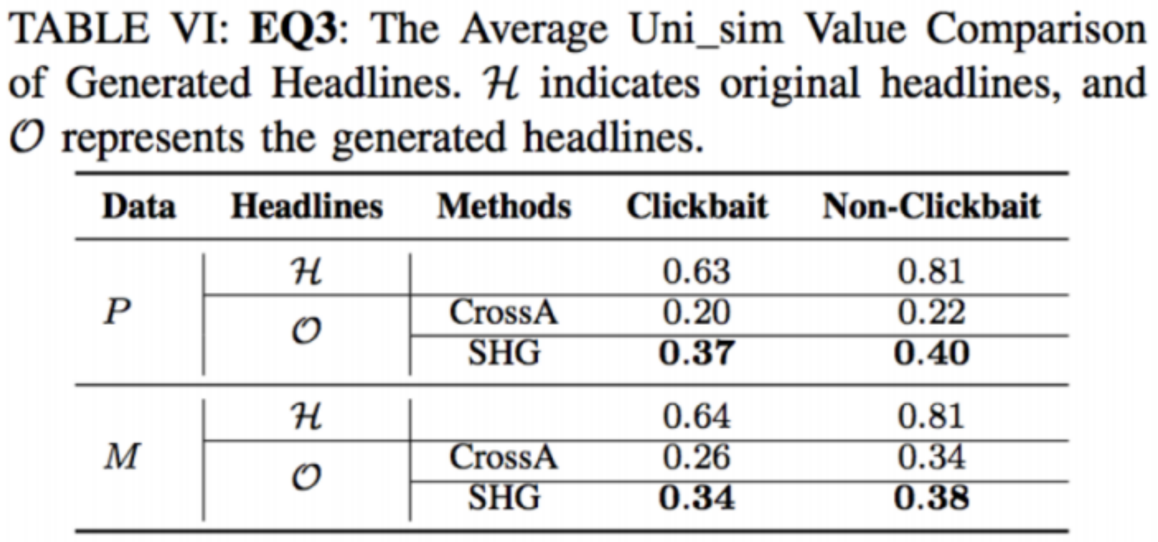
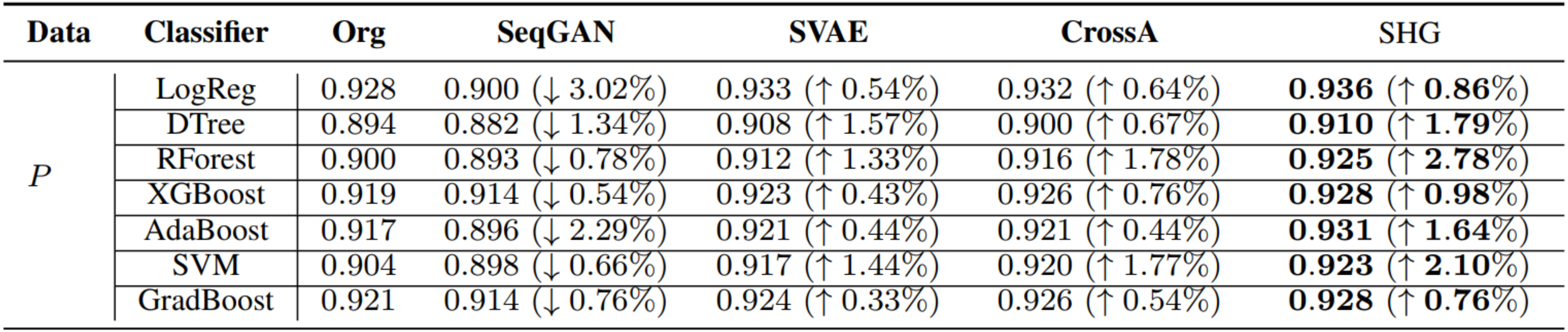


TABLE VI: **EQ3**: The Average Uni_sim Value Comparison of Generated Headlines. $\mathcal{H}$ indicates original headlines, and $\mathcal{O}$ represents the generated headlines.

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

Summary and Comparison for Fake News Detection

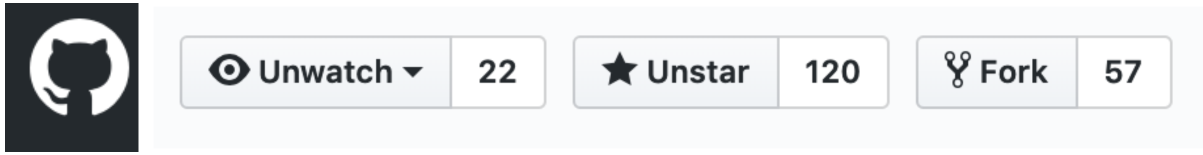
	Knowledge-based **fake news detection**	Style-based **fake news detection**	Propagation-based **fake news detection**	Credibility-based **fake news detection**
Information Utilized	News content		News content & Social context information	
Techniques	Graph models	Feature-based methods	Graph models & Feature-based methods	
Resources	Knowledge graphs	Fundamental theories		
Related Topic(s)	Fact-checking	Deception detection	Rumor detection	Clickbait/bot/review spam detection

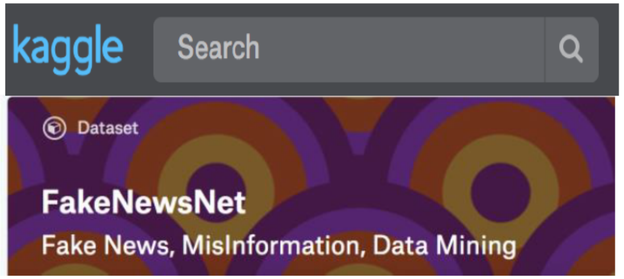
Fake News Detection

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

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

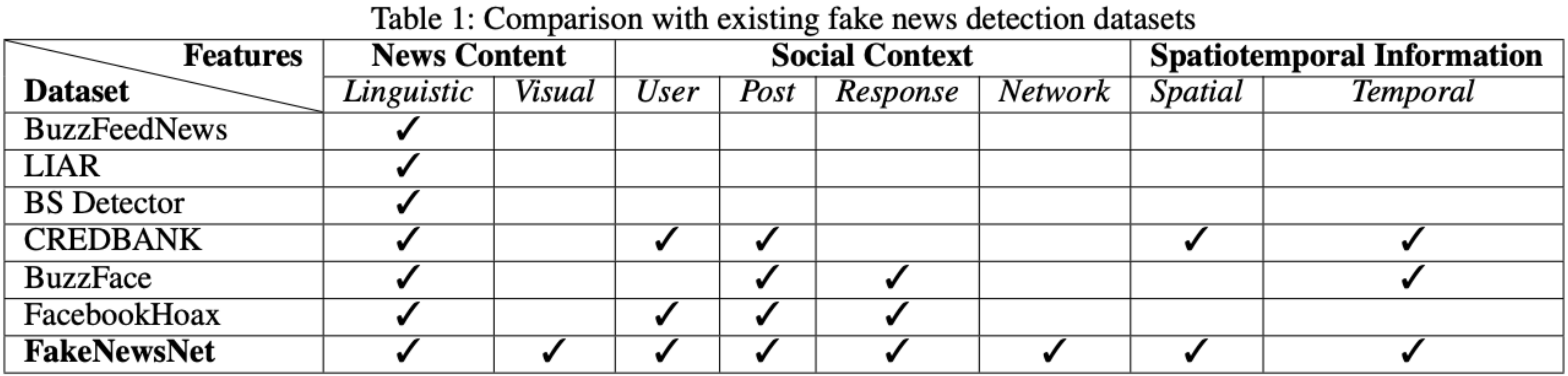
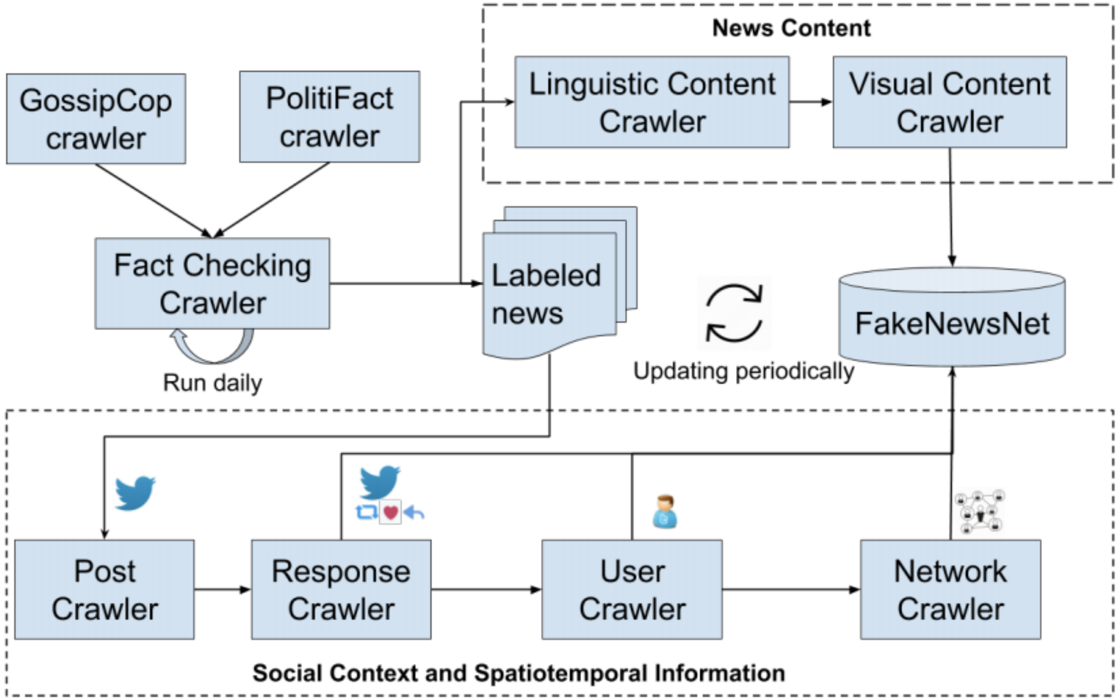


Table 1: Comparison with existing fake news detection datasets

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

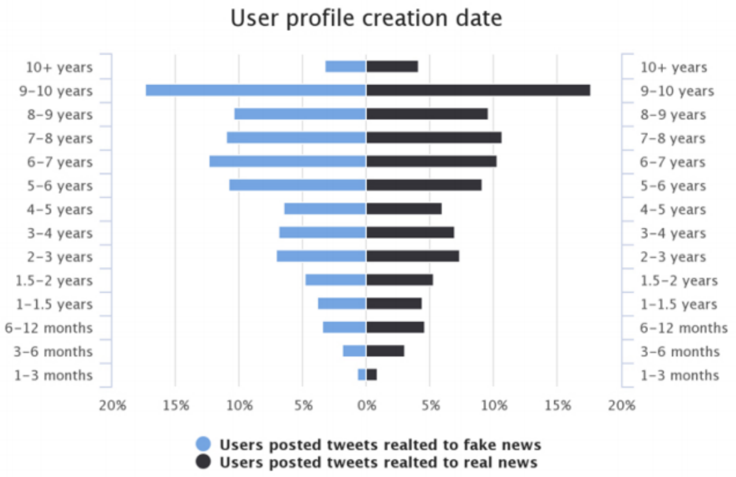
Data Integration

- **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 info and temporal data from meta data of Twitter

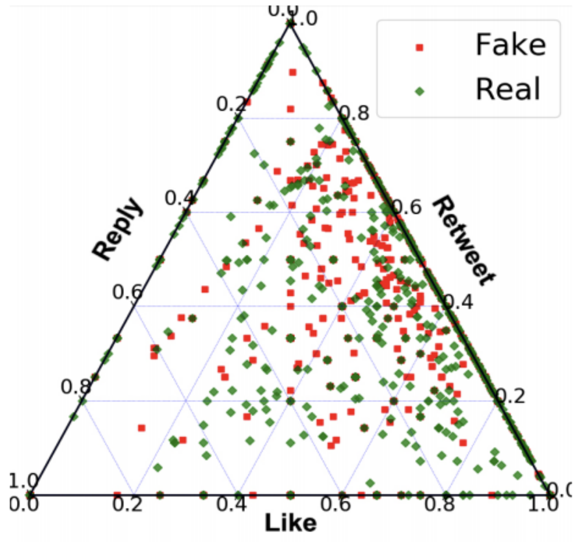


Data Analysis

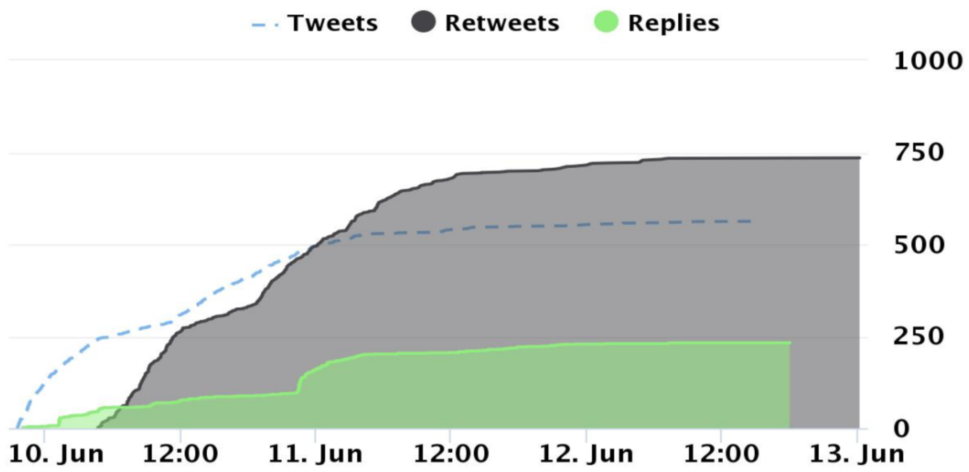
- **User profiles:** users who share real news pieces tend to have longer register time than those who share the fake news on average



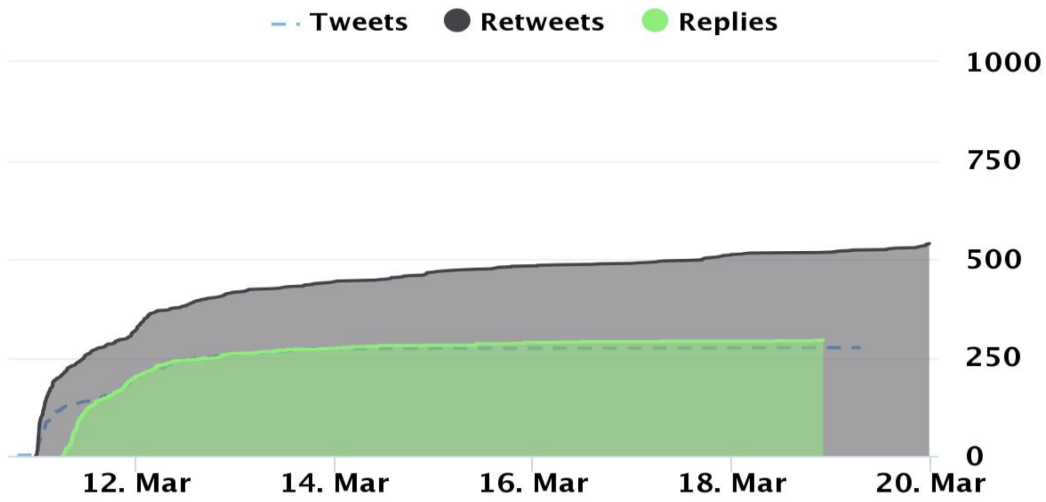
- **User engagements:** fake news pieces tend to have fewer replies and more retweets; real news pieces have more ratio of likes than fake news pieces do



- **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

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

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

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

Kai Shu, Deepak Mahudeswaran, and Huan Liu

2018 SBP-BRiMS

Challenge Problem Winner

Disinformation Challenge

THIS CERTIFICATE IS AWARDED TO

FAKENEWSTRACKER: TOWARDS FAKE NEWS COLLECTION, DETECTION, AND VISUALIZATION

BY

KAI SHU, DEEPAK MAHUDESWARAN AND HUAN LIU

2018 International Conference on Social Computing,
Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation
July 10-13, 2018
George Washington University
Washington, DC 20052

SBP 2018

SBP Disinformation Challenge Winner

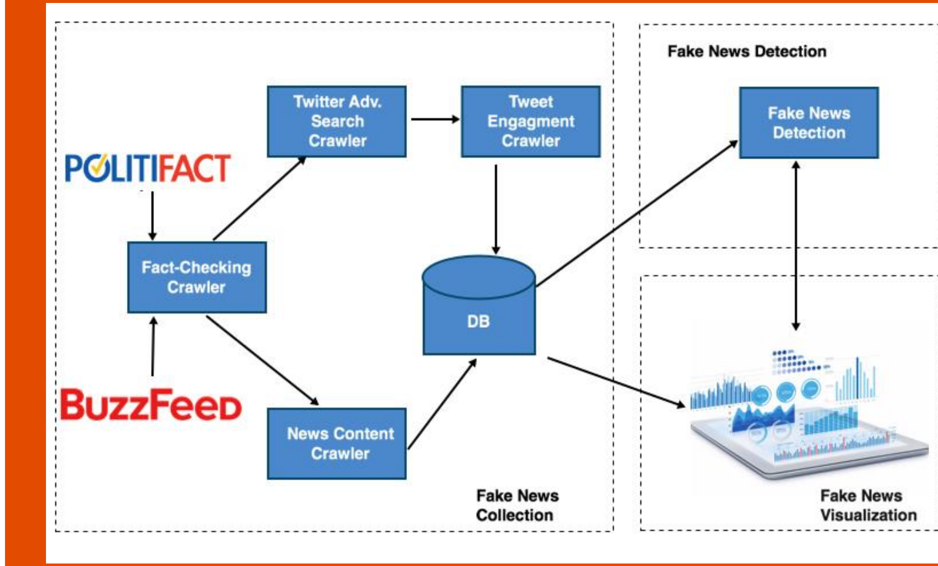
http://blogtrackers.fulton.asu.edu:3000

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

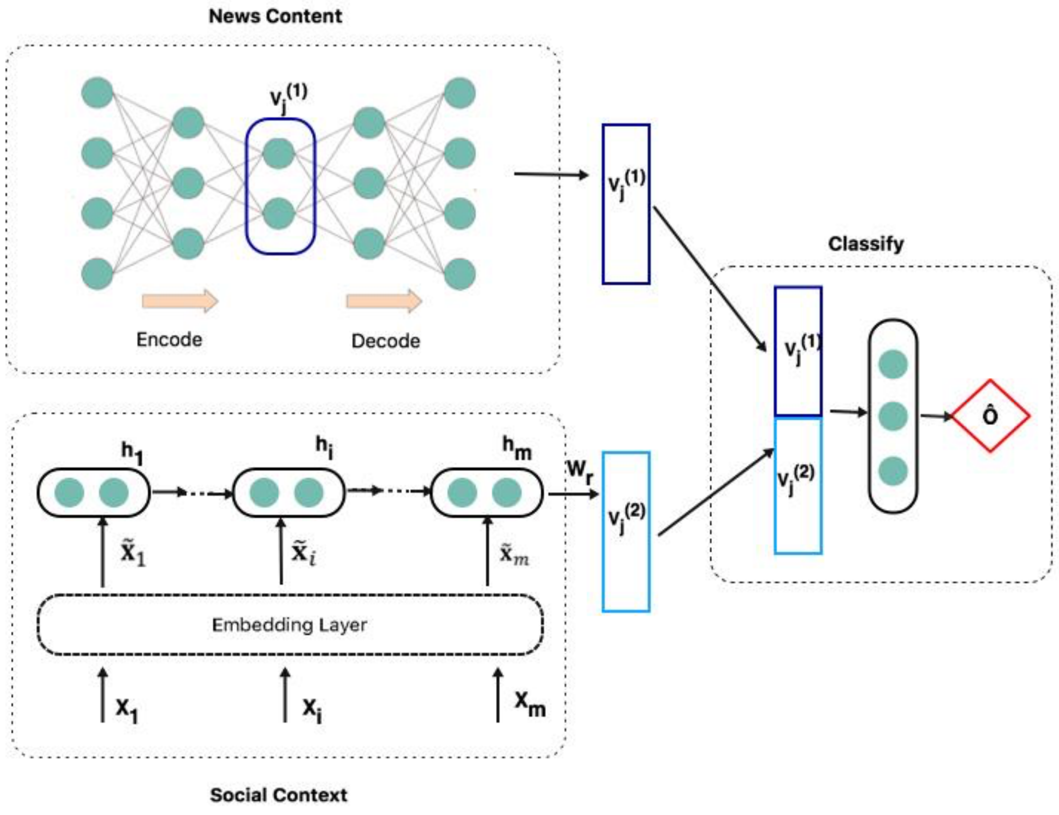
- **Fake News Detection:** finding fake news with advanced machine learning methods, such as deep neural networks

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



Fake News Detection

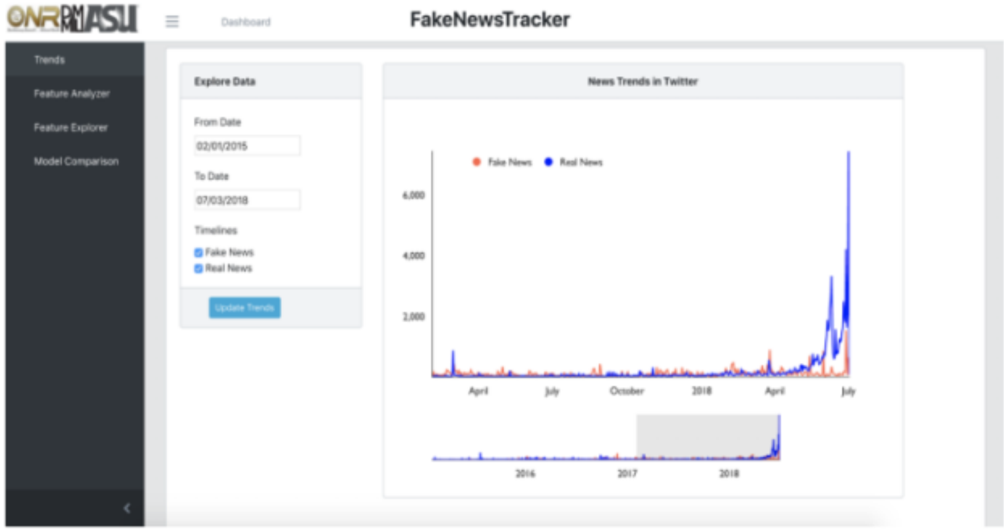
- Detect fake news with fusion of news content and social context
  - **News representation:** Represent news content using autoencoders
  - **Social engagement representation:** Represent social engagements using RNNs
  - **Social Article Fusion:** Combine both news and social engagement features to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



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

Fake News Visualization

Trends on Twitter

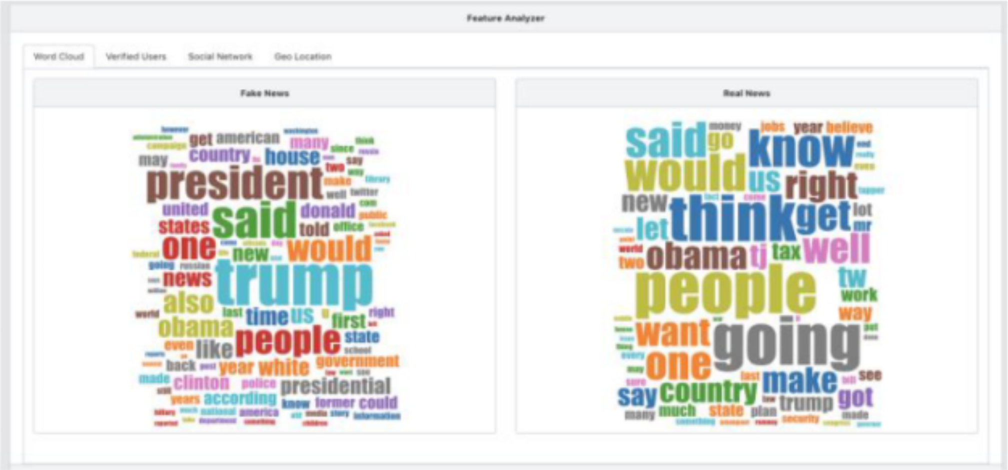


From Date
02/01/2015

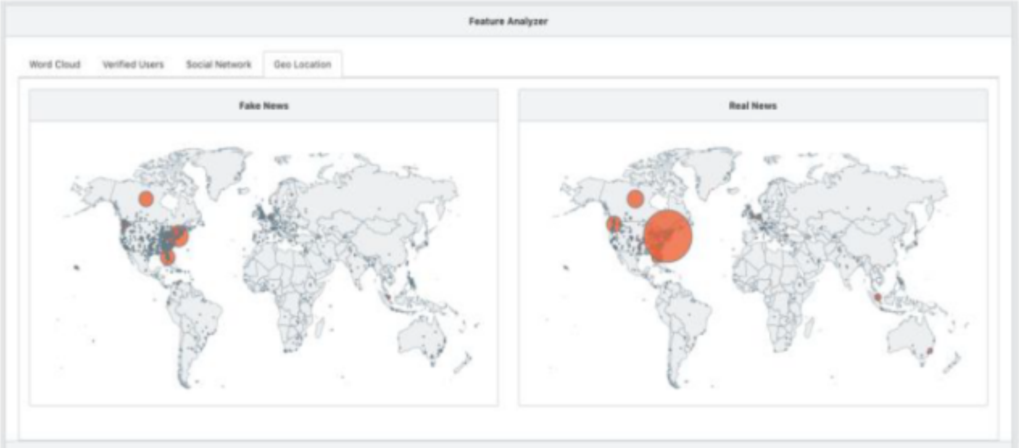
To Date
07/03/2018

Timelines
☑ Fake News
☑ Real News

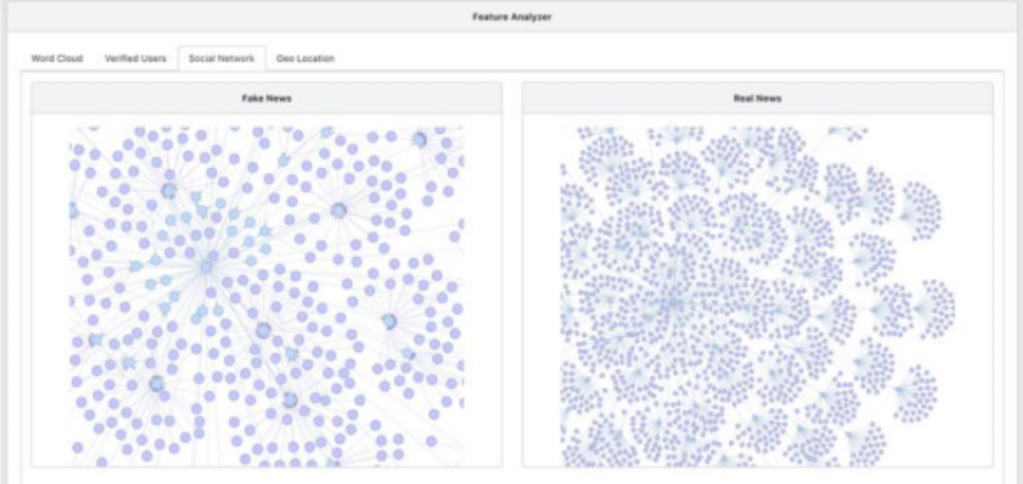
Topics of Fake news vs Real News



Geolocation of Fake News vs Real News



Social Network on Users Spreading Fake/Real news



Recent work at DMML on Fake News Detection

- 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/

Challenges and Highlights

- Fake News Early Detection
- Identify Check-worthy Content
- Cross-domain, -topic, -language Fake News Studies
- Deep Learning for Fake News Studies

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[9]
   - The evolution of deceptive content styles *within* domain, topic, and language

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

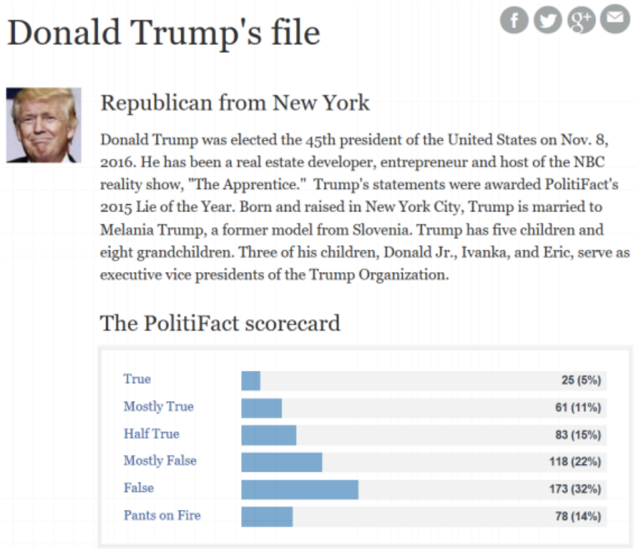
[9] W. Yaqing, et al., EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. KDD'18

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



Donald Trump's file

Republican from New York

Donald Trump was elected the 45th president of the United States on Nov. 8, 2016. He has been a real estate developer, entrepreneur and host of the NBC reality show, "The Apprentice." Trump's statements were awarded PolitiFact's 2015 Lie of the Year. Born and raised in New York City, Trump is married to Melania Trump, a former model from Slovenia. Trump has five children and eight grandchildren. Three of his children, Donald Jr., Ivanka, and Eric, serve as executive vice presidents of the Trump Organization.

The PolitiFact scorecard

True	25 (5%)
Mostly True	61 (11%)
Half True	83 (15%)
Mostly False	118 (22%)
False	173 (32%)
Pants on Fire	78 (14%)

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

Related Studies:

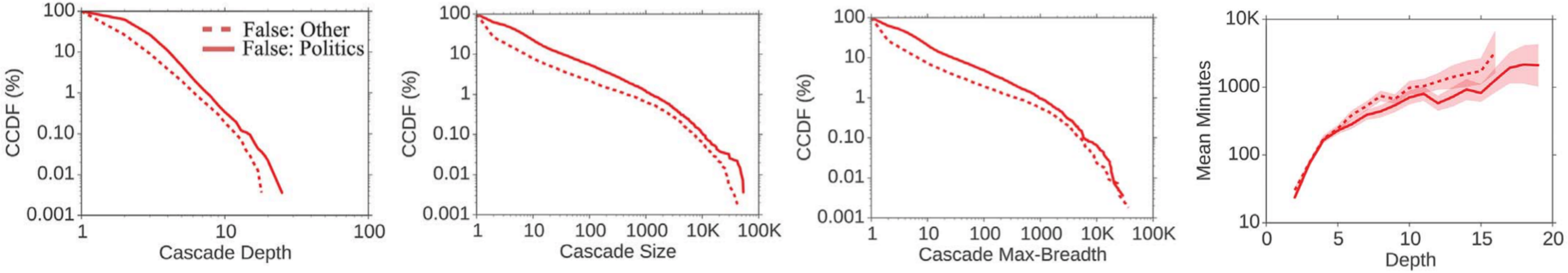
- N. Hassan, et al. Detecting Check-worthy Factual Claims in Presidential Debates, CIKM'15
- N. Hassan et al., Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster, KDD'17

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

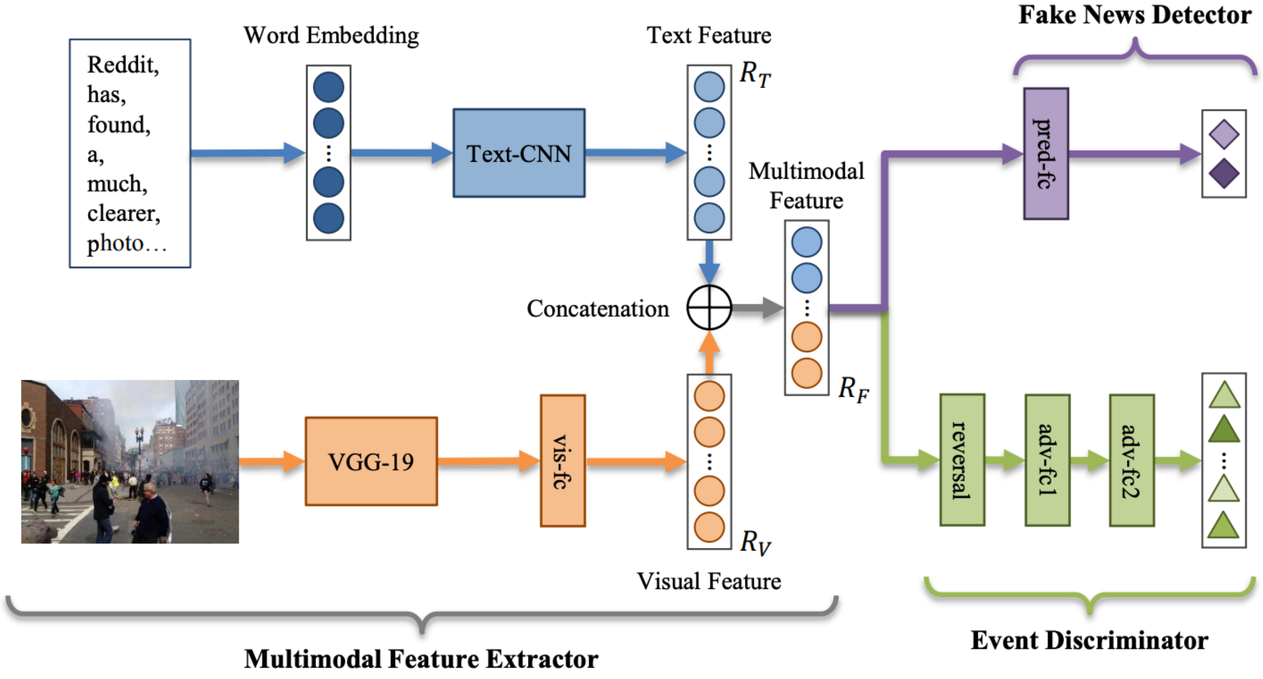
II. Explore **patterns** among fake news within different domains, topics and languages



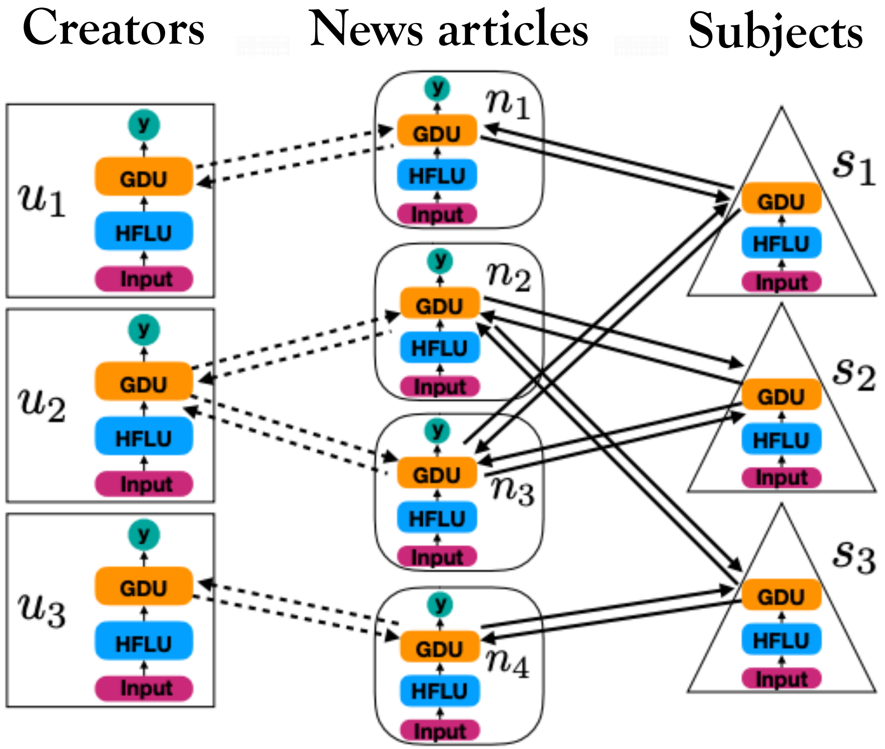
III. Develop **techniques** enables cross-domain, -topic, -language fake news detection

Figures are from: S. Vosoughi, et al. The spread of true and false news online. Science, 2018

Deep Learning for Fake News Detection



W. Yaqing, et al., EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. *KDD'18*

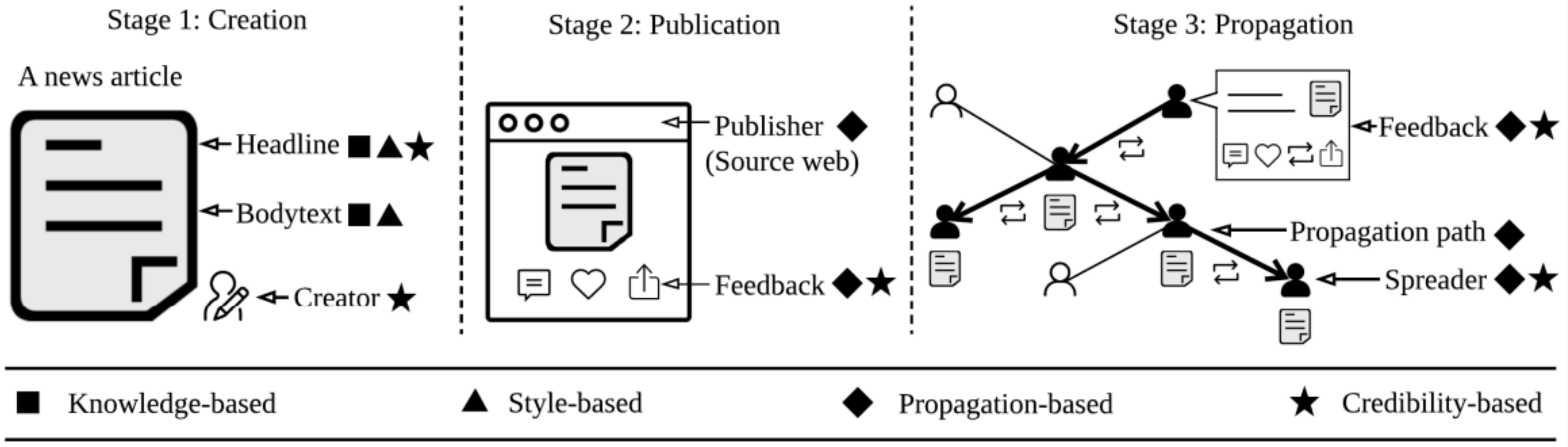


J. Zhang, et al. Fake News Detection with Deep Diffusive Network Model, arXiv: 1805.08751, 2018

Summary

I. **Fundamental Theories** encourage interdisciplinary research of fake news

II. **Fake News Detection** from various perspectives



III. **Challenges and Highlights** for potential research opportunities for fake news studies