#### We Dived Deep into Clickbaits: You won't Believe what Happened Next! The UNIVERSITY of OKLAHOMA THE UNIVERSITY of MISSISSIPPI. Md Main Uddin Rony<sup>1</sup>, Naeemul Hassan<sup>1</sup>, Mohammad Yousuf<sup>2</sup> dear.lab <sup>1</sup>The University of Mississippi, <sup>2</sup>The University of Oklahoma

## Motivation





How Many Of These 'Star Trek' Episodes Have You 'ou haven't lived long and prospered until you've watched these classic

CLICKHOLE COM

## **Clickbaiting has become widespread**

- All types of media practice clickbait.
- It has become a source of easy revenue.

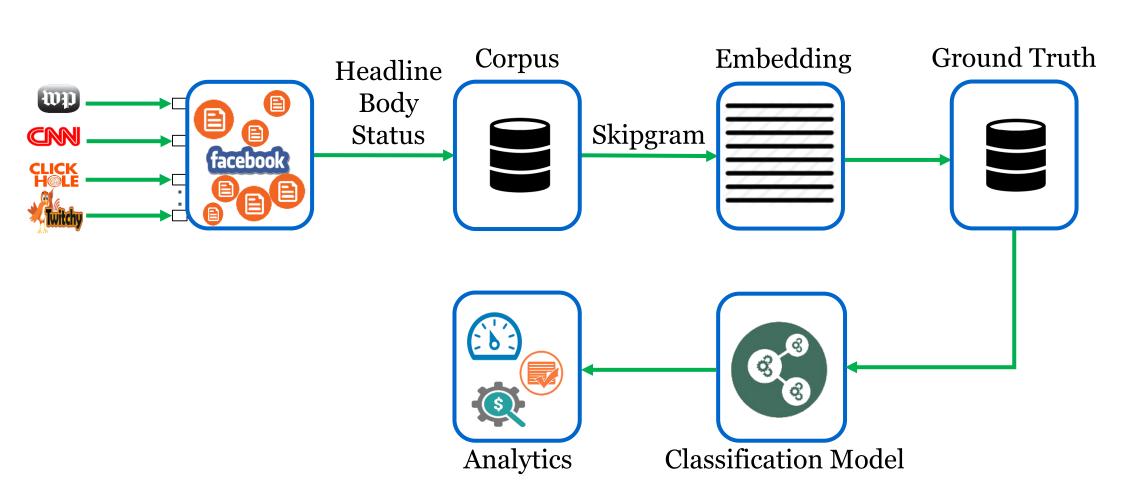
## **Negative impact on media ecosystem**

- Damages media-user relationship.
- Depletes media brand value.

### Lack of research on clickbait practice

- No large scale analysis on practice of clickbait by media organizations.
- No study to show its contribution to public engagement on social media.

# **Problem Formulation**



#### **Two categories of headlines** Clickbait

- Eek! What's Lurking in the Shadows?! I *Have to Know!*
- *I Left My Daughter And THIS Happened!* Non-Clickbait
- America's democracy has become illiberal
- Twitter Expands Fight Against Abuse

## Goal

- Building a clickbait detection model.
- Performing large-scale study on the clickbait practice by a range of media on social network (Facebook).

## Data Collection

#### **Ground Truth**

- 32,000 manually labeled news headlines. [1] Media Corpus
- About 1.7 million Facebook posts.
- From 68 mainstream and 85 unreliable media.
- Considered headlines, bodies, and the statuses.

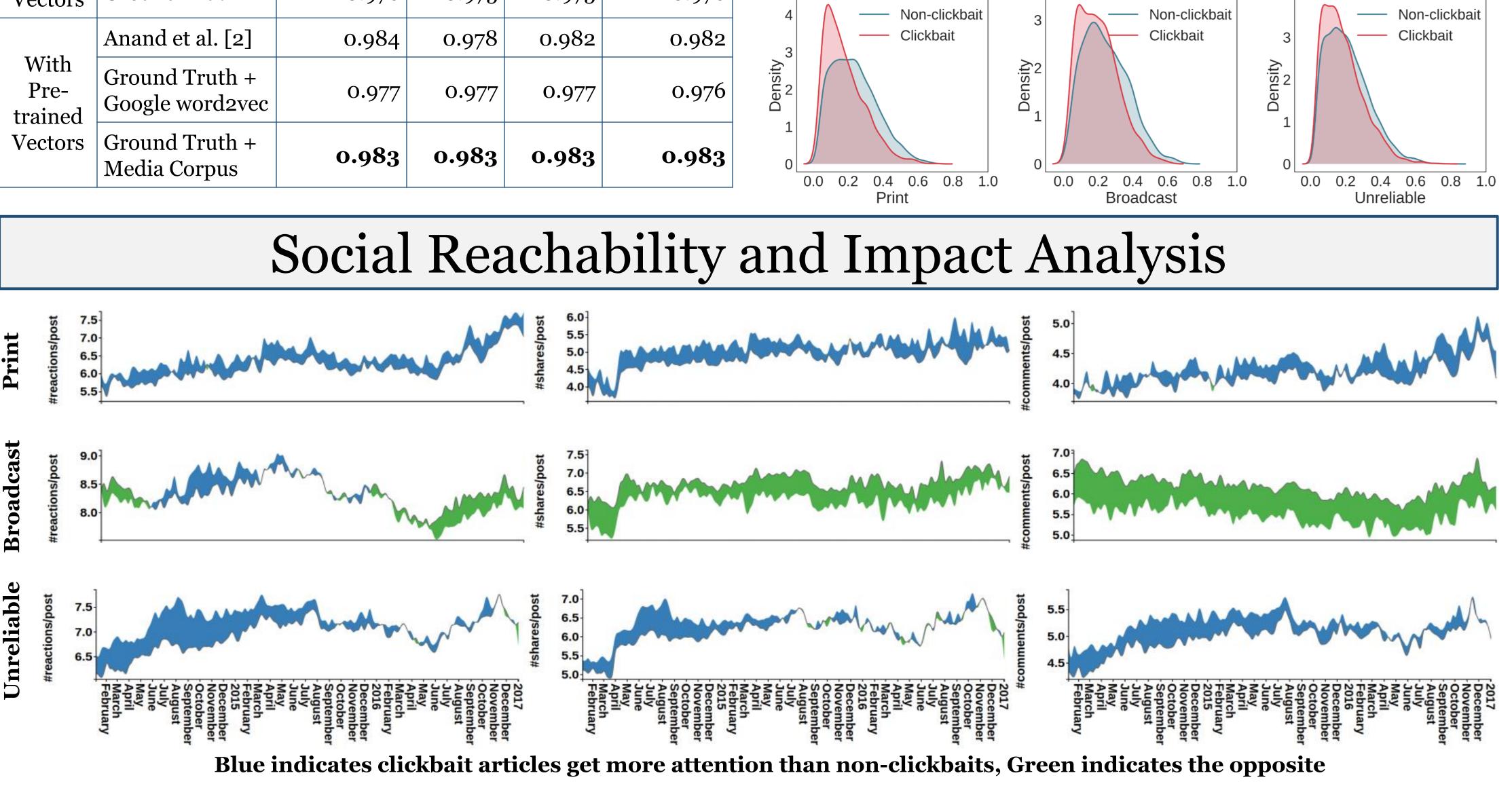
Media	Category	Link	Video	Total
Mainstream	Broadcast	324028	32924	356952
Mainstream	Print	516713	14129	530842
	Clickbait	371834	4099	375933
Unreliable	Conspiracy	309122	5841	314963
Unrenable	Junk Science	51923	649	52572
	Satire	41046	151	41197
]	Total	1614666	57793	1672459

# Clickbait Detection

Projection

- Distributed subword Input embedding was used to create sentence vectors.
- Pre-trained embeddings using 477236 words of  $sw_t$ the Media Corpus.
- Used Softmax function as the classifier.

	Method	Precision	Recall	F-score	Accuracy
Without Pre-	Chakroborty et al. [1]	0.95	0.90	0.93	0.93
trained Vectors	Ground Truth	0.976	0.975	0.975	0.976
	Anand et al. [2]	0.984	0.978	0.982	0.982
With Pre- trained	re- Google word 2vec 0.977 0.977	0.977	0.976		
Vectors	Ground Truth + Media Corpus	0.983	0.983	0.983	0.983



 Non-clickbait headlines highlight topics of collective problems such as public policies.

Prin

Prin (Nor

Broa

(Clie

Broa (Nor

Unr

Output

SW<sub>t-2</sub>

SW<sub>t-1</sub>

SW<sub>t+1</sub>

SW<sub>t+2</sub>

(Nor

## Qualitative Analysis

### **Topic Modeling**

 Used Biterm Topic Modeling (BTM). Clickbait headlines in print and broadcast personalized, media represent more sensationalized and entertaining topics.

nt ickbait)	vearthing new best 2015 day <sup>know</sup>	newthing woman trump make get know donald say	woman say new trump make chicago get star	get best make never woman hirump day kid	new 2016 trump boston donald say
nt on-clickbait)	get San police vega bay jose school	new hillary trump campaign obama clinton say	york woman nevyork police nevyork hawaiisay killed	clinton <b>NEW</b> <sub>take</sub> <b>trump</b> state world state world Say donald	trump soxnew state chronicle boston redsay win
oadcast ickbait)	make- best movie nevyet swift starike time		get episode Woman black new full trump	say history best trump make clinton know get thing	thing photo like photo ay food way dog national
oadcast on- clickbait)	say found death	sop sayu obama	dy year police found <b>nevy</b> dead say nbc killed	game first get Say back	new geographic stundisputed classic national home photo shark
reliable ickbait)	trump say get donald obama	white a merica obama	right people hillary hillary woman eagle <sup>get</sup>	woman likepeople man year thing way way way the people	reason reunionfather thing Way food <b>QAY</b> life health human
reliable on-clickbait)	eop <sup>police</sup> guneagleu obama musliman say	new hillary trump campaign obama clinton say	world world War New Video	say police truth never broadcast obama republican american	unew, world cancer human health food vaccine

#### **Headline-Body Relevance**

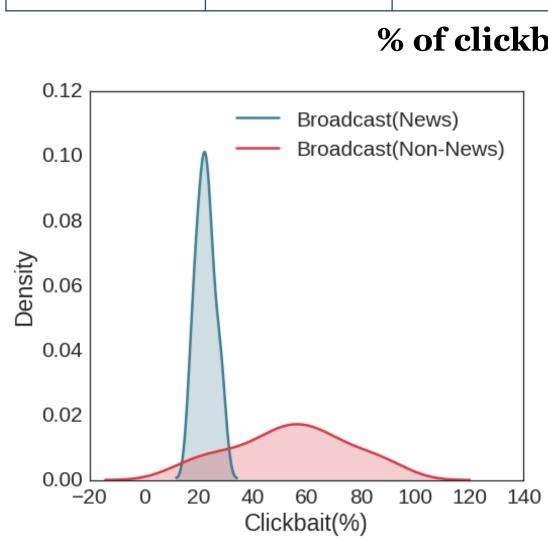
 Hypothesis: Clickbait headlines are less relevant to the body content. Cosine similarity was used to measure the

relevance between a headline and the body.

#### Media

Mainstrea

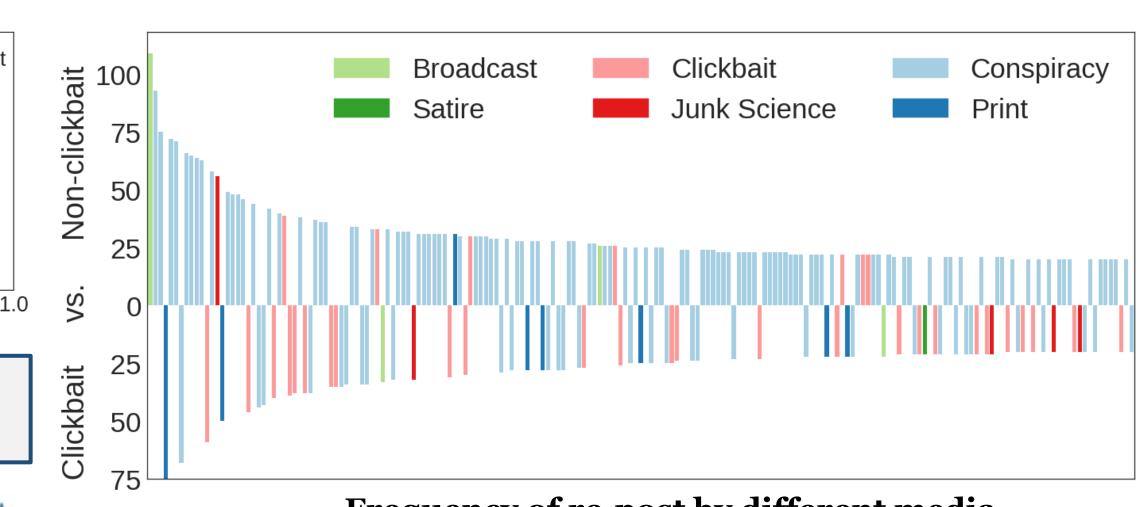
Unreliab



Media

Mainstream

Unreliable



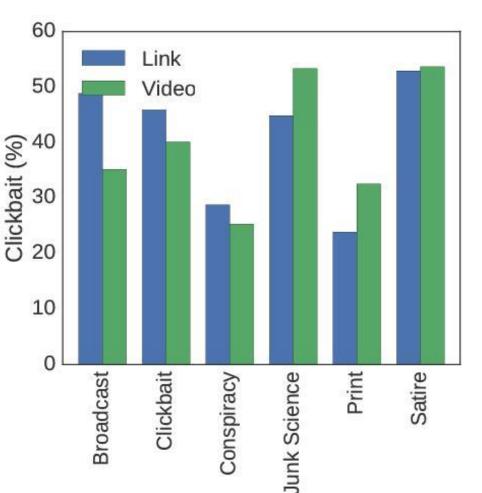
References 1. A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly. Stop clickbait: Detecting and preventing clickbaits in online news media. In Advances in Social Networks Analysis and Mining (ASONAM), 2016.

2 A. Anand, T. Chakraborty, and N. Park. We used neural networks to detect clickbaits: You won't believe what happened next! ArXiv preprint arXiv:1612.01340, 2016.

## Quantitative Analysis

a	Category	Clickbait	Non-Clickbait	Clcikbait(%)
	Broadcast	169752	187200	47.56
eam	Print	128022	402820	24.12
ble	Clickbait	172271	203662	45.82
	Conspiracy	90389	224574	28.7
	Junk Science	23637	28935	44.96
	Satire	21798	19399	52.91

% of clickbaits in various media



% of clickbait in news & non-news % of clickbait in link & video

	Category	Clickbait Status	Non-clickbait Link	Clickbait Status (%)
m	Broadcast	84192	176177	32.34
	Print	164669	379504	30.26
le	Clickbait	91747	157886	36.75
	Conspiracy	46851	190477	19.74
	Junk Science	12764	28349	31.05
	Satire	7425	14453	33.94

% of clickbait in Facebook Status

#### Frequency of re-post by different media