Developing Situational Awareness from Blogosphere: An Australian Case Study

Mainuddin Shaik(mxshaik@ualr.edu), Muhammad Nihal Hussain, Zachary Stine, and Nitin Agarwal

Collaboratorium for Social Media and Online Behavioral Studies (COSMOS)

cosmos.ualr.edu

SOTICS 2021

The Eleventh International Conference on Social Media Technologies, Communication, and Informatics
October 03, 2021 to October 07, 2021 - Barcelona, Spain

Dr. Nitin Agarwal

COSMOS Director - Leading projects with a combined funding of over $20 million from various agencies

- Department of Defense
- DARPA
- Department of State
- National Science Foundation

Dr. Muhammad Nihal Hussain

Data Scientist at Equifax

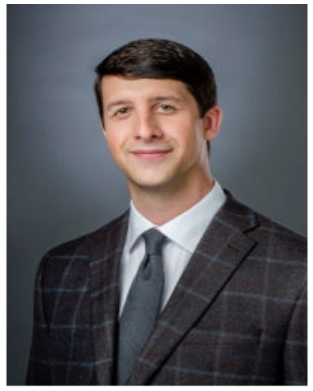
- Social Media Analysis
- Social computing
- Data and Web Mining
- Behavior Modelling

Mainuddin Shaik

COSMOS Doctoral Student

- Computational Social Science
- Mixed Methods
- Network Science
- Topic modelling

Dr. Zachary Stine



COSMOS Post-doctorate Fellow

- Complex systems
- Natural language processing
- Network Science
- Information Theory

Objective of this study

To study the discourse that deals with diplomacy, defense, trade tension surrounding Australia and China. Additionally, election related discourse in Australia.

Why?

- Help policy-makers make informed decisions
- Quickly visualize trending topics over time
- Discover influential topics
- Focus on influential discourse movers
- Useful for political scientists and sociologists to mine the key concerns from the influential discourse.

Agenda

1. Data collection
2. Methodology
3. Results
4. Future directions

Data collection

- Data collected from July 1, 2019, to December 31, 2020.

- Keywords used in the study were:
'Australia', 'China', 'South China Sea', 'Military Region', 'Fight', 'Tensions', 'Election', 'India' 'Road Belt Initiative', 'Thai', 'Defence' 'Defense' 'Indo pacific', 'Asia Pacific'.

- Curated 20,066 relevant blog posts in total.

- Data into two groups a dataset including all 20,066 blog posts (i.e., All data) and 2) a dataset that excludes any COVID-19 related data (i.e., No COVID data).

- Used the following keywords *'+Australia - (covid covid-19 coronavirus pandemic vaccine)'* to exclude the COVID related posts.

Data can be accessed through ‘Australian DOD tracker’

- Available at https://btracker.host.ualr.edu

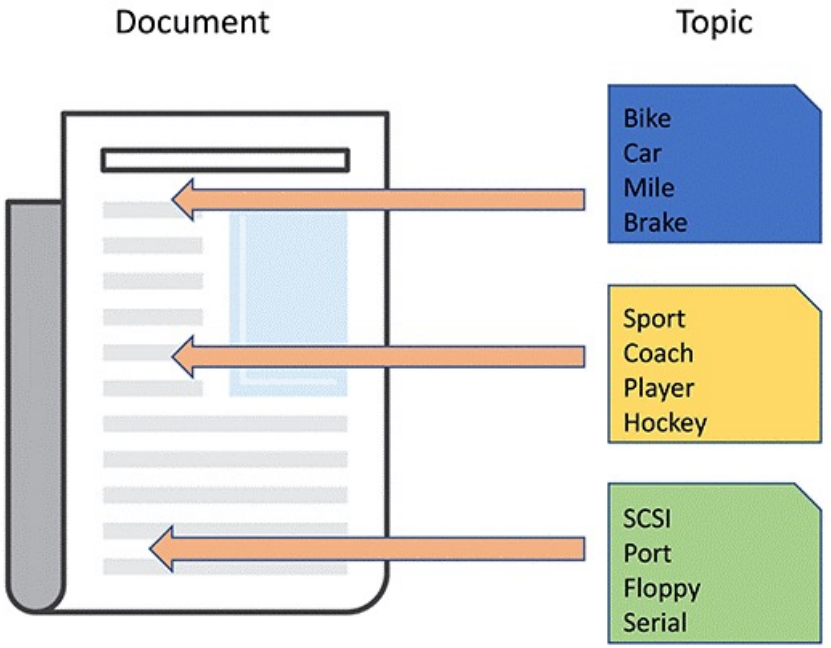
Data elements	All data	No COVID data
Total posts	20,066	10,113
Total blog domains	679	344
Total blog authors	4,217	2,494
Comments	417,050	84,798

Methodology

- Topic Modelling
- Influence
- Network analysis (will explain in ‘Results’ section)

Topic Models

Latent Dirichlet Allocation (LDA) is a probabilistic generative model



Source: https://aws.amazon.com/blogs/machine-learning/introduction-to-the-amazon-sagemaker-neural-topic-model

LDA TOPICS FROM 'ALL DATA' AND 'NO COVID DATA'

Topic 1. china, chinese, military, india, war, states, security, power, trump, united
Topic 4. covid, vaccine, health, coronavirus, virus, cases, pandemic, public, deaths, care
Topic 2. climate, global, market, trade, energy, economics, china, change, australian
Topic 3. trump, see, america, great, biden, news, get, html, video, know
Topic 0. election, party, vote, policy, bank, zealand, voters, per, labor, think

Topic 3. china, chinese, states, war, government, military, power, united, may, use
Topic 0. women, get, much, life, well, made, see, way, back, know
Topic 1. government, australian, china, state, climate, media, per, much, news, change
Topic 2. china, australian, india, south, water, sea, government, pacific, country, fire
Topic 4. party, government, election, vote, labor, australian, votes, parties, preferences, candidates

MAPPING TOPIC THEMES BETWEEN 'ALL DATA' AND 'NO COVID DATA'

Theme of the topic	All data	No COVID data	Sample blogposts
a) South China Sea, China, Australia, USA, Defense tension	**Topic 1**	**Topic 2, Topic 3**	'Beijing's Hong Kong plans may lead to an exodus, and Australia must be ready'. 'huh, fuck the quad'. 'The massive, systemic and grave crimes committed over the years by fighters of Australian elite'. 'Singapore-Australia exercise activities involves Singapore's F-15SGs'.
b) Climate change, Economy, and Trade	**Topic 2**	**Topic 1, Topic 2**	'Too much fuel causes extreme bush fires, not climate change'.
c) Federal and Regional elections in Australia	**Topic 0**	**Topic 4**	'Scott Morrison is up two on approval to 66% and down two on disapproval to 30%, while Anthony Albanese is up one to 44% and up two to 41%, with Morrison's lead as preferred prime minister out from 58-29 to 60-28'

Influence

Influence of a blogpost can be visualized as an influence graph or i-graph.

1) Recognition (number of in-links, represented by ι)
2) Activity generation (number of comments, represented by γ)
3) Novelty (number of outlinks, represented by θ), and Eloquence (length of the blogpost, represented by λ). In a directed i-graph, if $f(x)$ denotes the influence flow of a blogpost *x* with *I* being the influence of *x*. The influence of each blogpost $I(x)$, can be calculated as

$$I(x) = w(\lambda)\times\left(w_{com}\gamma_x + f(x)\right) \quad \text{--(1)}$$

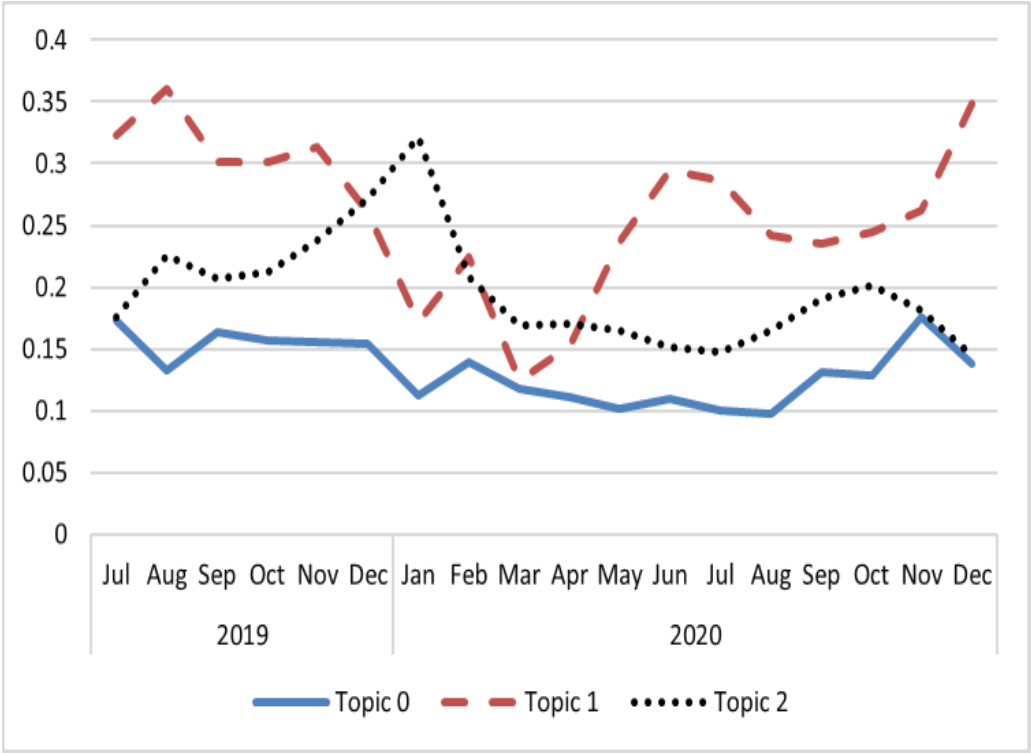
$$f(x) = w_{in}\sum_{m=1}^{|\iota|} I(x_m) - w_{out}\sum_{n=1}^{|\theta|} I(x_n) - -(2)$$

$$m(x) = I(x)\times p(topic|blogpost) \quad \text{--(3)}$$

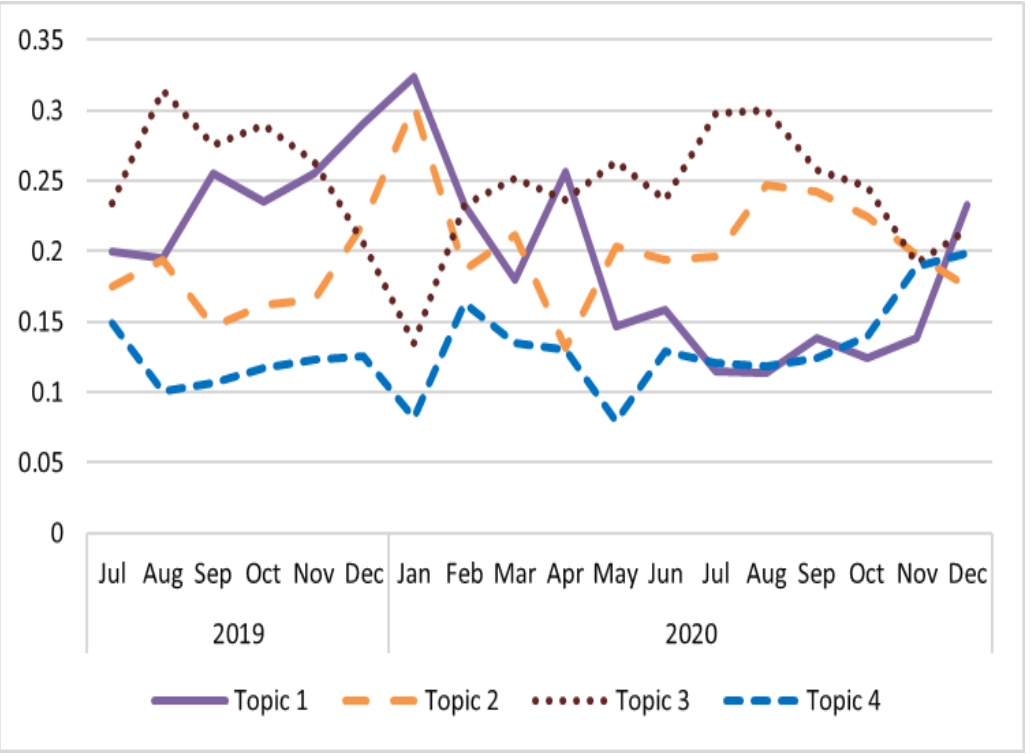
A	B	C	D	E
blogpost_id	infleunce	**date**	**0**	**Influence_0**
4348983	**8.7**	2019-07-01 00:00:00	0.372759	3.243004981

RESULTS

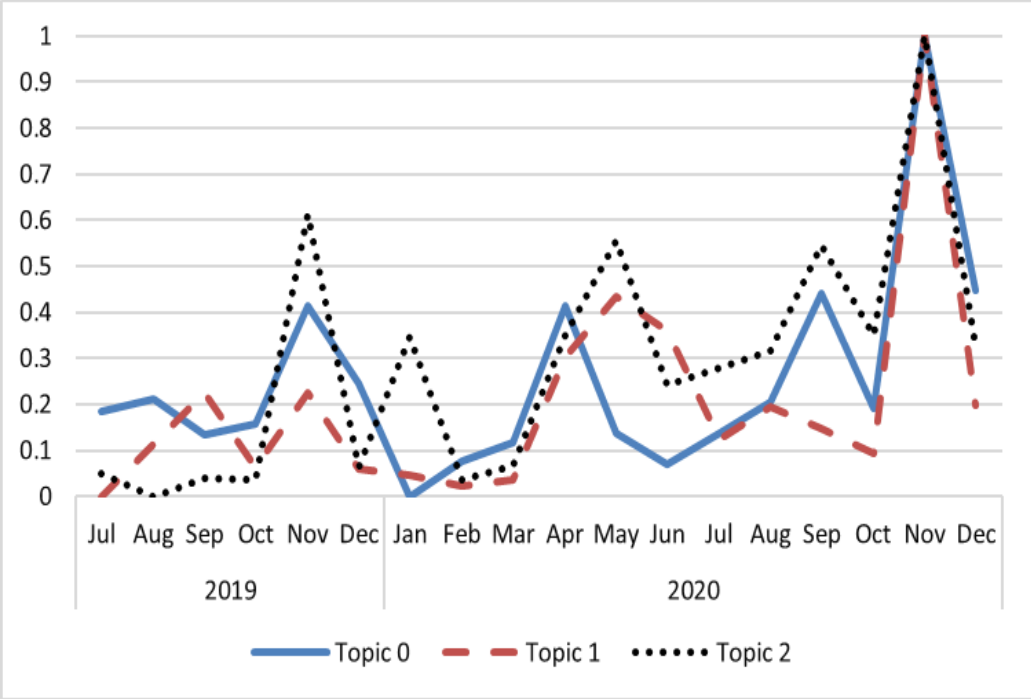
All data



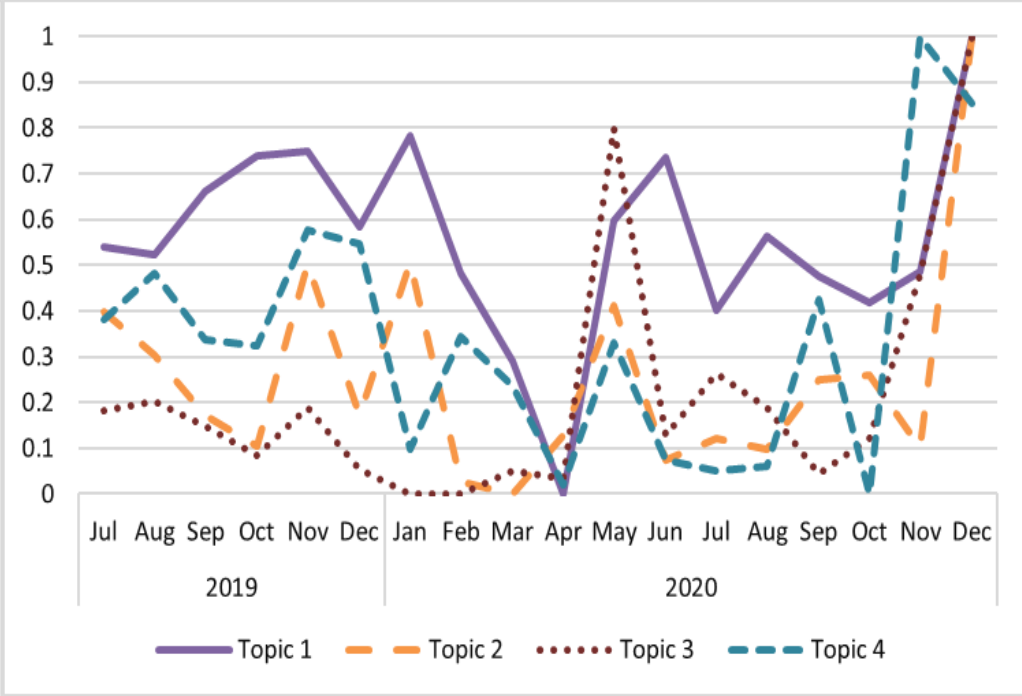
No COVID data



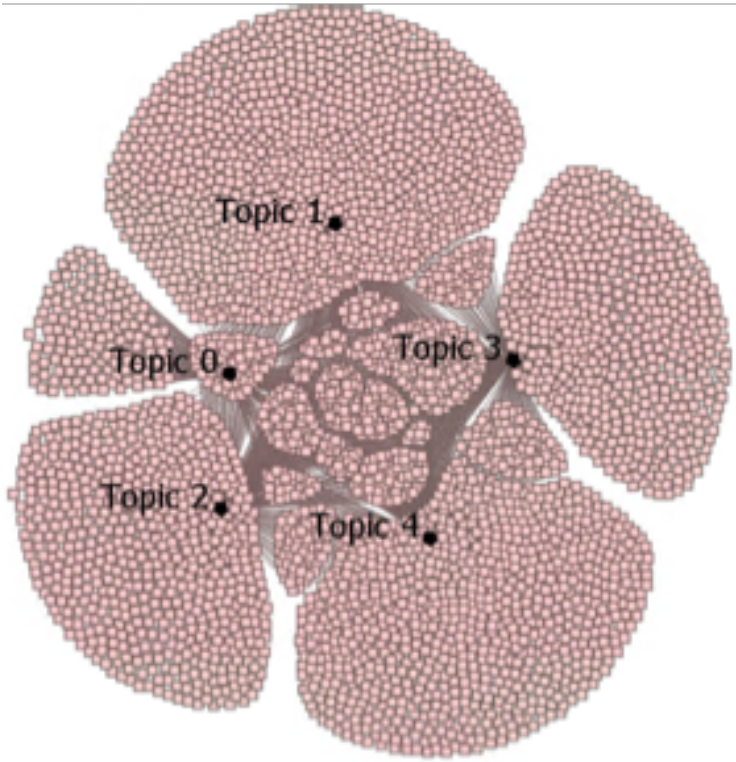
All data influence



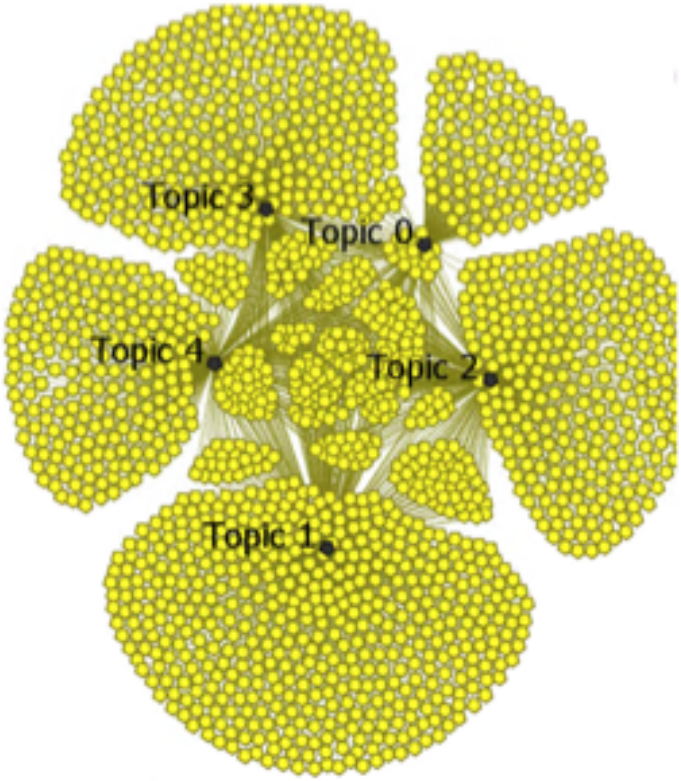
No COVID data influence



Topic blogger Network

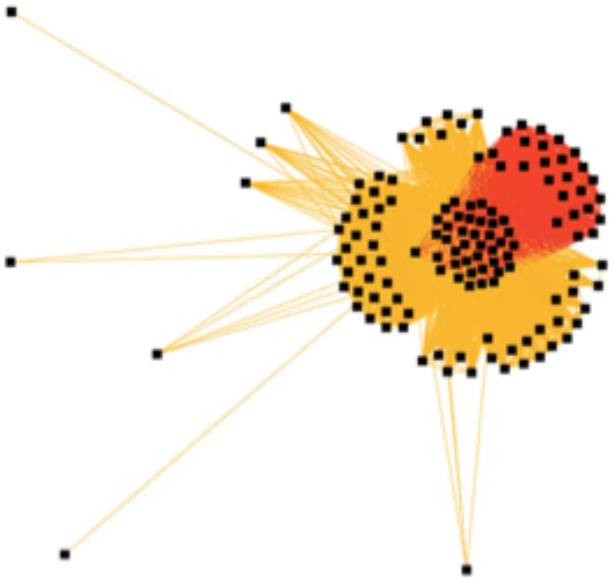


All Data

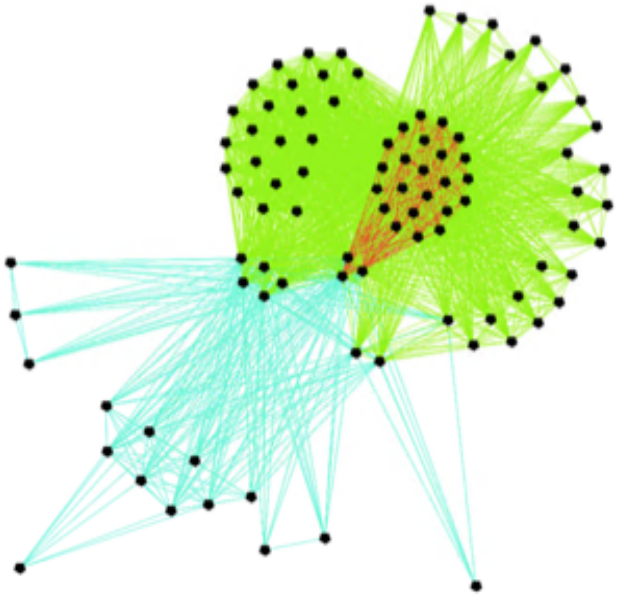


No Covid Data

Folded Topics and Blogger network



All Data



No Covid Data

CORE BLOGGERS AND THEIR INFLUENCE SCORE

	'All Data'	
Blog	Blogger	Influence score
catallaxyfiles.com	Sinclair Davidson	11100.6
catallaxyfiles.com	currencylad	5260.5
theaimn.com	The AIM Network	4151.4
quadrant.org.au	quadrant	1566.5
catallaxyfiles.com	Guest Author	1536.6
theaimn.com	Dr Binoy Kampmark	1488.3
pngattitude.com	Keith Jackson	198.6
crikey.com.au	Charlie Lewis	6.6
	'No COVID data'	
Blog	Blogger	Influence score
zerohedge.com	Tyler Durden	23678.7
catallaxyfiles.com	currencylad	2560.5
theaimn.com	The AIM Network	1925
crikey.com.au	John Quiggin	753.9
quadrant.org.au	quadrant	613.5
theaimn.com	Dr Binoy Kampmark	553.8

COSMOS Tools:

- **Blogtrackers** - https://btracker.host.ualr.edu
- **VTracker** - https://vtracker.host.ualr.edu
- **Focal Structure Analysis** - http://fsa.host.ualr.edu/
- **COVID-19 MISINFO - https://cosmos.ualr.edu/covid-19**

Blogtrackers YouTubeTracker

FOCAL STRUCTURE ANALYSIS

cosmos.ualr.edu

Follow **cosmographers** on

ACKNOWLEDGEMENTS

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-20-1-0262, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

COSMOS Collaboratorium for Social Media and Online Behavioral Studies

National Science Foundation · Office of Naval Research · Air Force Research Laboratory · Defense Advanced Research Projects Agency · Army Research Office · Department of Homeland Security · Arkansas Research Alliance