Developing Socio-computational Approaches to Mitigate Socio-cognitive Security Threats in a Multi-platform Multimedia-rich Information Environment

Nitin Agarwal, Ph.D.
Maulden-Entergy Chair and Distinguished Professor
Director, COSMOS Research Center
University of Arkansas – Little Rock
nxagarwal@ualr.edu

COSMOS

Collaboratorium for Social Media and Online Behavioral Studies

Statewide research center headquartered in the University of Arkansas - Little Rock, USA

78 members (Canada, USA, St. Vincent & The Grenadines, Germany, France, Turkey, Nigeria, Iraq, Pakistan, India, Nepal, Bangladesh)

Follow **cosmographers** on

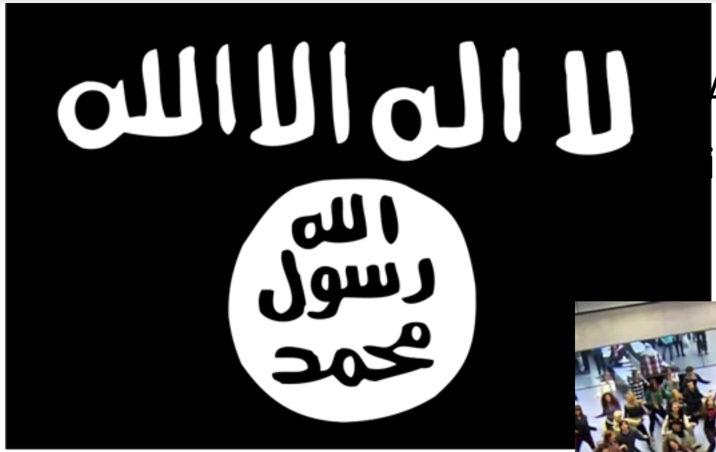
https://cosmos.ualr.edu/

Agenda

- Characterizing multimedia information environment tactics and impact assessment
- Collective action-based framework to characterize coordinated cognitive attacks
- Characterizing information actors (producers and consumers) engaged in cognitive attacks

Problem, Solution, and Usefulness

- ATO cyber influence operation
- ises to assist public affairs and



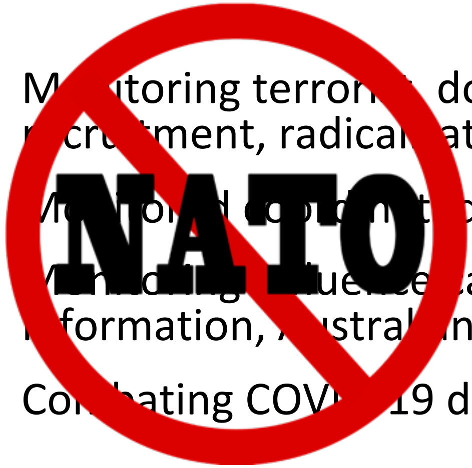
The Islamic State in Iraq & Leva
(ISIL/ISIS/Daesh)

Dragoon Ride
2015

Hackers Networks (DHNs)

Baltic Operations
2019

- M itoring terror domes
  r cru tment, radical ation, a
- Mo ito d coordi t d inau
- M ndoring ue ce ampaigns in ngap
  i formation, A stral n DoD Strateg Univ
- Co bating COV 19 disinformatio torne



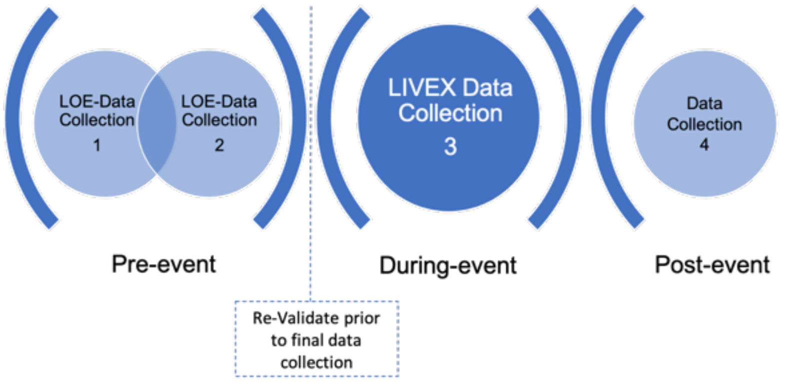
Anti-NATO Propagandist

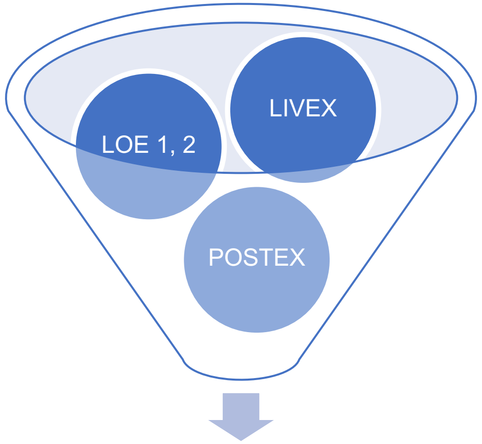
Flash Mobs (deviant)

Internet Trolls

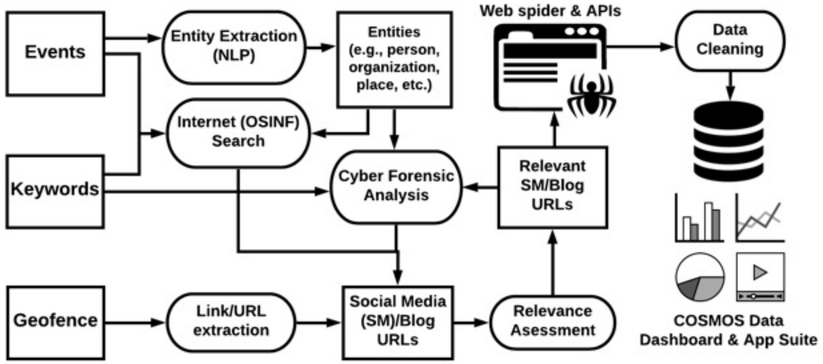
Violent Protests

Precise Data Collection





Bottom-line Up Front (BLUF)



Methodology

Dashboard

Sources include
- Events
- Keywords
- Geofence
- Entities
- OSINF/metadata through cyber forensics

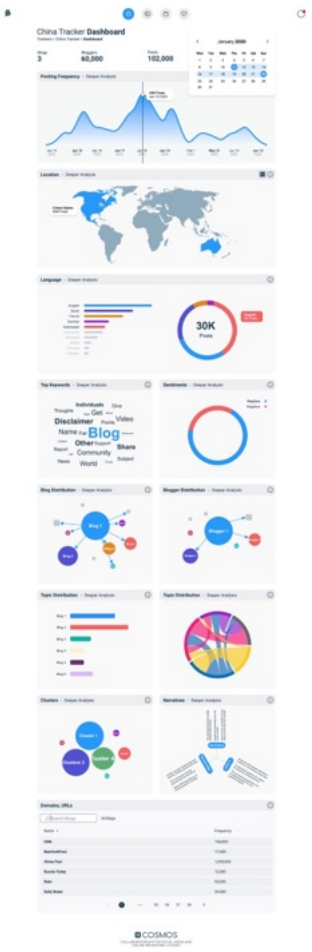
Over 300 GB of data collected every day consisting of
- Text
- Images
- Audio
- Video
- Networks
- Metadata

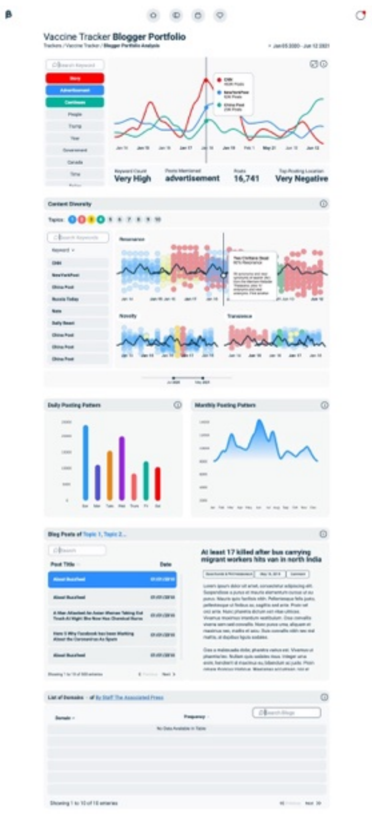
Multi-threaded, distributed, resilient, and scalable data collection framework has been developed, evaluated, and deployed.

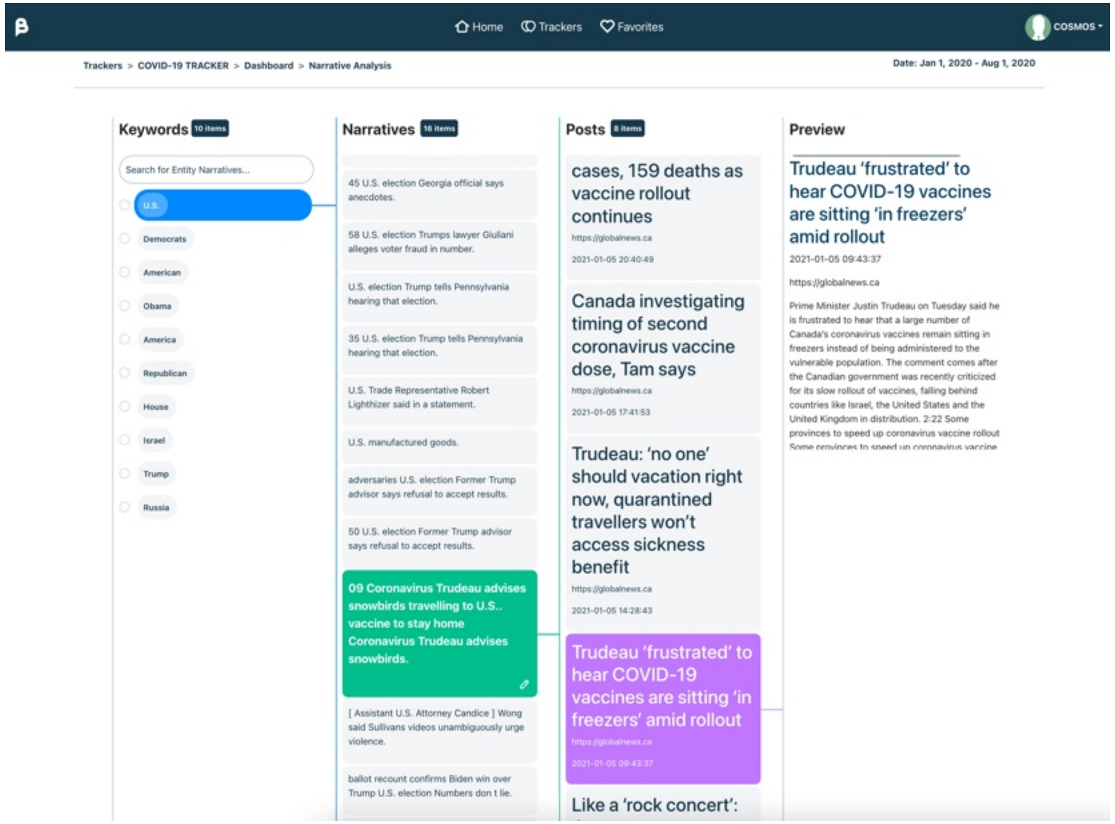
Information Filtering

Basic Filters - Dashboard

Advanced Filters – Narrative extraction and resonance with community, alarming shift in discourse of a blogger







- Environmental Impact
- Disturbs Local Economy
- Threatening Stability
- Mocking/ridiculing NATO
- NATO is aggressive
- NATO poking 'Russian Bear'
- NATO is a joke – Not a real threat
- 'Stop NATO' Protests/Astroturfing
- Mistreatment by NATO Troops

Set up indicators and warnings (I&W system)

Influence Assessment

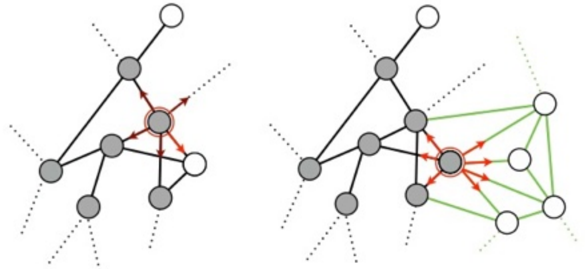
Real-time influence assessment

$$InfluenceFlow(p) = w_{in} \sum_{m=1}^{|\iota|} I(p_m) - w_{out} \sum_{n=1}^{|\theta|} I(p_n)$$

$$I(p) \propto w_{comm} \gamma_p + InfluenceFlow(p)$$

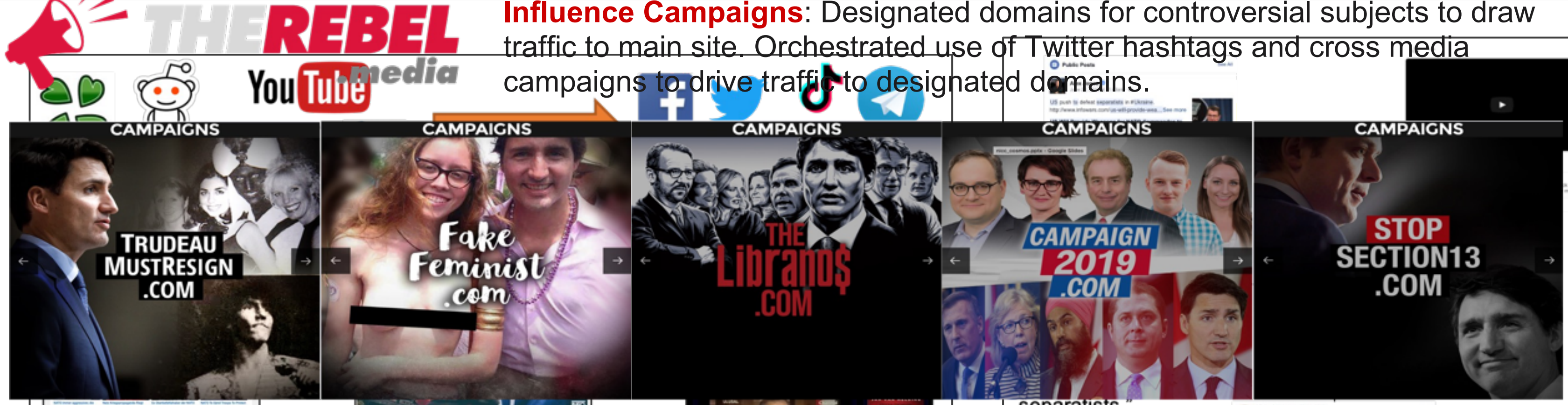
$$I(p) = w(\lambda) \times (w_{comm} \gamma_p + InfluenceFlow(p))$$

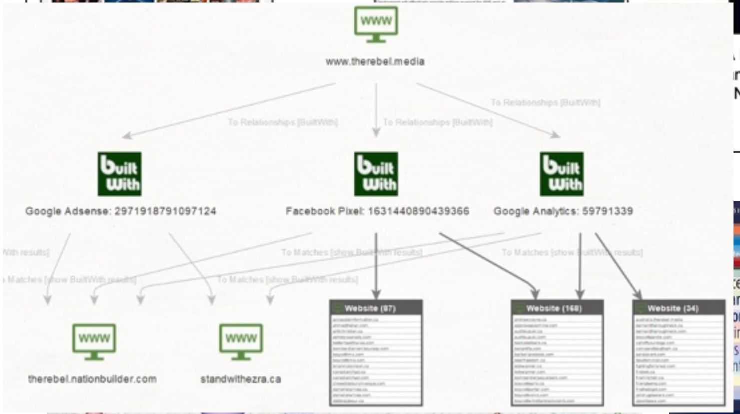
$$iIndex(B) = \max(I(p_i))$$

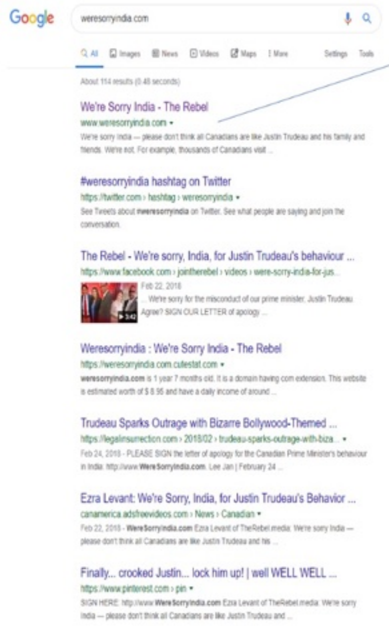


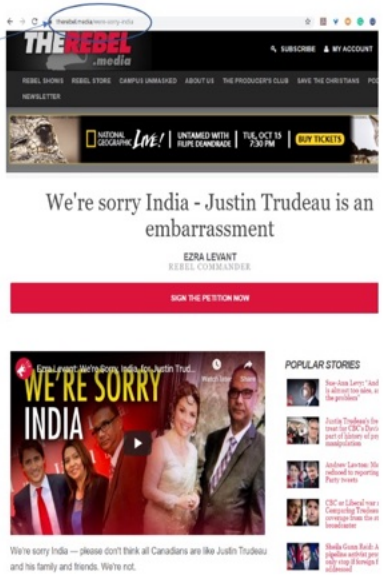
Multiplatform Influence Campaigns

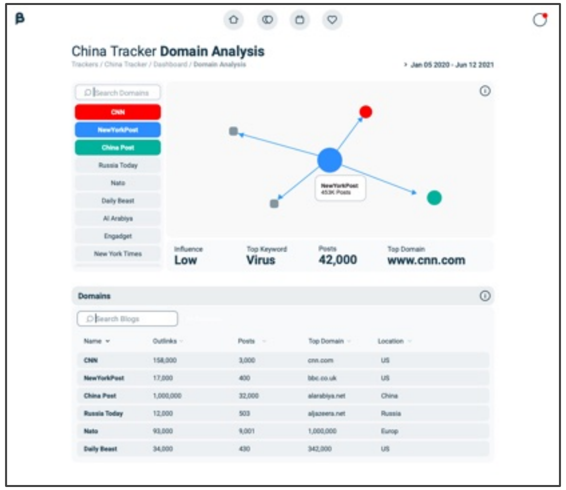
Influence Campaigns: Designated domains for controversial subjects to draw traffic to main site. Orchestrated use of Twitter hashtags and cross media campaigns to drive traffic to designated domains.

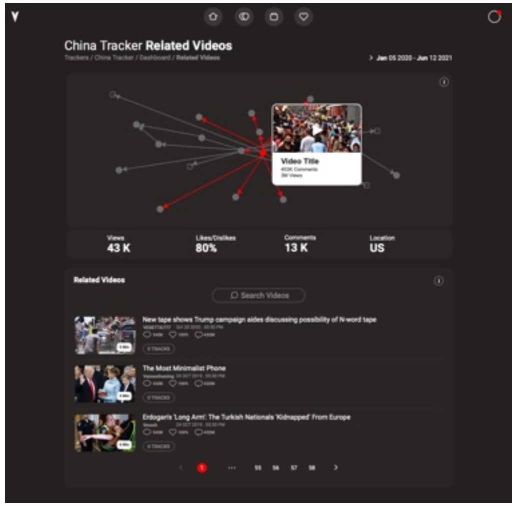








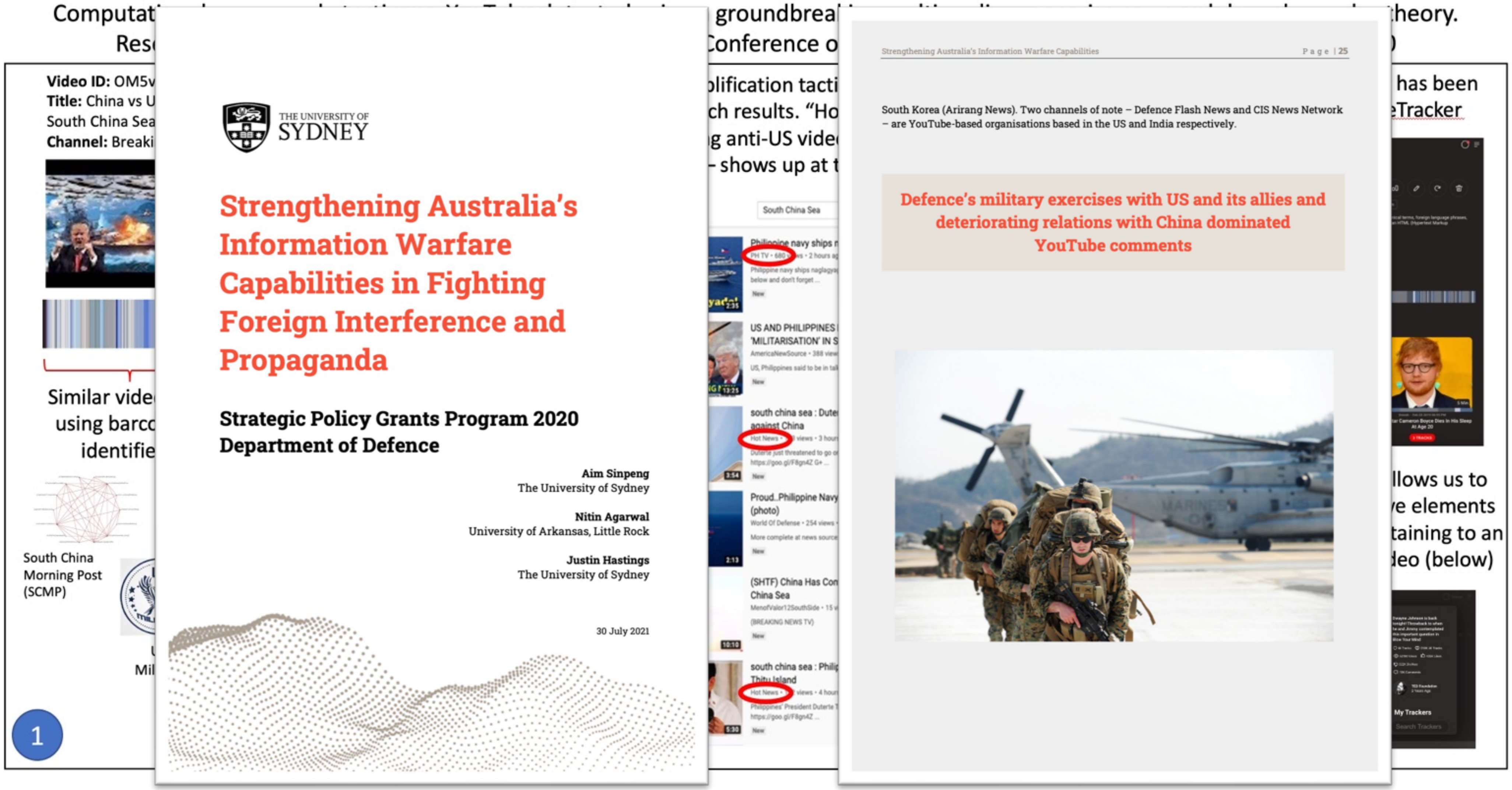




DRDC
RDDC
Canada

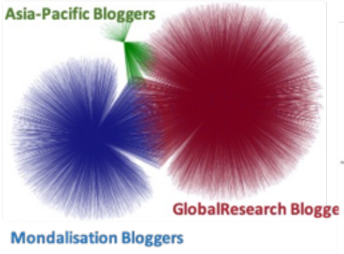
291 fake websites created using “Built with API” service with anti-Trudeau narrative. One such fake website “weresorryindia.com” amplifies anti-Trudeau narrative suggesting he is not a friend of the Canadian-Indian community. The website tops the results in search engines

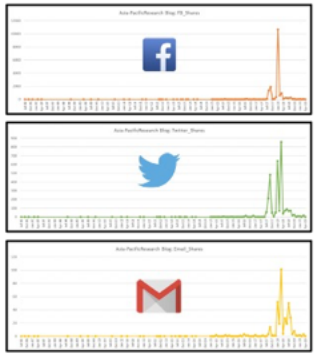
Multimedia Influence Campaigns

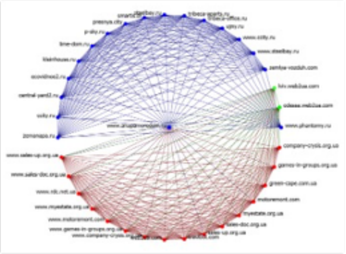


Coordinated Campaigns

Coordinated bloggers



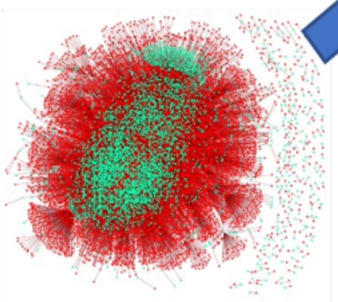




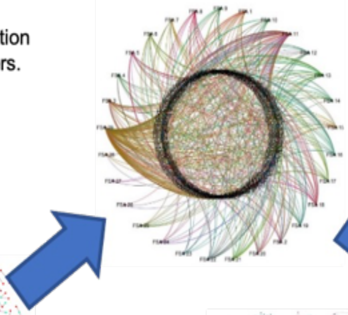
Anti-West/anti-US Indo-pacific bloggers

Coordinated YouTubers

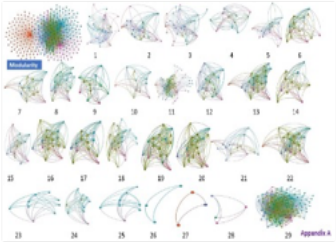
Multi criteria optimization formulation to identify disinformation spreaders. (Alassad, Agarwal, et al., 2021) Central European Journal of Operations Research, Springer



YouTube COVID-19 misinformation video-user network



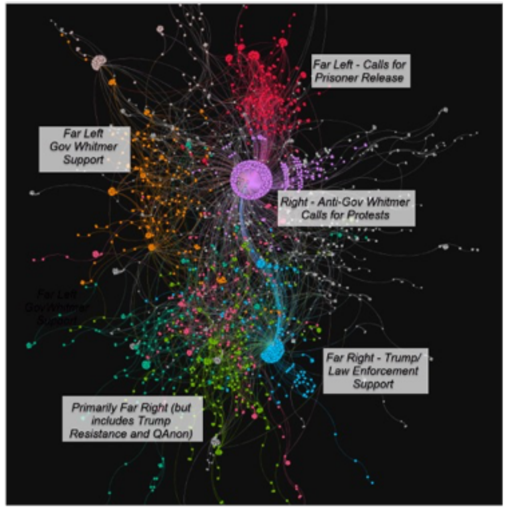
Identifying misinformation spreaders



Sorted on effectiveness

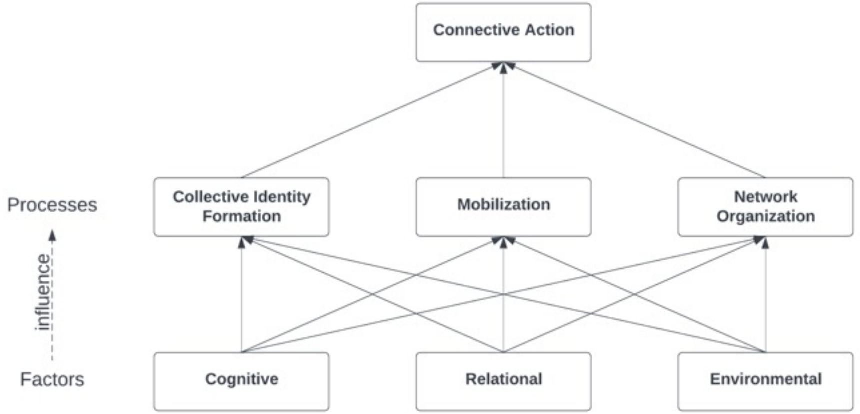
COVID-19 misinformation and conspiracy theory YouTubers

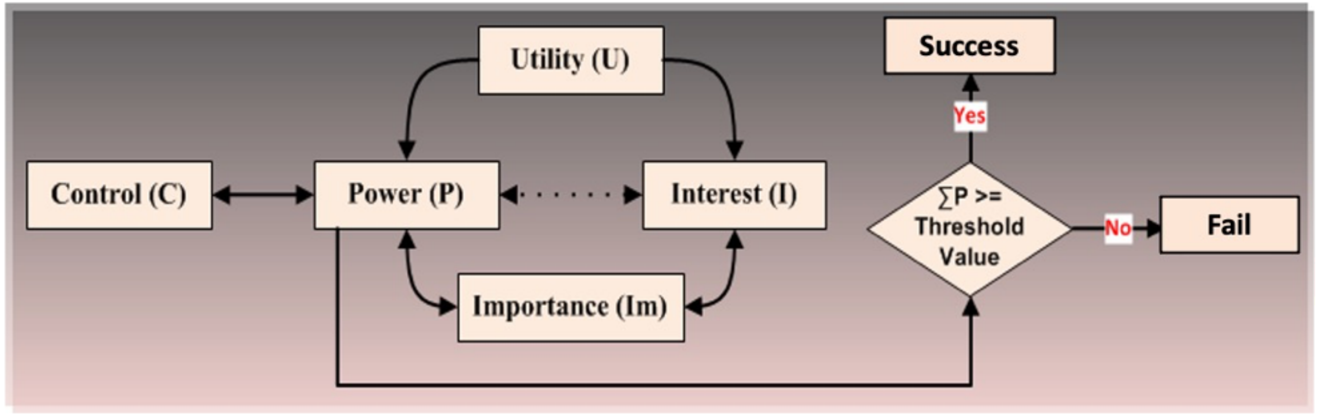
Coordinated Twitter users



Michigan lockdown protest network
#LetMiPeopleGo, #MiLeg, #Endthelockdown, #MichiganProtest; April 1 to May 20; 16,383 tweets; 3,632 nodes; 382 groups (focused on 5 most powerful groups)
Research showed powerful coordination among conspiracy groups calling for protest and actions against Gov. Whitmer. FBI later unraveled a plot to kidnap Gov. Whitmer.

Coordinated Campaigns





$$\max \sum_{i=1}^{n} \delta_i \tag{1}$$

Subject to

$$\delta_i = \{d\vec{c}_1 \preccurlyeq d\vec{c}_2 \preccurlyeq d\vec{c}_3 \preccurlyeq \cdots \preccurlyeq d\vec{c}_i\} - \overline{dc_j^Q} \quad \forall i,j \tag{2}$$

$$d_i^c = \sum_j m_{ij} \quad \forall i \tag{3}$$

$$d_i^c \geq 2 \quad \forall i \tag{4}$$

$$D_G^L = \frac{1}{n}\sum_{i=1}^{n} d_i^c \tag{5}$$

$$D_G^L < d_i^c \leq D_G^U \quad \forall i \tag{6}$$

$$a_i^c = \frac{(\#\ of\ Triangles) \times 3}{\#\ of\ Connected\ Triples\ of\ Nodes} \quad \forall i \tag{7}$$

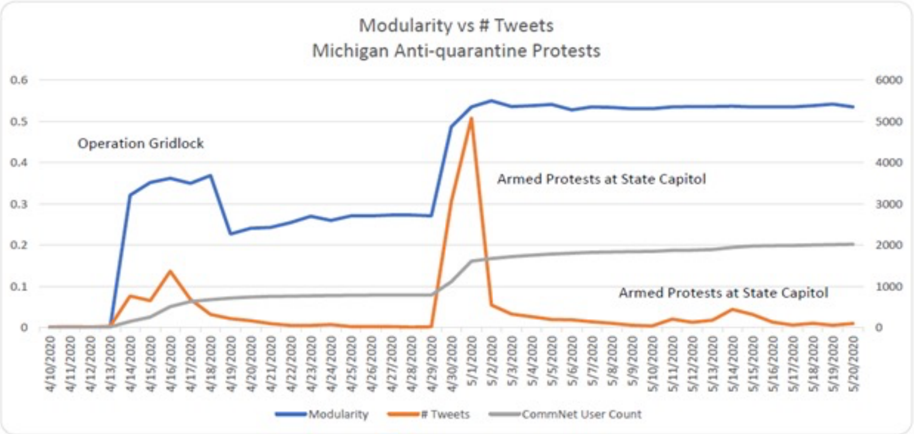
$$AC_G^L = \frac{1}{n}\sum_{i=1}^{n} a_i^c \tag{8}$$

$$AC_G^L < a_i^c \leq AC_G^U \quad \forall i \tag{9}$$

$$\overrightarrow{C_v} = \{\overrightarrow{c_1}, \overrightarrow{c_2}, \overrightarrow{c_3}, \ldots, \overrightarrow{c_i}\} - \overline{c_j^Q} \quad \forall i,j \tag{10}$$

$$\overline{\vec{c}\delta_{i}}_{n\times k} = \overrightarrow{C_{\delta_i}} \quad \forall i \tag{11}$$

$$F = \left\{c_0, \overline{c_j^Q}, \overline{c_{j+1}^Q}, \ldots, \overline{c_\kappa^Q}\right\} \quad \forall j,\kappa \tag{12}$$



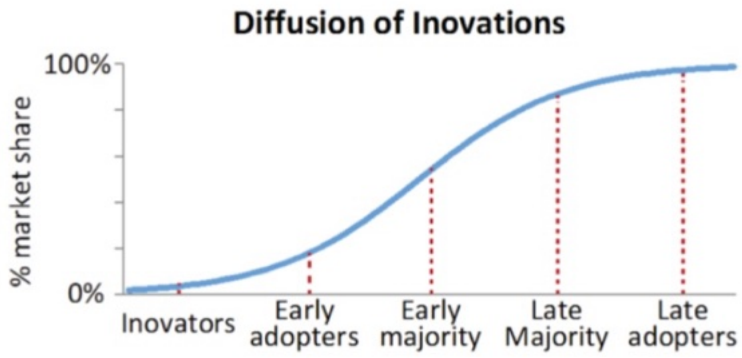
Communities	Political Category	No. of Users in Each Community	No. of Nodes	No. of Edges	Modularity	Average Weighted Degree	Average Betweenness Centrality	DCFM Power
Largest Community	Right	459	510	526	0.294	1.457	0	90.88
Second Community	Right	152	284	419	0.578	2.771	0	77.12
Third Community	Right	212	269	322	0.468	1.792	0	280.98
Fourth Community	Left	78	204	491	0.422	9.779	24.26	10.36
Fifth Community	Left	115	243	339	0.608	7.683	0.16	28.94

Table Y1 - Top 5 Largest Modularity-Based Communities

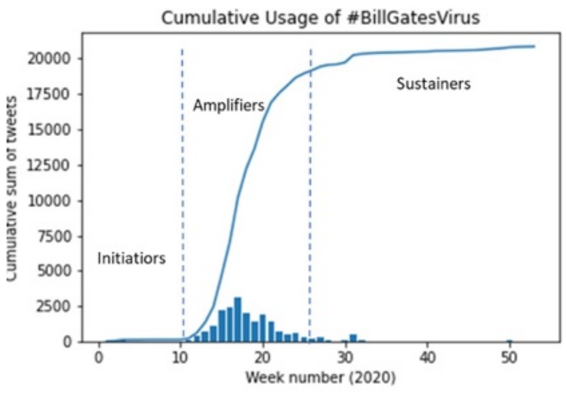
Agarwal, N. et al. (2020) Combining Advanced Computational Social Science and Graph Theoretic Techniques to Reveal Adversarial Information Operations . Journal of Information Processing and Management. Elsevier.
Agarwal, N. et al. (2021) Decomposition Optimization Method for Locating Key Sets of Commenters Spreading Conspiracy Theory in Complex Social Networks. Central European Journal of Operations Research. Springer.

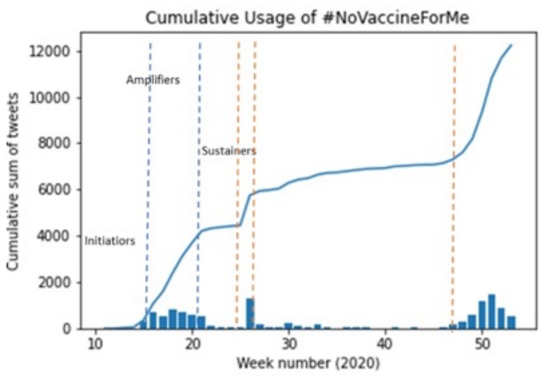
Information Actor Characterization

Characterization of information producers



Rogers (1962). Diffusion of innovations.



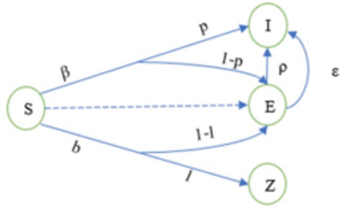


Spann & Agarwal (2022), OSNEM, Elsevier.

Characterization of information consumers

How misinformation spreads? Who should be inoculated?
Leveraging epidemiological modeling. Treat misinformation like an epidemic.

- S: Susceptible
- I: Infected
- Z: Skeptic
- E: Exposed

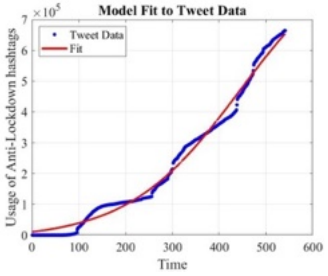


$$\frac{dS}{dt} = -\beta S\frac{I}{N} - bS\frac{Z}{N}$$

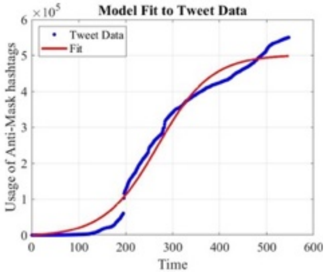
$$\frac{dE}{dt} = (1-p)\,\beta S\frac{I}{N} + (1-l)bS\frac{Z}{N} - \rho E\frac{I}{N} - \varepsilon E$$

$$\frac{dI}{dt} = p\beta S\frac{I}{N} + \rho E\frac{I}{N} + \varepsilon E$$

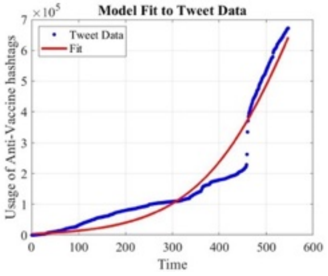
$$\frac{dZ}{dt} = lbS\frac{Z}{N}$$



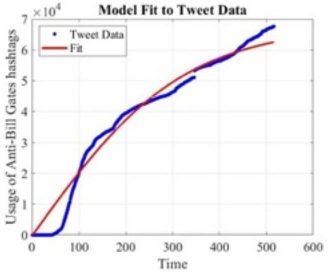
Anti lockdown 6.9% error



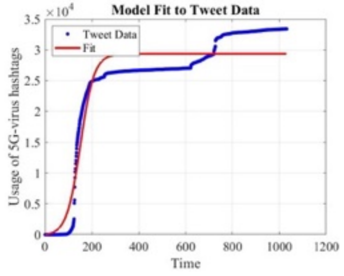
Anti mask 8.2% error



Anti vaccine 12.3% error



Anti Bill Gates 7.9% error

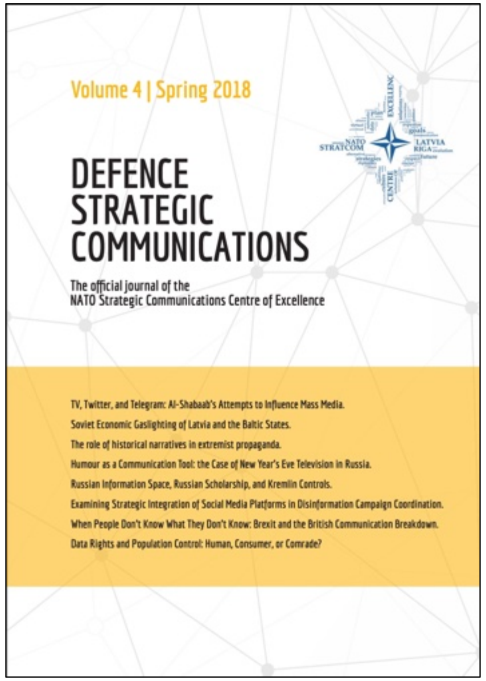


Anti 5G 9.5% error

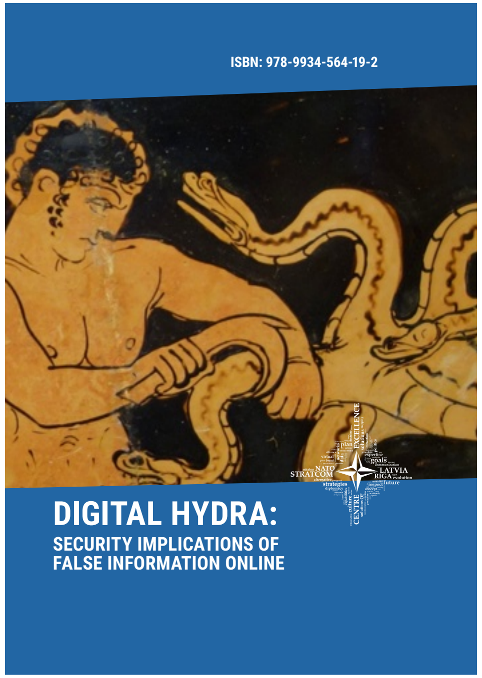
Maleki & Agarwal, HICSS 2022

Coordinated Campaigns

Published by Army University Press, in 7 volume book set on ***Large-Scale Combat Operations***, in the book titled "Perceptions are Reality: Information Operations" (AUSA)



Examining Strategic Integration of Social Media Platforms in Disinformation Campaign Coordination. **Journal of NATO Defence Strategic Communications**

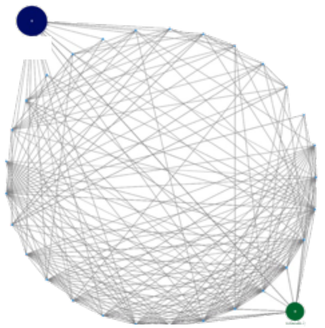


Blogs, Fake News, and Influence Operations. Digital Hydra: False Information Online as a Weapon, **NATO StratCom COE.**

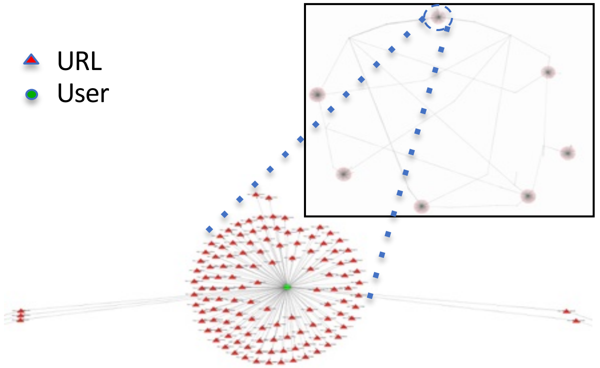
Human-AI Agent Coordination Evolution

From bots → botnets → social bots

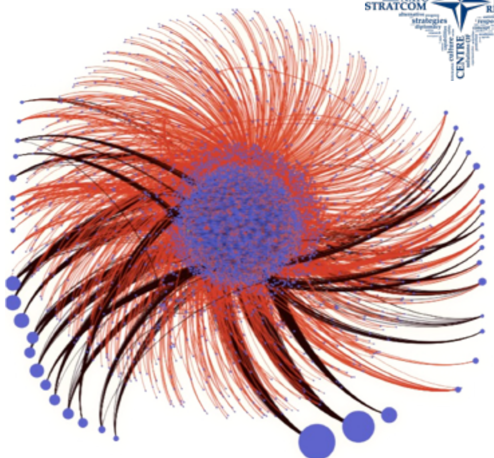
Published by NATO StratCom COE



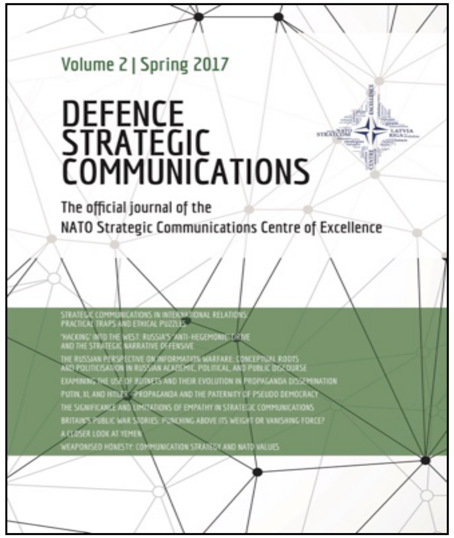
Mutual reciprocity.
IFYFM & *FMIFY*



No mutual reciprocity.
Coordination in information network



Organizational hierarchy.
Core and peripheral bots


Volume 2 | Spring 2017

DEFENCE STRATEGIC COMMUNICATIONS

The official journal of the NATO Strategic Communications Centre of Excellence

Examining the Use of Botnets and their Evolution in Propaganda Dissemination. **Journal of NATO Defence Strategic Communications**

COORDINATION/BEHAVIOR COMPLEXITY

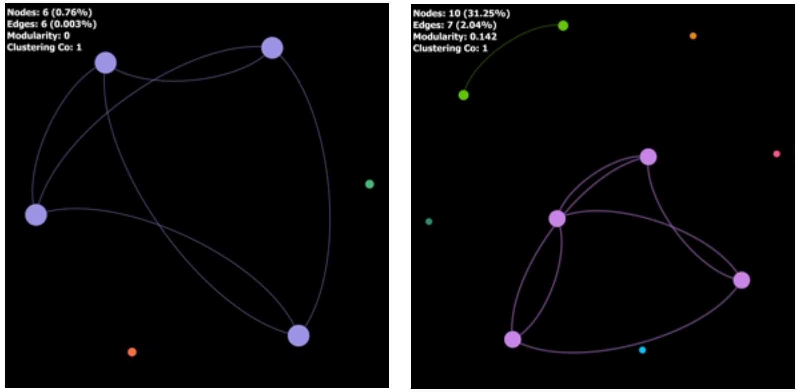
Crimean Invasion 2014

Dragoon Ride 2015

Trident Juncture 2015

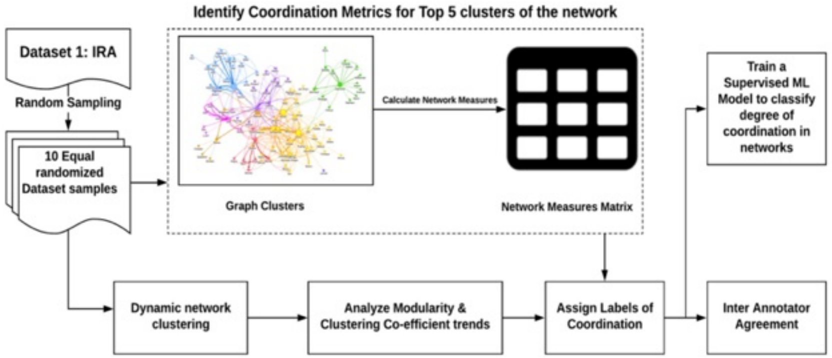
ISIS Propaganda 2016

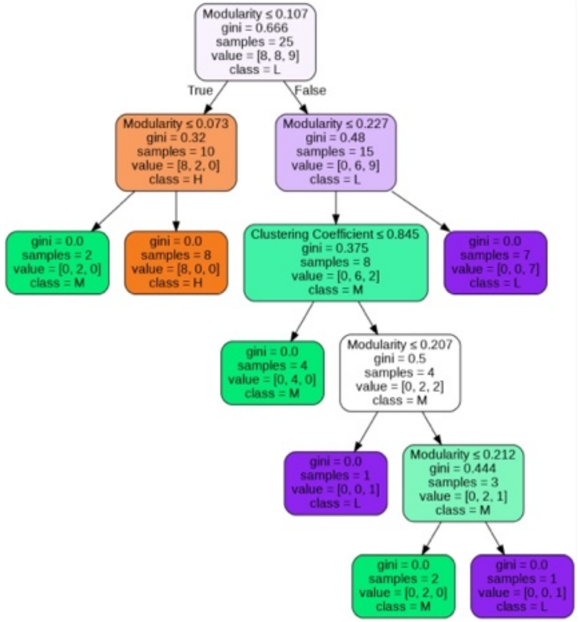
Modeling Botnet Coordination

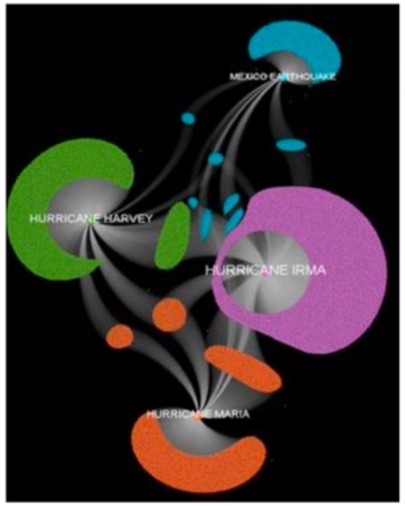


IRA Twitter bot data released by US Intelligence Agencies

Detecting coordination among Twitter social bots (Khaund & Agarwal, 2021) IEEE Transactions on Computational Social Systems (TCSS).

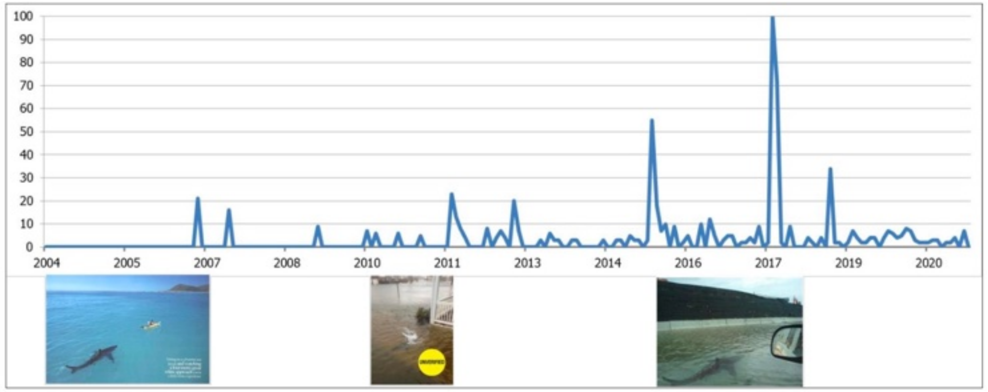


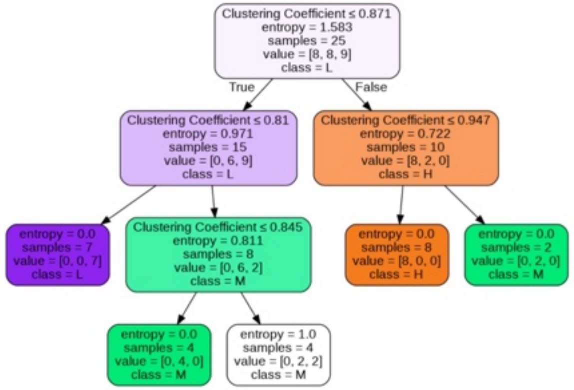




Language	Hashtag (Translation)
English	#DACA, #BlackLivesMatter
Spanish	#VenezuelaDemocraciaYDiálogo (Venezuela Democracy and Dialogue), #Cáncer (Cancer)
Arabic	(The demise Of Israel) #زوال_إسرائيل, (The Jews) #اليهود
French	#Nucléaire (Nuclear), #GendarmerieEnOpération (Gendarmerie Special Operations)
Mandarin	金正恩 (Kim Jong-un), 核试验 (Nuclear Test)

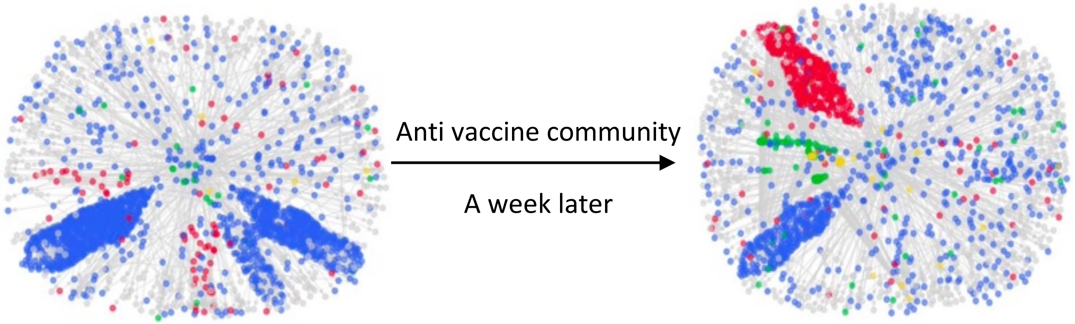
Bot coordinated misinformation during Hurricanes

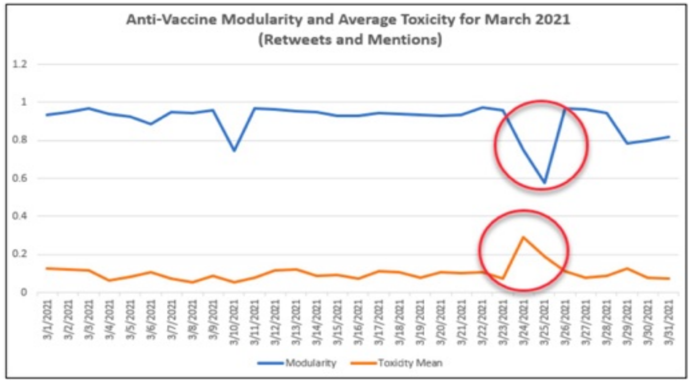


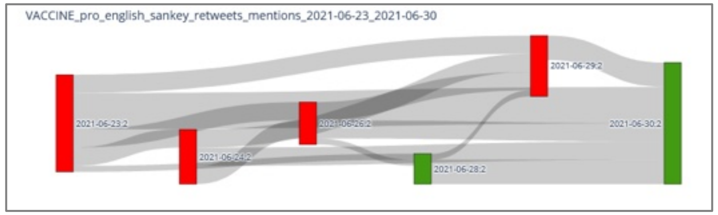


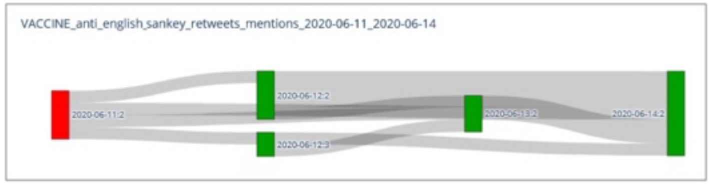
Toxicity, Community Dynamics, and BOTS!

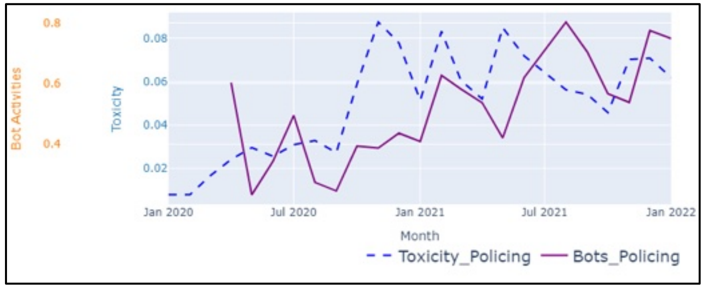
- Toxic discourse causes disruption and polarization/ segregation among communities.
- Community splinters when toxicity rises.
- Granger causality test suggests the effect is strongest after two days.
- Bot activity found to be positively correlated with toxicity.
- Higher toxicity leads to community fracturing.
- This work shows a way to measure impact of bots.

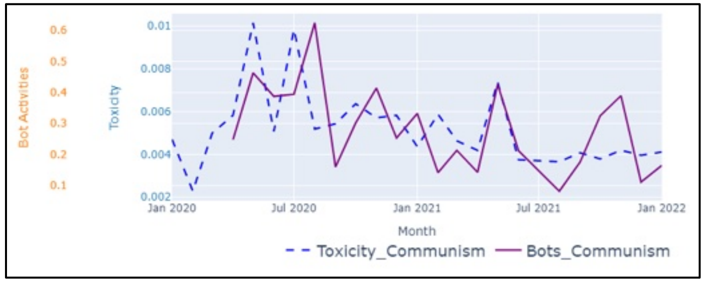




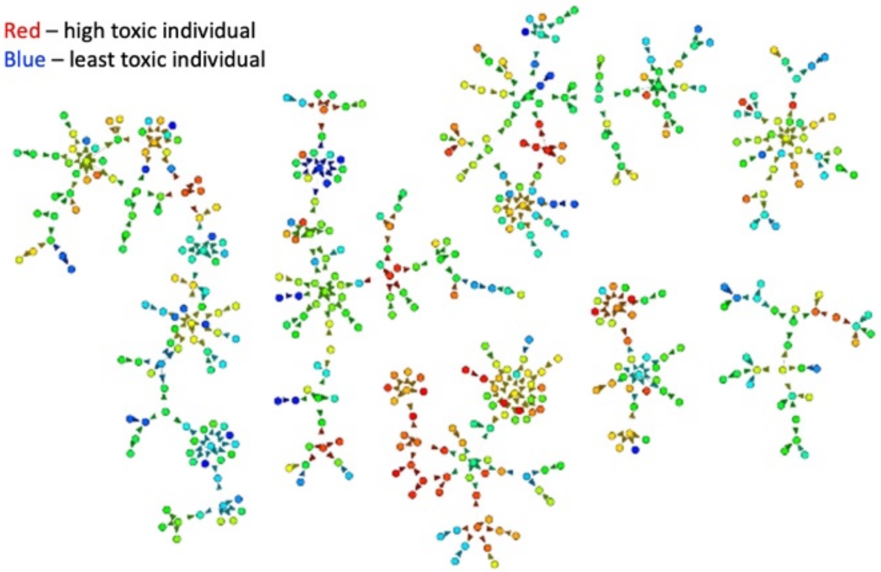








Networks and Toxicity



Toxicity analysis on YouTube commenters. Toxic discourse causes disruption and polarization/segregation among communities, as seen above. We demonstrate that by removing highly toxic users from a network, hate speech reduces, online discourse improves, and fractured communities heal. Our findings offer guidance to policymakers within each online social network to make informed decisions about the information environment and derive appropriate and timely countermeasures to continue providing a healthy platform for their users.

STRS epidemiological model for toxicity propagation – Susceptible (S); Toxic (T); Recovered (R); Susceptible (S)

We demonstrate that by removing highly toxic users from a network, hate speech reduces, online discourse improves, and fractured communities heal. Our findings offer guidance to policymakers within each online social network to make informed decisions about the information environment and derive appropriate and timely countermeasures to continue providing a healthy platform for their users.

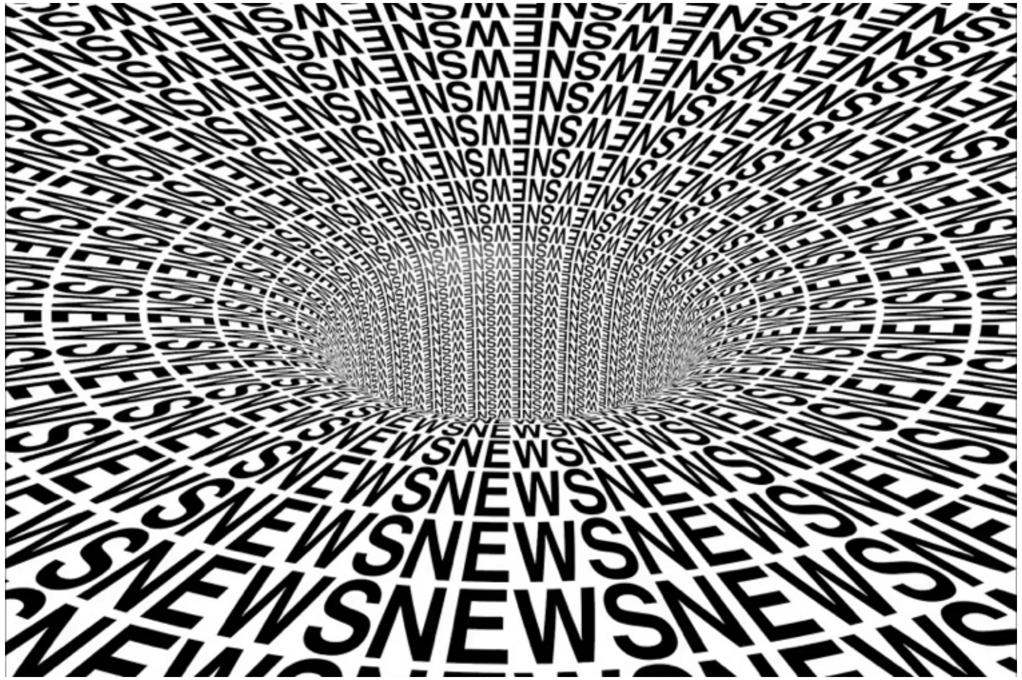
Experimental simulation	Toxicity score	Percentage reduction
Removal of top 10 users with high Betweenness centrality	0.720981759	0.21
Removal of top 10 users with high PageRank centrality	0.722317191	0.02
Removal of users with toxicity scores greater than 0.8	0.641927323	11.15

Agarwal et al. (2022). Applying an Epidemiological Model to Evaluate the Propagation of Toxicity related to COVID-19 on Twitter. The 55th Hawai'i International Conference on System Sciences (HICSS), January 4-7,, Maui.
Agarwal, N. et al. (2021) Developing a Socio-Computational Approach to Examine Toxicity Propagation and Regulation in COVID-19 Discourse on YouTube. *Information Processing and Management Special issue on Dis/Misinformation Mining from Social Media.* Vol. 58, Issue 5, 2021. Elsevier.

Working with LinkedIn and Arkansas Office of the Attorney General

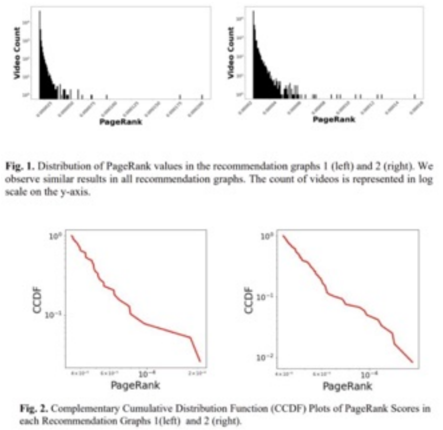
Platform Algorithmic 'Recommender' Bias

- AI-based recommendation algorithms that predict our shopping behaviors, books and articles to read, videos to watch lack transparency.
- Recommendation algorithm learns from behavioral data and perpetuates the underlying bias in its recommendations.
  - YouTube's recommendation algorithm is known to push its viewers down the conspiratorial rabbit hole by suggesting related videos.
  - On Facebook, ads to recruit delivery drivers for Domino's Pizza Inc. were disproportionately shown to men, while women were more likely to receive notices in recruiting shoppers for grocery-delivery service Instacart Inc.
- Explainable model could help in identifying causes of biased recommendations thereby enhancing the model's transparency.

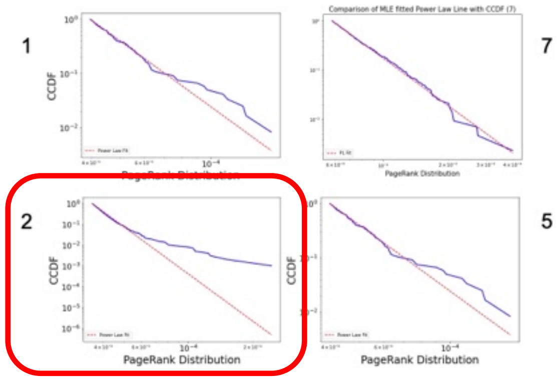


Characterizing Bias and Impact Assessment

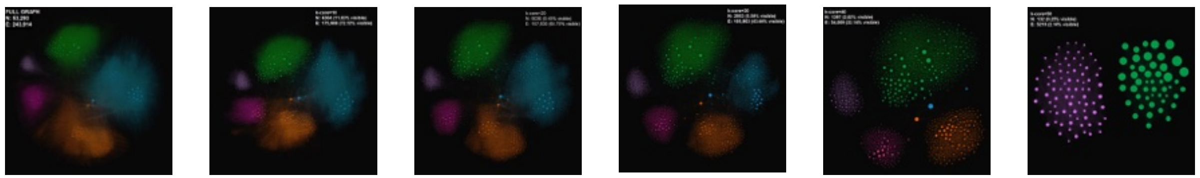
- Channel characterization based on implicit algorithmic bias requires identification and measurement of bias.
  - Power law distribution fitting (exponent)
  - Gini coefficient - a single number that demonstrates a degree of inequality in a distribution of income/wealth.
- Characterization based on <u>context</u> -> COVID-19 context showed largest implicit algorithmic bias, marked in red.
- Characterization based on <u>impact</u> -> Implicit bias resulted in extremist content communities (far left and far right information bubbles).



Topic drift and decrease in relevance was observed.



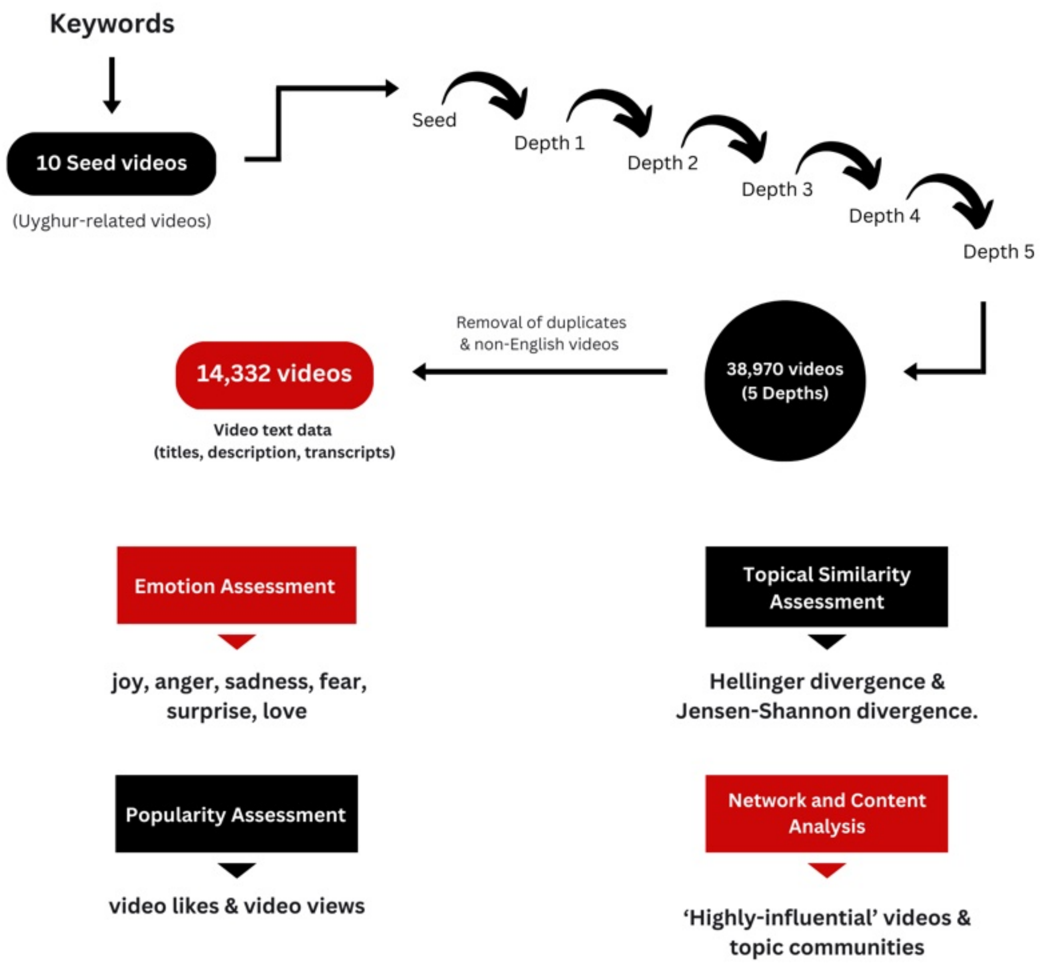
Top PageRank videos were removed weeks or months after their appearance in the recommendation network. Reason for content removal is violation of platform terms and services.



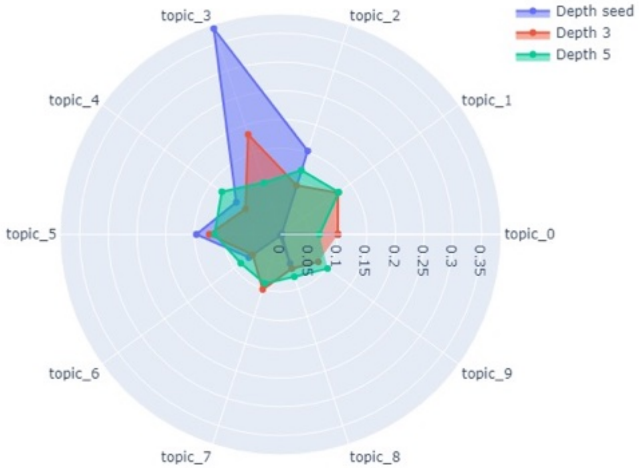
Far-left and far-right content communities/information bubbles resulting due to biased recommendations

Baris Kirdemir and Nitin Agarwal. Exploring Bias and Information Bubbles in YouTube's Video Recommendation Networks. The 10th International Conference on Complex Networks and their Applications (COMPLEX NETWORKS 2021), November 30 – December 2, 2021. Madrid, Spain

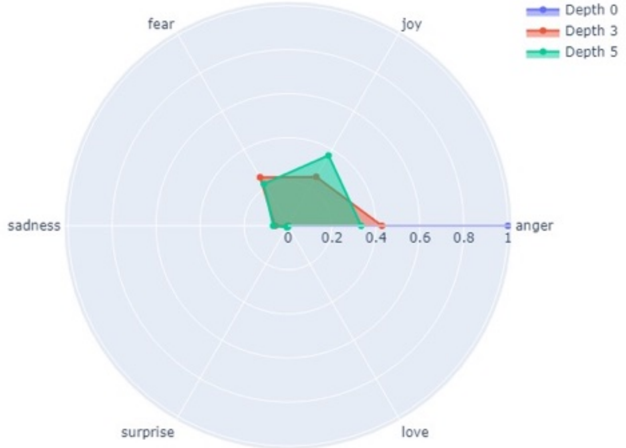
Exploiting Algorithmic Bias – Uyghur Narrative



Topic Drift



Emotion Drift

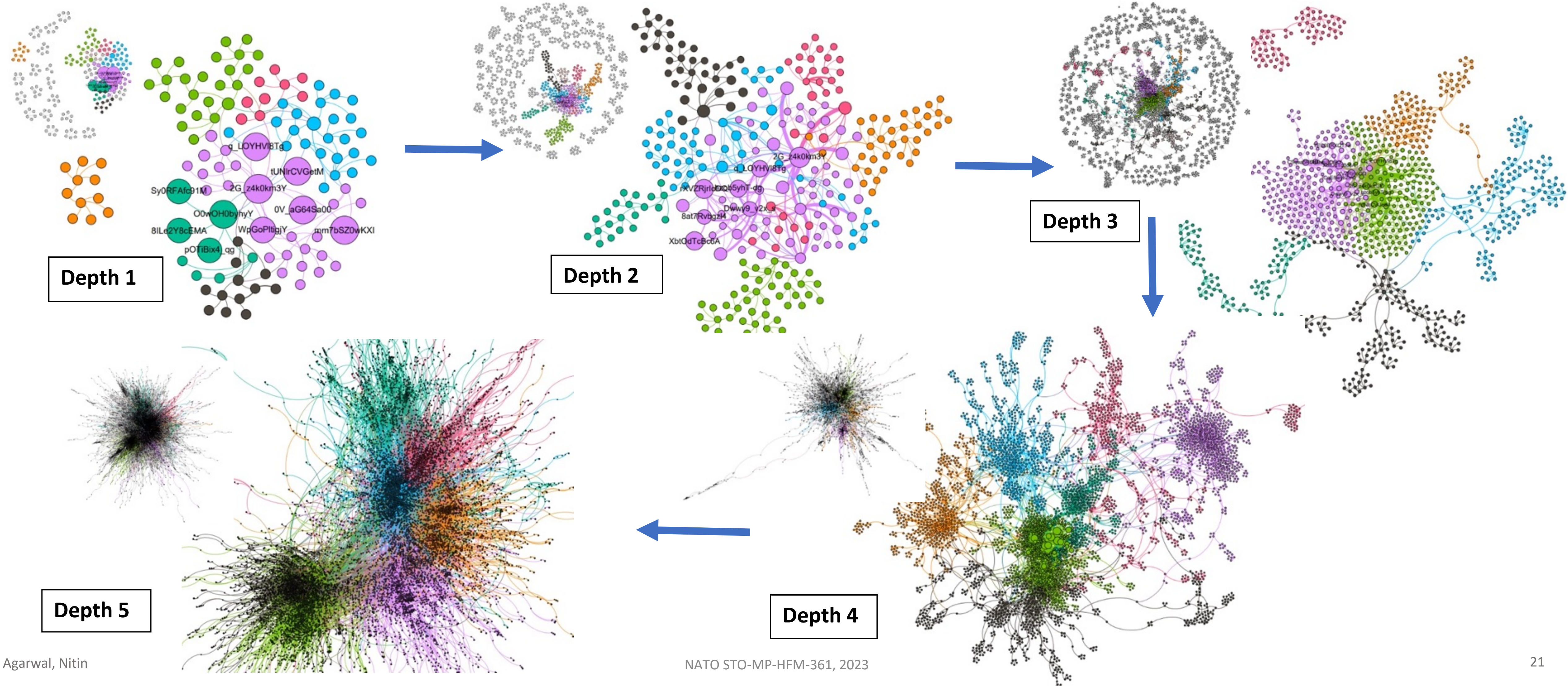


Agarwal et al. (2023). 45th European Conference on Information Retrieval (ECIR 2023), April 2 – 6, 2023, Dublin, Ireland.
Agarwal et al. (2023). 9th International Conference on Human and Social Analytics (HUSO 2023), March 13 – 17, 2023, Barcelona, Spain.

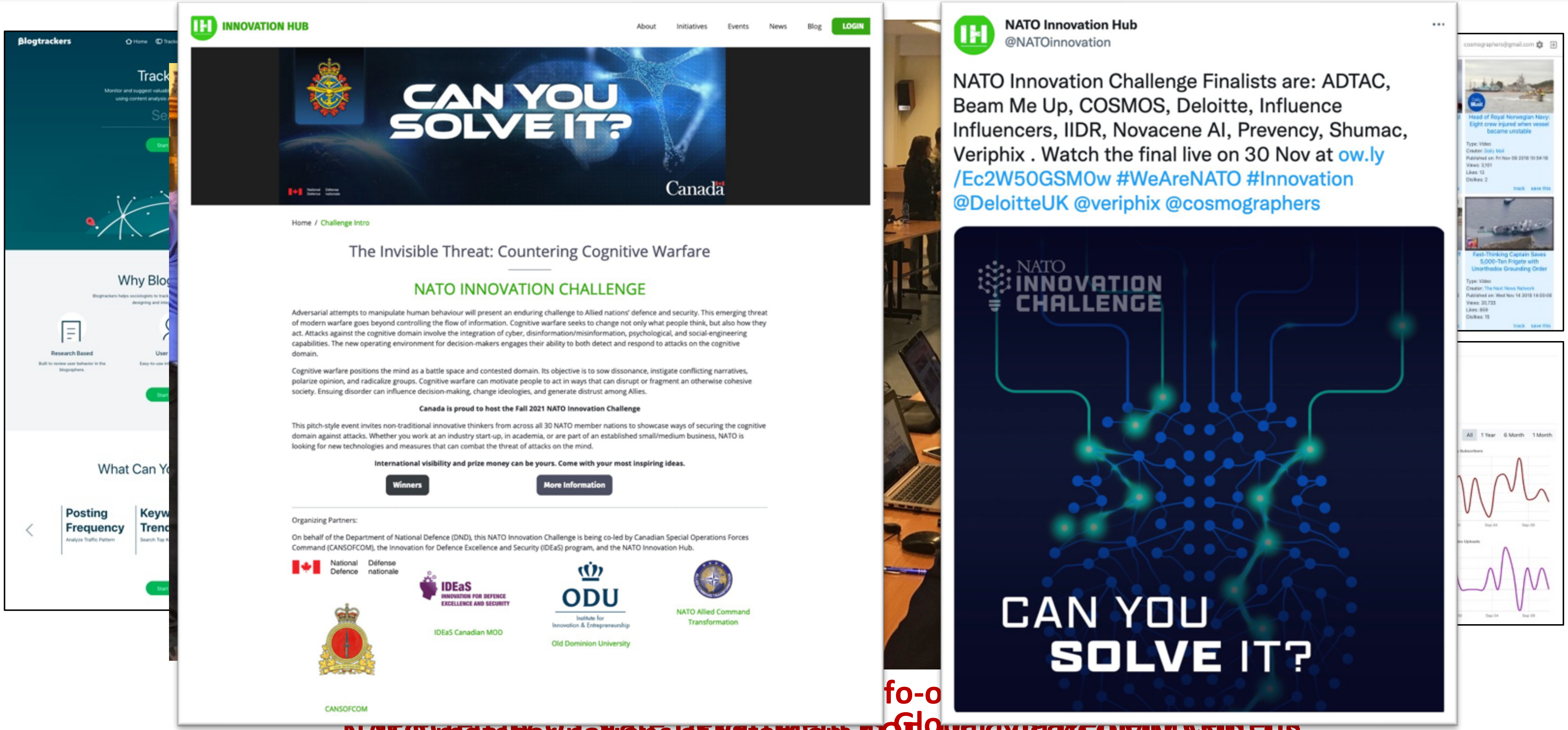
UALR

Recommendation Network Analysis

COSMOS



Technologies & Innovation

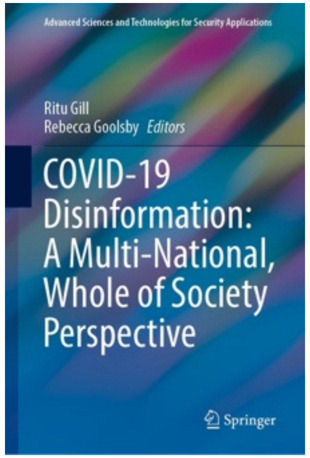


NATO member nations at StratCom COE, US CYBERCOMMAND, US Marine Information Group (3-MIGs)

Recognized as top 10 COSMOS in the 2021 NATO Innovation Hub Challenge on "The Invisible Threat: Countering Cognitive Warfare" from 132 teams across the 30 NATO member nations.

Technologies, Innovation, & Policymaking

- COVID-19 Misinformation Tracker recognized by the **World Health Organization (WHO)** as one of the key technological innovations developed across the world to address COVID-19 pandemic.
- The capability was developed in partnership with the **Arkansas Office of the Attorney General.**
- The application leverages our work on socio-cognitive threat modeling, education, and awareness efforts to assist policymakers.



NATO Research and Technology Group (RTG HFM-293), 2022

Shared research with

- **U.S. Senator John Boozman** (Member, Senate Appropriations Committee)
- **U.S. Senator Ed Markey** (Member, Senate Committee on Foreign Affairs; Chairman, Subcommittee On East Asia, The Pacific, And International Cybersecurity Policy; Member, Subcommittee On Near East, South Asia, Central Asia, And Counterterrorism)
- **U.S. Senator Elizabeth Warren** (Member, Senate Committee on Armed Services, Subcommittee on Emerging Threats and Capabilities)

Thank You!

COSMOS Tools:

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

Follow **cosmographers** on

https://cosmos.ualr.edu/

Nitin Agarwal, Ph.D.
Director, COSMOS Research Center
nxagarwal@ualr.edu