

# The BotBuster Universe

Lynnette Hui Xian Ng

The CMU centers for:

Informed DEMocracy And Social cyber-security

Computational Analysis of Social and Organizational Systems



**Carnegie Mellon University**



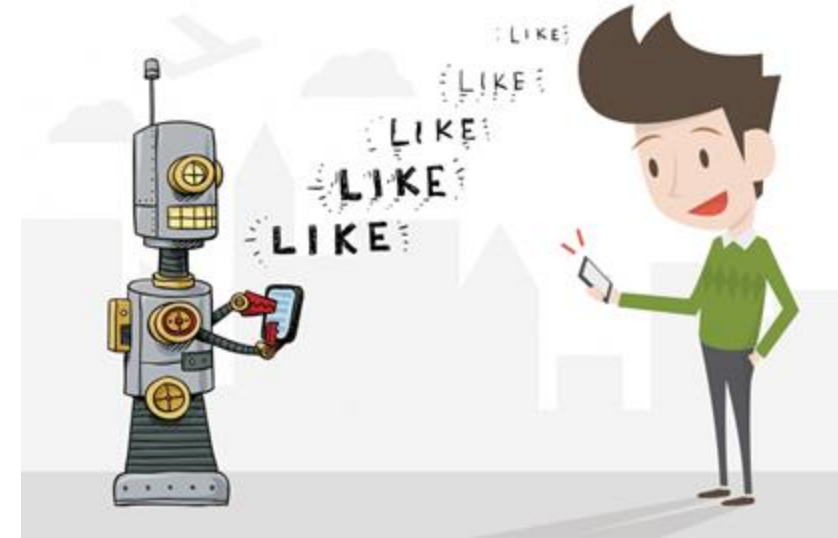
# What is a Bot?



A Robot



LLM ChatBot

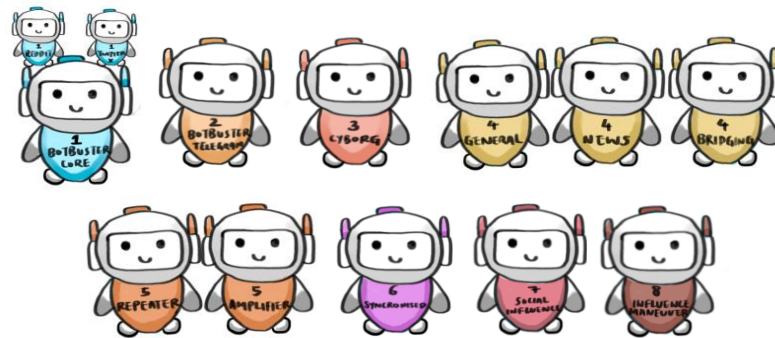


Social Media Bot

# What are social media bots?

- ❑ Social Media Bot, or Bot, *noun*.

A **programmable account** that automatically carries out a **series of mechanics** on **social media platforms**



# Why do we care about Bots?

- ❑ Bots actively participate in information campaigns, drawing attention to their cause
- ❑ Inauthentic account detection (bot) is crucial to social media analysis because:
  - ❑ Find mis/dis-information narratives
  - ❑ Study conversation manipulation
  - ❑ Characterize information operations
  - ❑ Identify coordinated behavior

# Bots are more active than humans

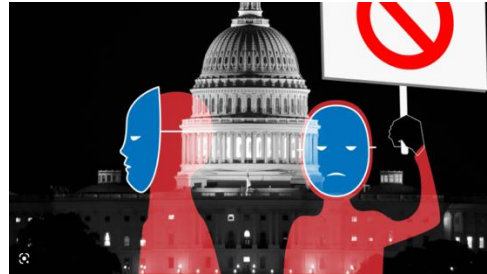
- ❑ Bot posts 2x more than humans
- ❑ Bot posts resonate  $\sim 1.16x$  more
- ❑ Bot posts are engaged with for up to  $\sim 6.89x$

**Ng, Lynnette Hui Xian**, Wenqi Zhou, and Kathleen M. Carley. "Exploring Cognitive Bias Triggers in COVID-19 Misinformation Tweets: A Bot vs. Human Perspective." *arXiv preprint arXiv:2406.07293* (2024).

# Social Media Bots influence our digital conversations



Bots affect businesses

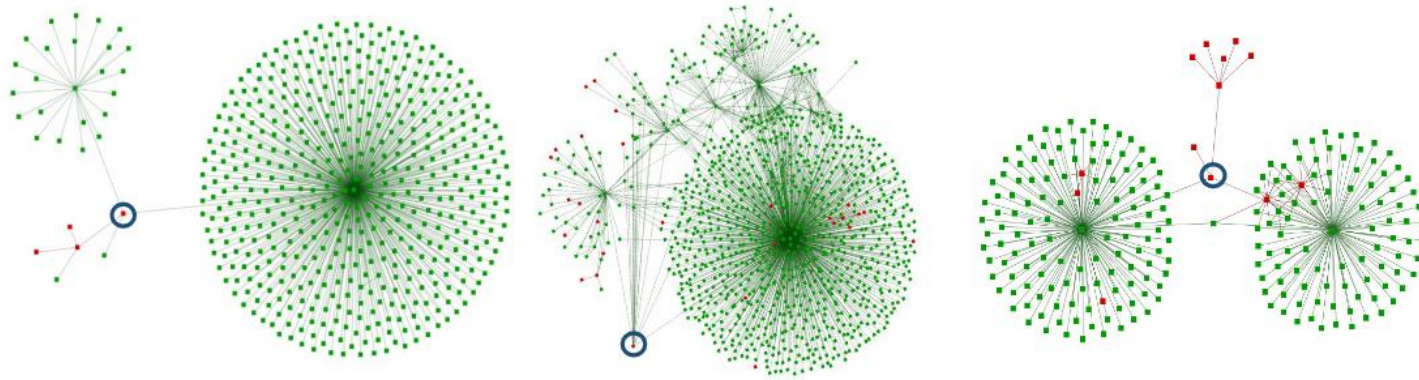


Bots affect elections



Bots affect entertainment

# Bots can pressure you to change your stance

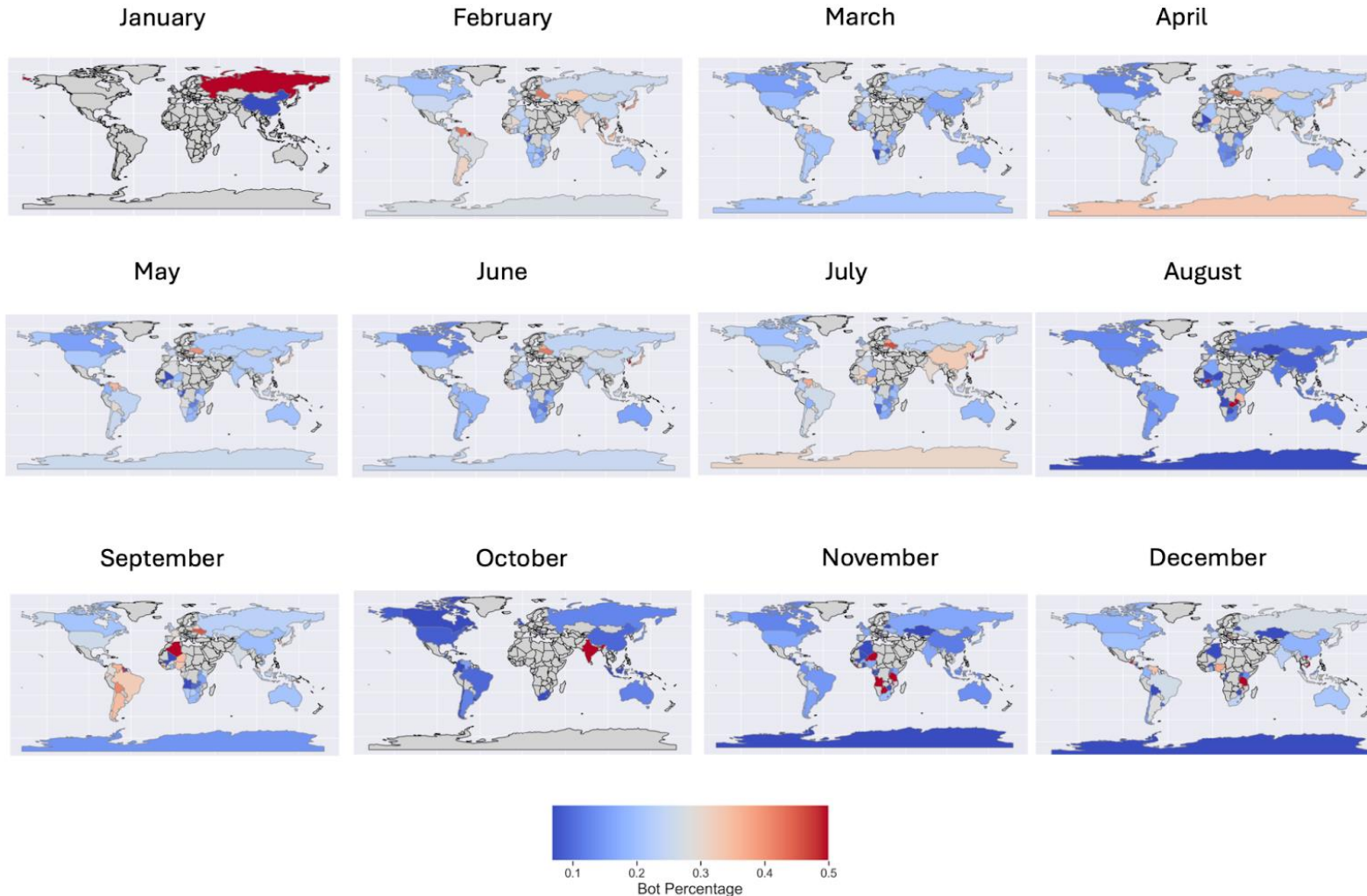


**Ng, Lynnette Hui Xian,** and Kathleen M. Carley. "Pro or anti? A social influence model of online stance flipping." *IEEE Transactions on Network Science and Engineering* 10, no. 1 (2022): 3-19.



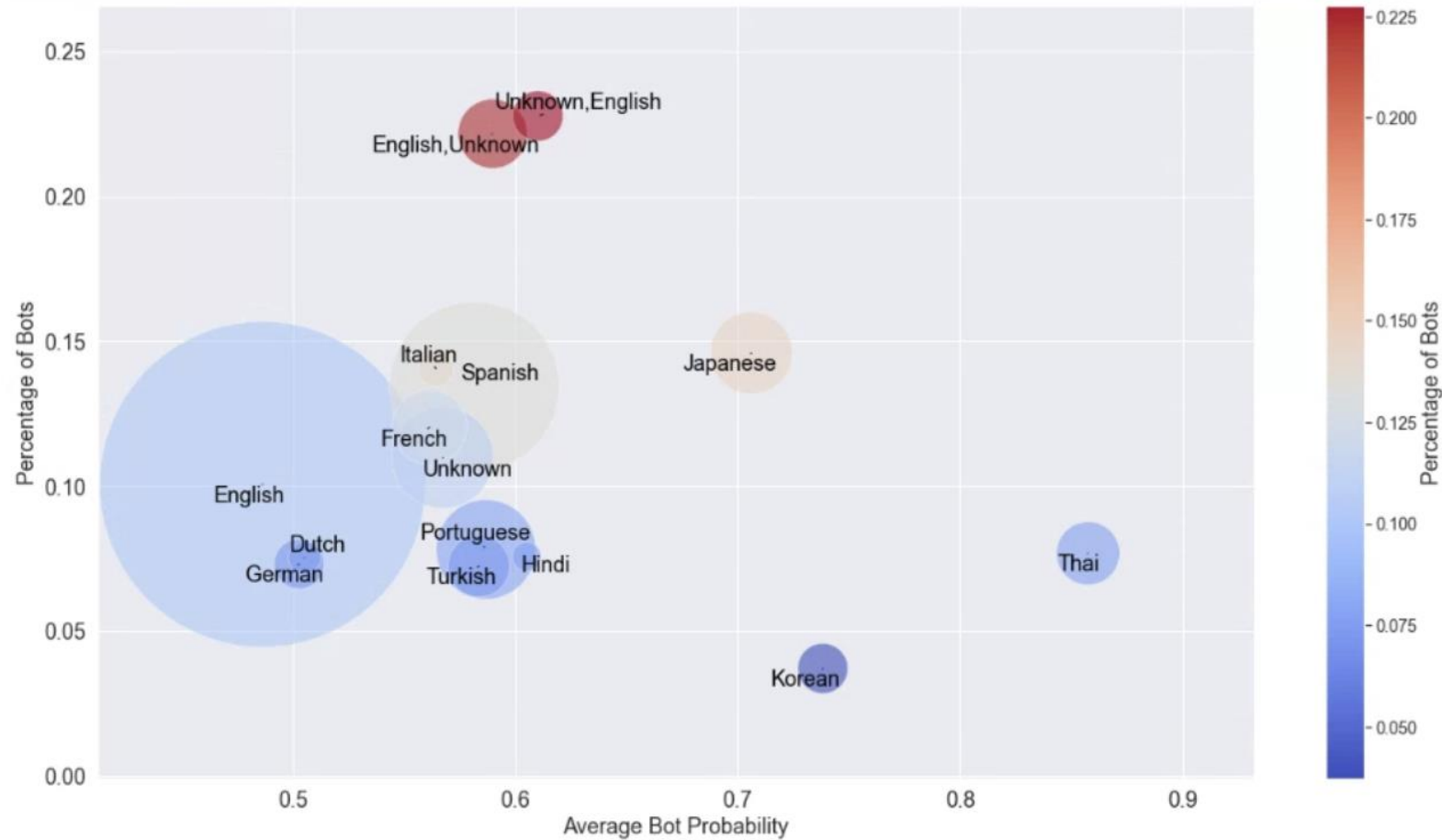
# Where are Bots found?

Bot Percentage by Country in 2021

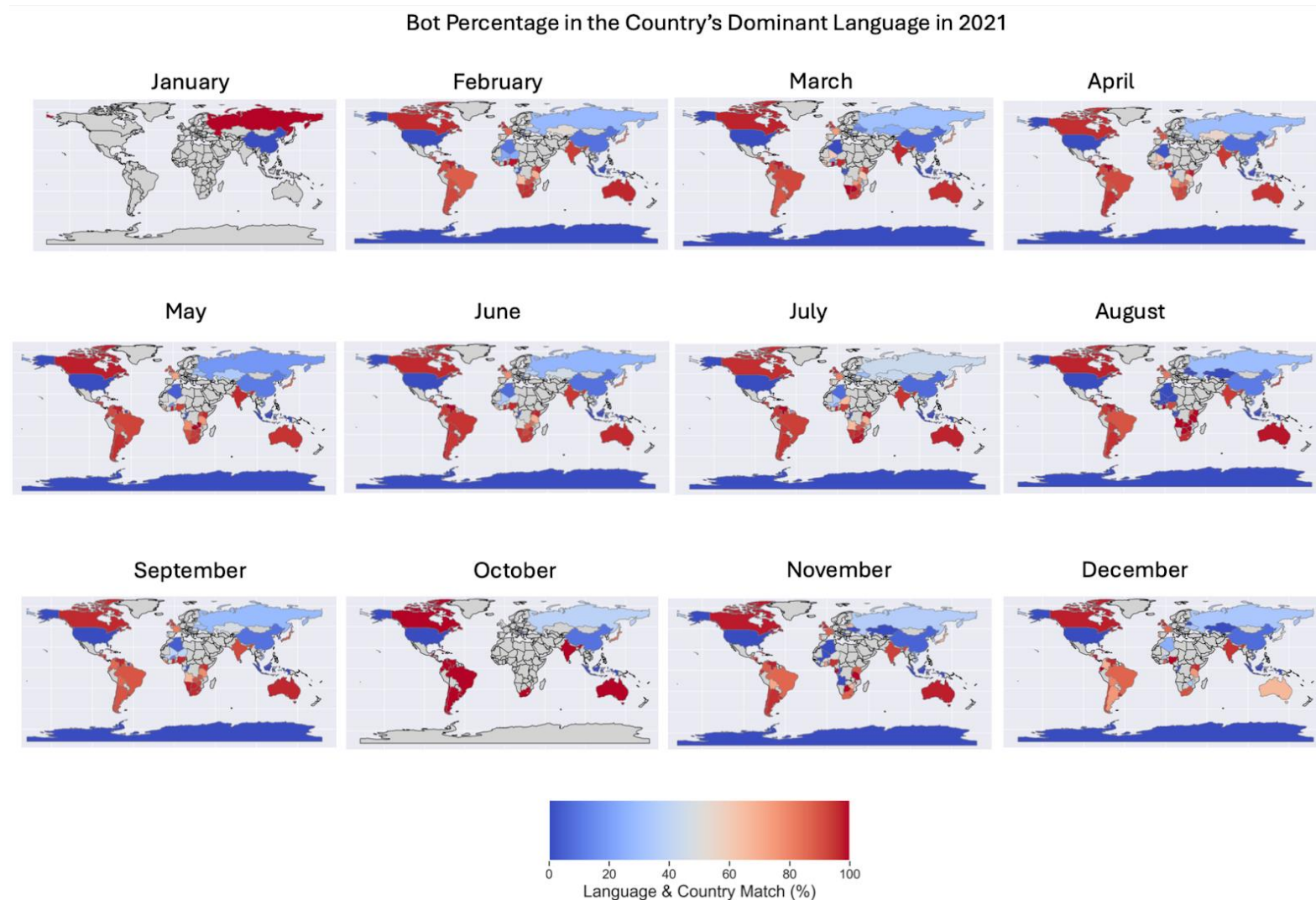




# What do Bots speak?

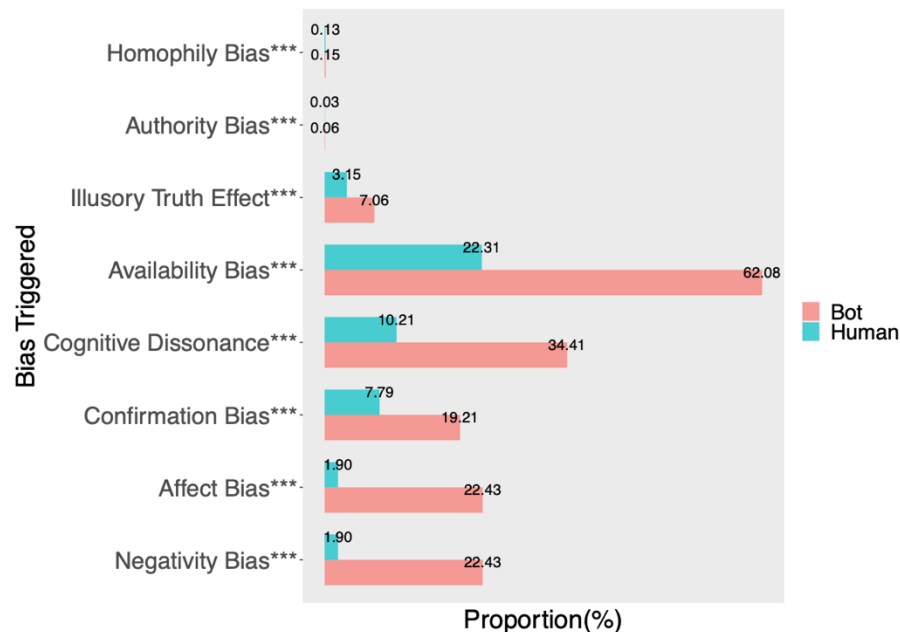


# What do bots speak?



# Why are Bots successful?

- Bots use more triggers of cognitive biases in their posts

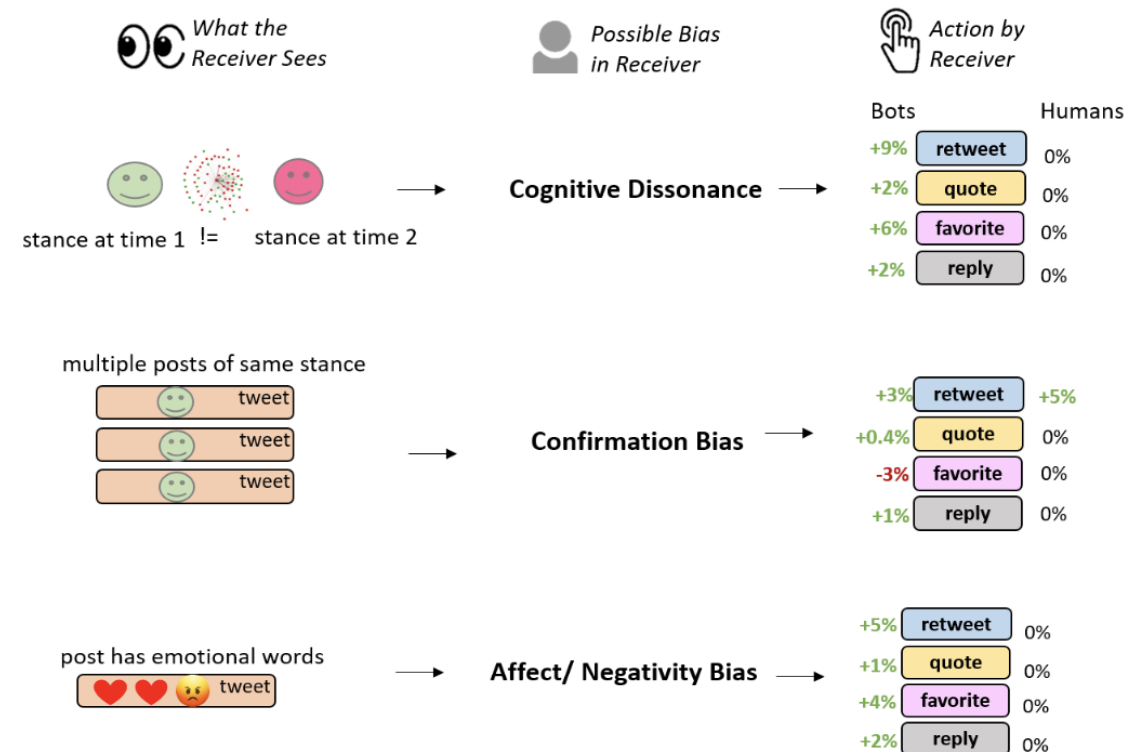


Judgment by Representativeness	The similarity between two groups represent the degree which one is representative of the other
Homophily Bias	Increased acceptance of a message if it is shared to the same community
Authority Bias	Perception that message is credible if it is from reputable sources
Judgment by Availability	An event is judged likely if it is easy to imagine or memorable
Availability Cascade	Collective belief gains more plausibility through increased repetition
Illusory Truth Effect	Tendency to believe in a story more easily when it is repeated multiple times
Affect Bias	Emotions rather than information have a disproportionate effect on judgment
Negativity Bias	Higher belief in negative than positive news
Judgment by Anchoring	Where an anchor is used as an approximation towards the judgment
Cognitive Dissonance	Tendency to avoid having conflicting beliefs and attitudes
Confirmation Bias	Favor information that conforms and strengthen prior beliefs

Ng, Lynnette Hui Xian, Wenqi Zhou, and Kathleen M. Carley. "Exploring Cognitive Bias Triggers in COVID-19 Misinformation Tweets: A Bot vs. Human Perspective." *arXiv preprint arXiv:2406.07293* (2024).

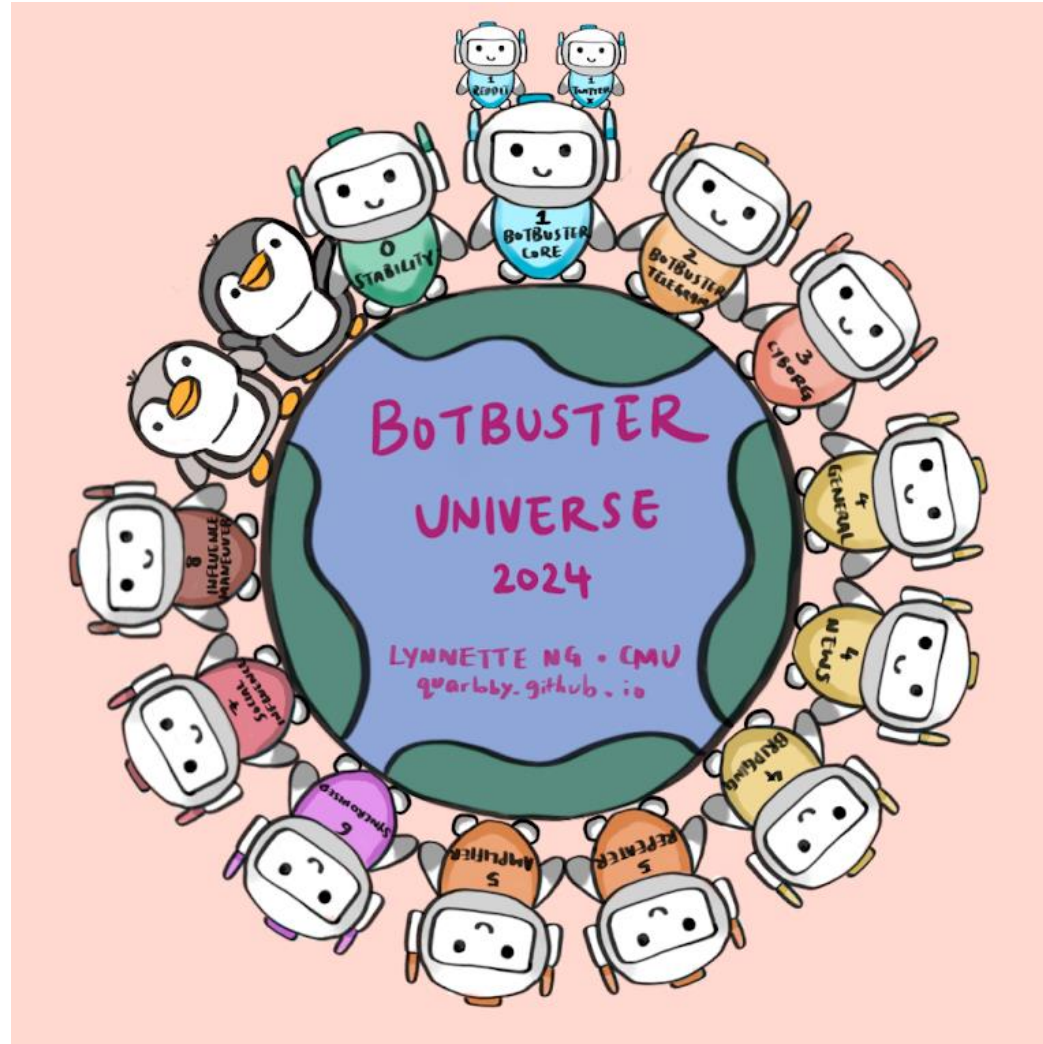
# Why are Bots successful?

- ❑ Bots use more triggers of cognitive biases in their posts
- ❑ Resulting in increased engagement



Ng, Lynnette Hui Xian, Wenqi Zhou, and Kathleen M. Carley. "Exploring Cognitive Bias Triggers in COVID-19 Misinformation Tweets: A Bot vs. Human Perspective." *arXiv preprint arXiv:2406.07293* (2024).

# The BotBuster Universe



*Also available in a  
postcard!*

*Come get one!*

*(While stock lasts)*

# The BotBuster Algorithm

- ❑ Multi-platform mixture-of-experts bot detection algorithm
- ❑ Twitter (X), Reddit, Instagram, Telegram

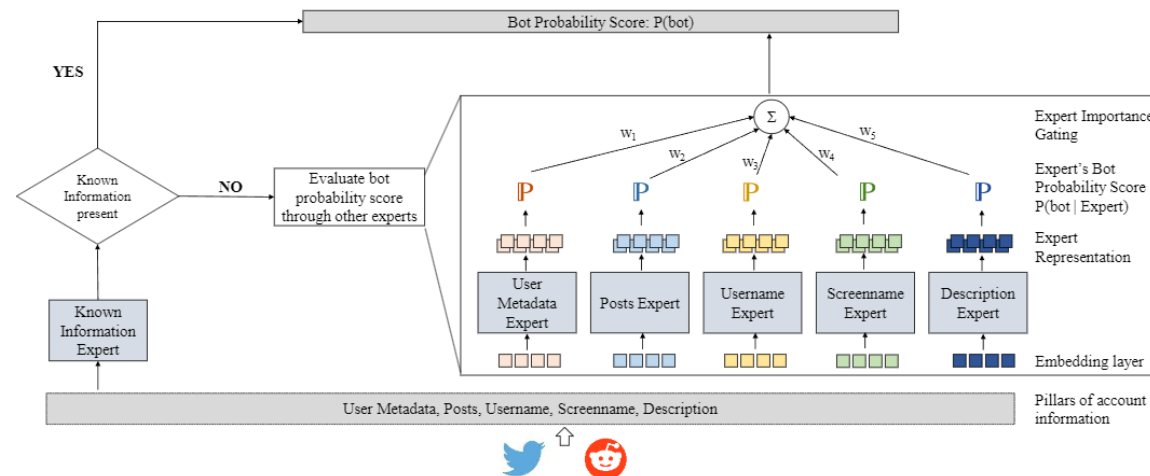


Figure 1: Diagram of the BotBuster Architecture.

Ng, Lynnette Hui Xian, and Kathleen M. Carley. "Botbuster: Multi-platform bot detection using a mixture of experts." In *Proceedings of the international AAAI conference on web and social media*, vol. 17, pp. 686-697. 2023.



# The BotBuster Algorithm

- ❑ Each data field is trained/tested by a separate expert
- ❑ Expert predictions are aggregated together to form final prediction
- ❑ Handles incomplete data
  - ❑ Expert is not activated and probability is based on remaining fields
  - ❑ Can analyze 100% of the dataset

**Ng, Lynnette Hui Xian,** and Kathleen M. Carley. "Botbuster: Multi-platform bot detection using a mixture of experts." In *Proceedings of the international AAAI conference on web and social media*, vol. 17, pp. 686-697. 2023.



# Bot Detection Algorithms are constantly being improved

- Because bot detection is a hard problem
  - Bots and humans are very similar
  - Bots get more similar to humans over time
  - Who copied who?

**Ng, Lynnette Hui Xian, and Kathleen M. Carley.** "Botbuster: Multi-platform bot detection using a mixture of experts." In *Proceedings of the international AAAI conference on web and social media*, vol. 17, pp. 686-697. 2023.

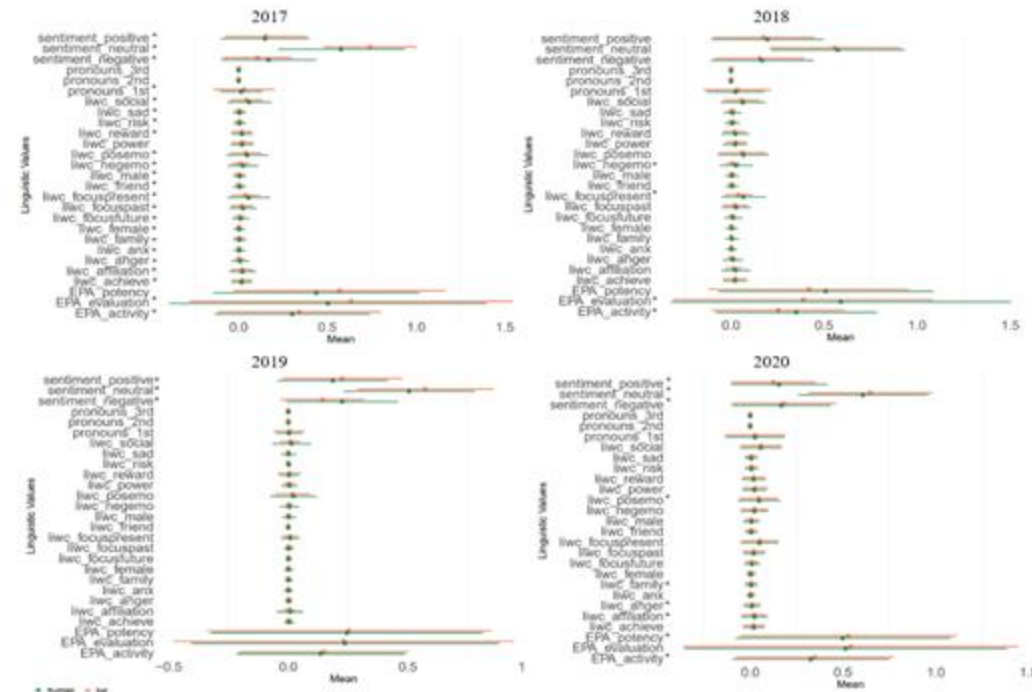
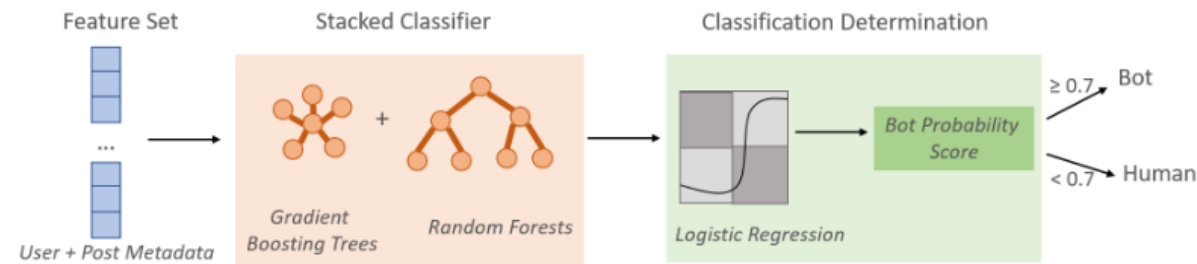


Figure 5: Change in linguistic features over years of bot datasets. \* denotes significant difference ( $p < 0.05$ ) between bot/human

# Tiny-BotBuster

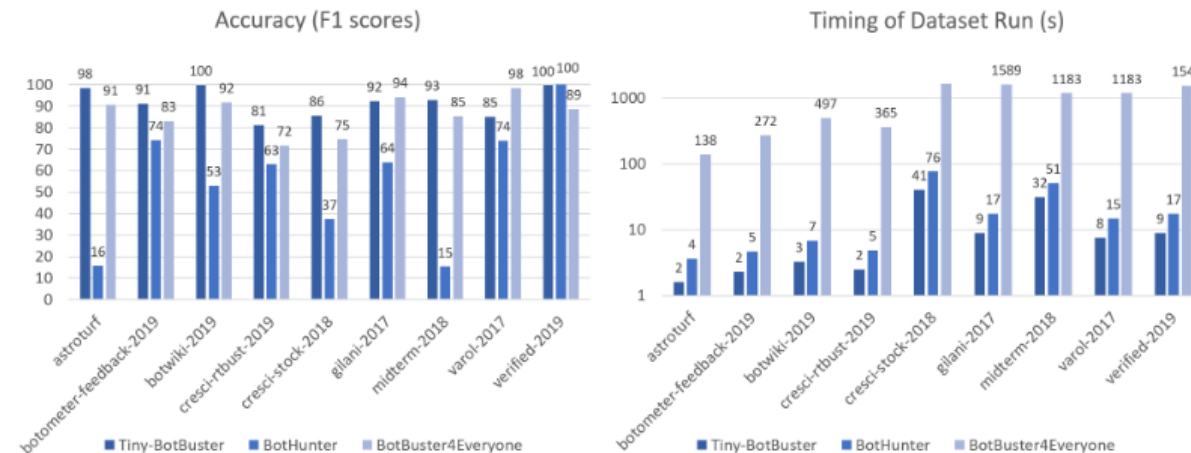
- Better, smaller and faster



**Ng, Lynnette Hui Xian,** Mihovil Bartulovic and Kathleen M. Carley. "Tiny-BotBuster: Identifying Automated Political Coordination in Digital Campaigns" In *Proceedings of the international AAAI conference on web and social media*, vol. 17, pp. 686-697. 2023.

# Tiny-BotBuster

- Better, smaller and faster



Ng, Lynnette Hui Xian, Mihovil Bartulovic and Kathleen M. Carley. "Tiny-BotBuster: Identifying Automated Political Coordination in Digital Campaigns" In *Proceedings of the international AAAI conference on web and social media*, vol. 17, pp. 686-697. 2023.

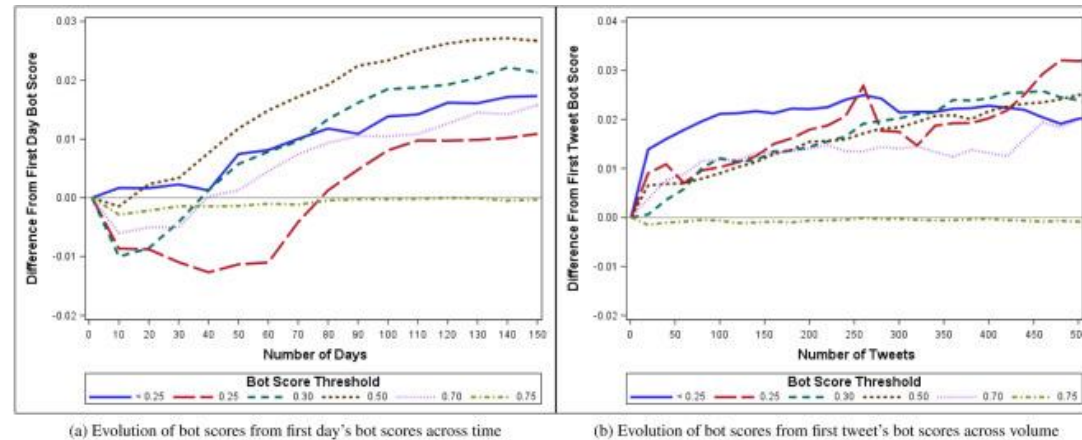
# What value is a Bot?

- ❑ Bot detection algorithms return a value between  $[0, 1]$
- ❑ Represents probability of the user being a bot
- ❑ But...
- ❑ What probability is a good threshold?
  - ❑ What is a good threshold for a consistent score?
  - ❑ What is a minimum number of tweets for a consistent score?

Ng, Lynnette Hui Xian, Dawn C. Robertson, and Kathleen M. Carley. "Stabilizing a supervised bot detection algorithm: How much data is needed for consistent predictions?." *Online Social Networks and Media* 28 (2022): 100198.

# Bot Detection Thresholds

- ❑ Large-scale longitudinal analysis of tweets and probability scores
- ❑ A good threshold is 0.70
- ❑ Minimally requires 20 tweets for a stable score



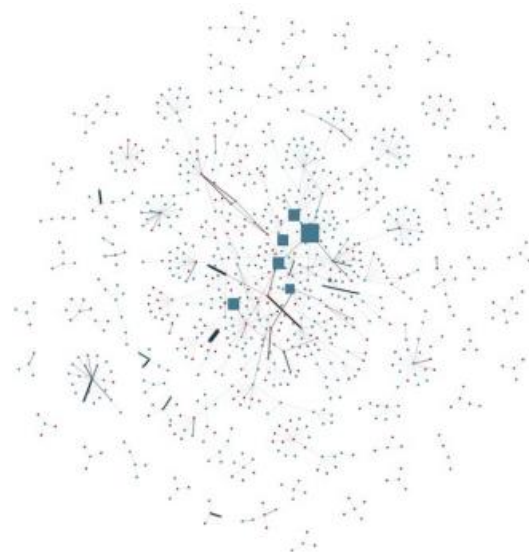
Ng, Lynnette Hui Xian, Dawn C. Robertson, and Kathleen M. Carley. "Stabilizing a supervised bot detection algorithm: How much data is needed for consistent predictions?." *Online Social Networks and Media* 28 (2022): 100198.

# BotBuster Telegram

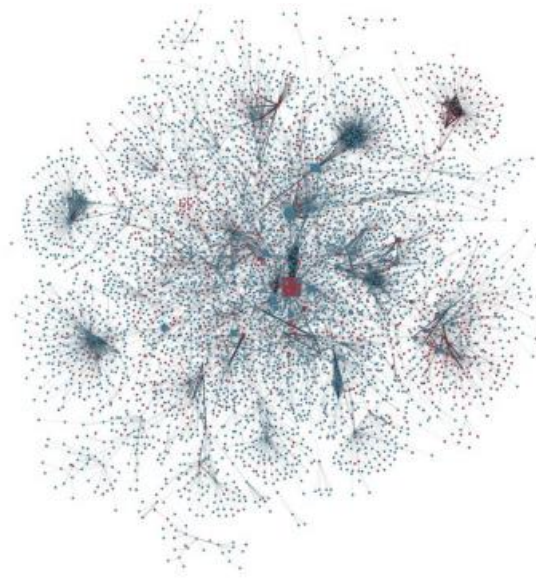
- ❑ Perception-based bot detection algorithm
  - ❑ Relies mostly on the text posted
  - ❑ Mimics how human users determine bots: i.e., through visible information (texts)
- ❑ Works with 73% accuracy
- ❑ Examples of bot-like messages/ users
  - ❑ Copying & pasting links in multiple messages, where bulk of the links are same except for URL parameters and query strings
  - ❑ Posting exact same links
  - ❑ Repeating same messages multiple time
  - ❑ Capital letters & same number of exclamation marks for message sent (“THAT’S RIGHT, THANK YOU!!!”)
  - ❑ Repetitive and short messages (“Inbox me please”)

# BotBuster Telegram

- ❑ Disinformation spread in COVID data
- ❑ Bot users essential in sustaining conversations
- ❑ Human users crucial in forwarding disinformation



(a) User x User network by Message Forwarding



(b) User x User network by Message Reply



# Cyborgs

- ❑ Half-human, half-bots
  - ❑ Appears as bot in one timeframe, appears as humans in another
  - ❑ Controlled by human operators in some instances, and automated scripts in others
- ❑ Used by activists and influential people
- ❑ Used for strategic communication purposes

# Cyborgs

- ❑ Identified with two key properties:
- ❑ (1) Changes in bot classification output
  - ❑ Frequent flipping of bot classification agent, thus changing their bot/human labels from timeframe to timeframe
  - ❑ At least 3 flips
- ❑ (2) Difference in bot-likeness scores that provide definitive bot/human classification
  - ❑ Large difference in bot likelihood scores between flips
  - ❑  $\geq 0.10$  change

# Cyborgs

- ❑ Lower proportion suspended
- ❑ Demonstrates ability to evade detection, or the awareness of the platform

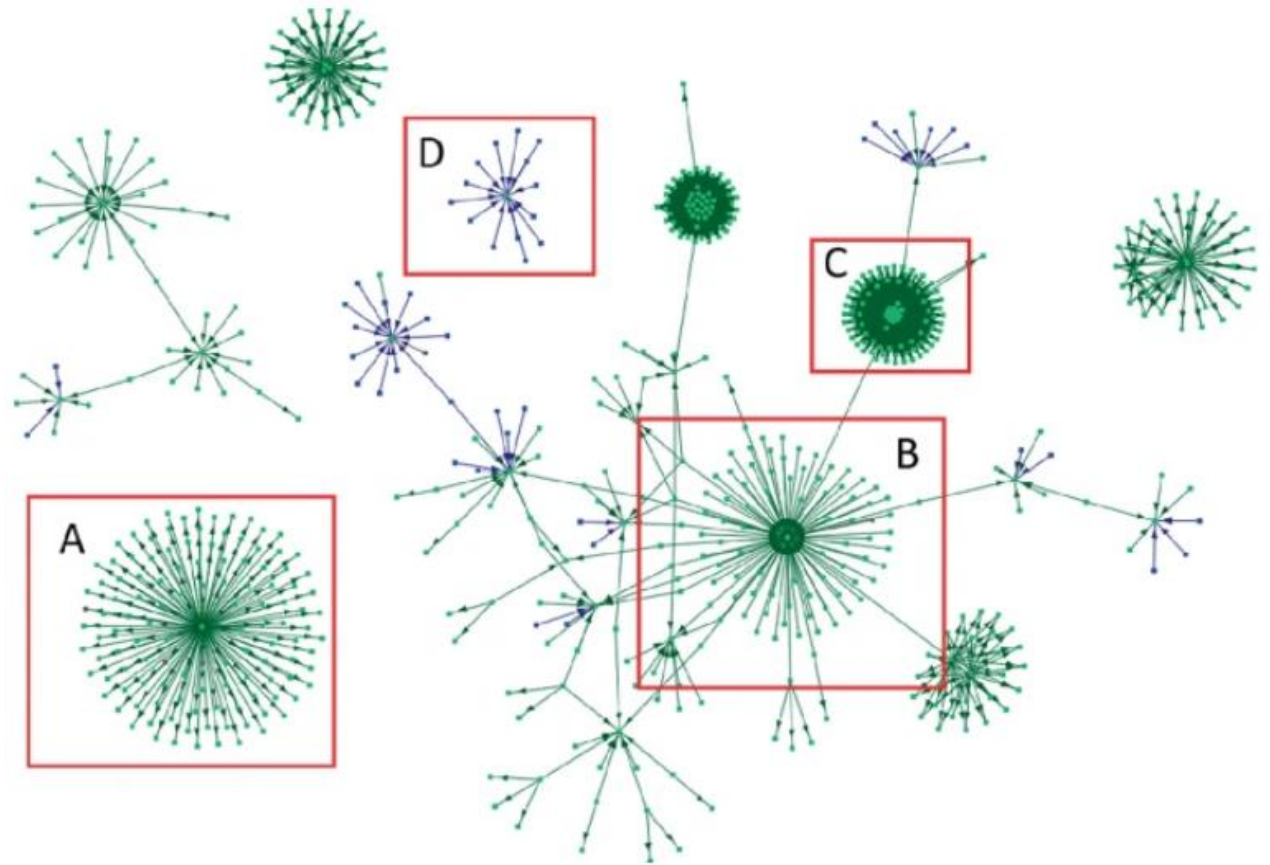
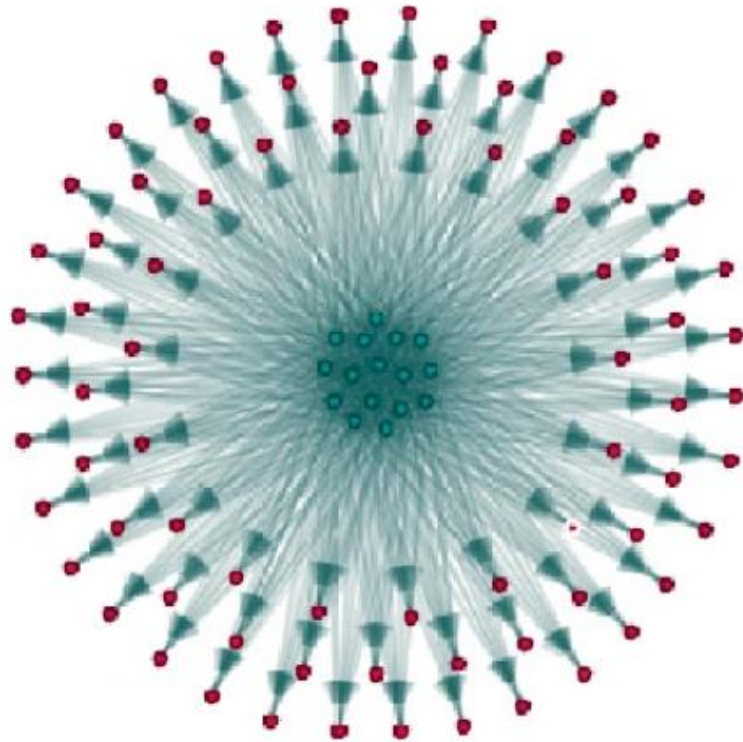
Table 8. Comparison of proportion of users suspended and lifespan of still-alive users.

	Coronavirus dataset			US Elections dataset		
	Bot	Cyborg	Human	Bot	Cyborg	Human
Proportion suspended (%)	89.2	56.0	19.5	76.9	49.2	19.9
Avg length of acct (days)	2751±1226	3663±1141	2901±1294	2643±1304	3437±1173	2715±1255
p-value of ANOVA for length of acct	$3.452 \times 10^{-5}$			$3.79 \times 10^{-142}$		
Graph of length of accounts against user type						

# Repeater, Amplifier Bots

- ❑ Repeater bots: Sharing same message multiple times
- ❑ Amplifier bots: Excessive retweets of specific accounts
  - ❑ Source bots: reply on other bots for retweet amplification
  - ❑ Overt amplifier bots: retweet isolated repeater bots
  - ❑ Periphery amplifier bots: round-robin scheme to promote the same accounts
  - ❑ Covert amplifier bots: conceal their bot-iness

# Repeater, Amplifier Bots



# Coordinated Bots

- ❑ Coordinated groups of user accounts work together in online social media
  - ❑ Can manipulate online discourse
- ❑ Ways to coordinate group online
  - ❑ Semantic: common hashtags, common phrases, common text
  - ❑ Social: common @mentions, common user tags
  - ❑ Referral: common URLs

# Coordinated Bots

- ❑ Combined Synchronized Index
  - ❑ Separates events, pairs of actors based on extent of synchronicity
  - ❑ Will be demonstrated later
- ❑ Identification of common online actions (e.g., hashtags, retweets, mentions) that are synchronous in time
- ❑ Bots are active participants in political/social activism discussions, synchronize to prop up messages
- ❑ Bot-Human pair have highest synchronization index
  - ❑ Bots can have an effect on human users, and possibly manipulate their opinions

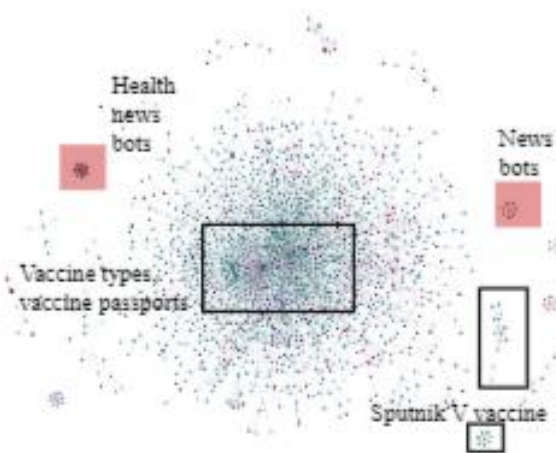


# Coordinated Bots

## Coordination Type

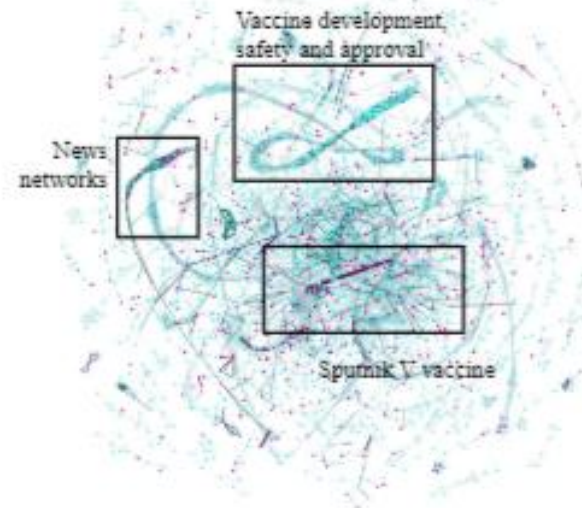
Semantic  
(via Hashtags)

COVID Vaccine  
Release 2021  
#covid #vaccine



D=0.001

Referral  
(via URLs)



D=0.004

Social  
(via Mentions)



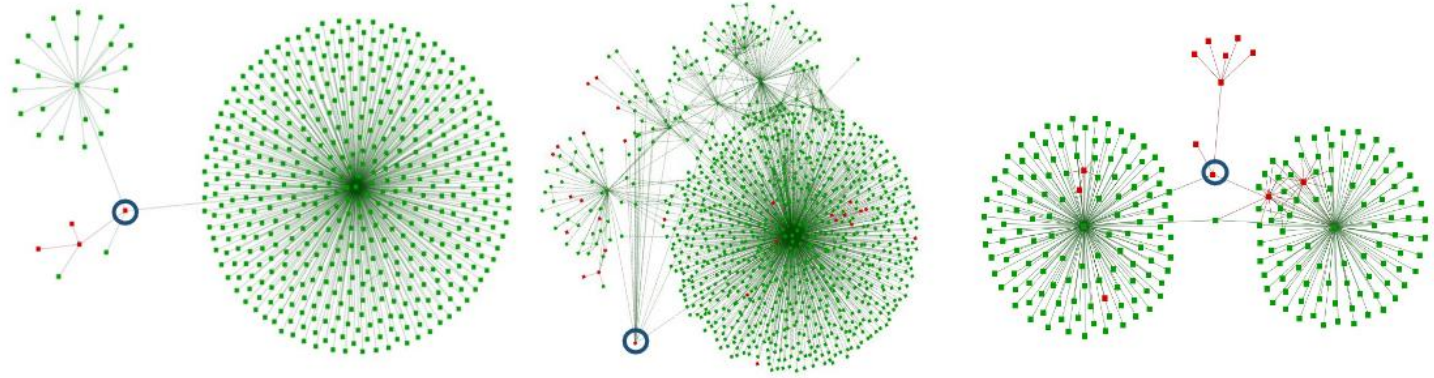
D=0.002

# Social Influence and Bots

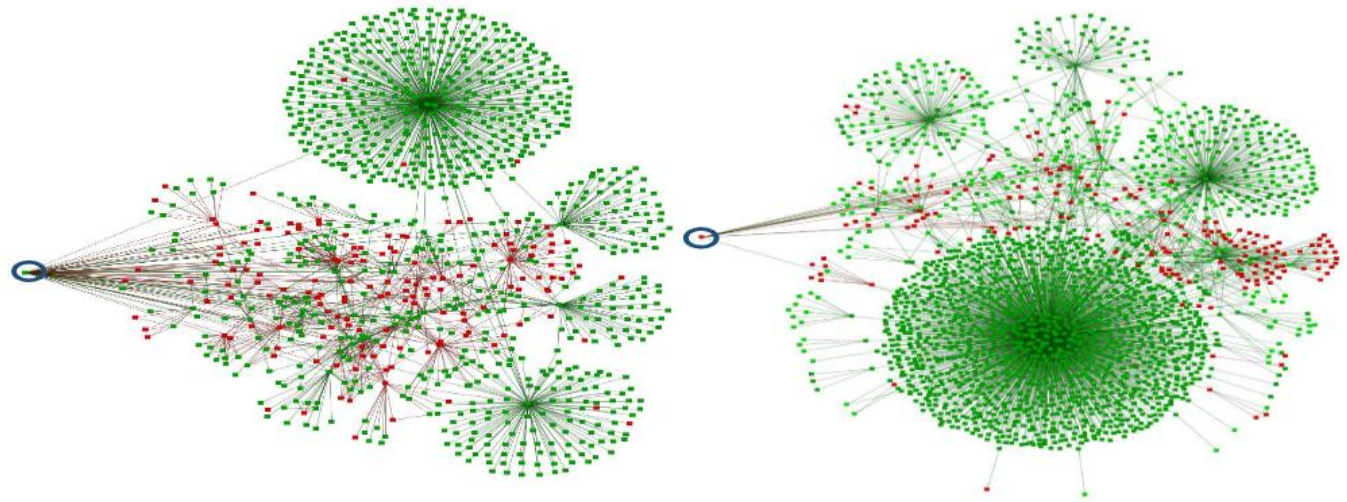
- ❑ Social influence: a user can be influenced by
  - ❑ Endogenous (innate static variables)
  - ❑ Exogeneous (influences from other neighboring users)
- ❑ Characterizes change of an individual's stances in a complex social environment towards a topic
- ❑ Study social influence through phenomenon of flipping stance towards covid vaccine
- ❑ Bots require less influence to flip
  - ❑ Fewer neighbors in opposite stances
  - ❑ Less conviction, more desire to fit in

# Social Influence and Bots

Bots



Humans

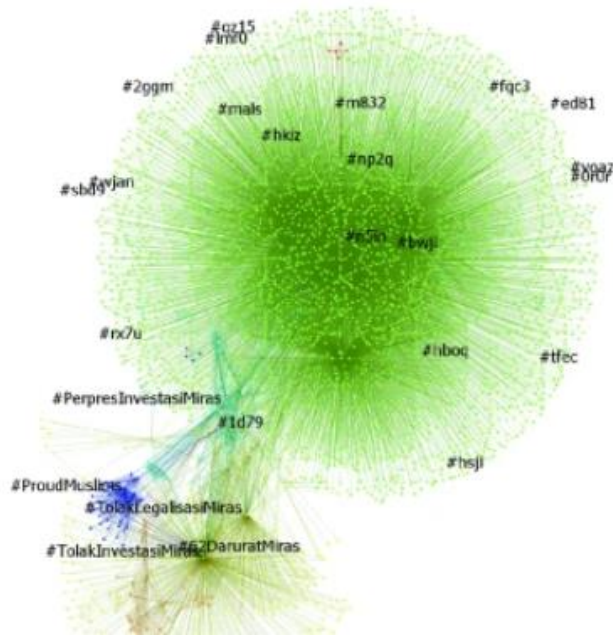


# Information Maneuvers

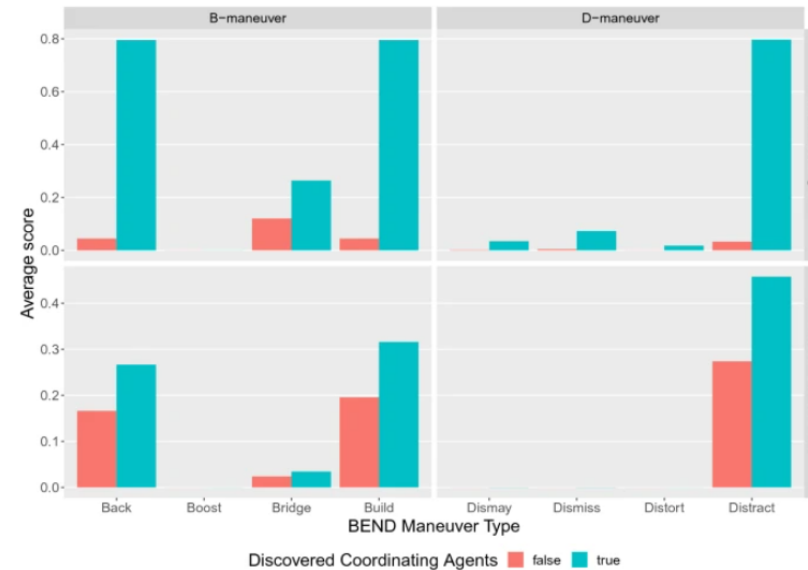
	Manipulating the narrative		Manipulating the social network	
Positive	<b>Engage</b>	Messages that bring up a related but relevant topic	<b>Back</b>	Actions that increase the importance of the opinion leader or create a new opinion leader
	<b>Explain</b>	Messages that provides details on or elaborate the topic	<b>Build</b>	Actions that create a group or the appearance of a group
	<b>Excite</b>	messages that elicit a positive emotion such as joy or excitement	<b>Bridge</b>	Actions that build a connection between two or more groups
	<b>Enhance</b>	Messages that encourage the topic-group to continue with the topic	<b>Boost</b>	Actions that grow the size of the group or make it appear that it has grown
Negative	<b>Dismiss</b>	Messages about why the topic is not important	<b>Neutralize</b>	Actions decrease the importance of the opinion leader
	<b>Distort</b>	Messages that alter the main message of the topic	<b>Nuke</b>	Actions that lead to a group being dismantled or breaking up, or appearing to be broken up
	<b>Dismay</b>	Messages that elicit a negative emotion such as sadness or anger	<b>Narrow</b>	Actions that lead to a group becoming sequestered from other groups or marginalized
	<b>Distract</b>	Discussion about a totally different topic and irrelevant	<b>Neglect</b>	Actions that reduce the size of the group or make it appear that the group has grown smaller



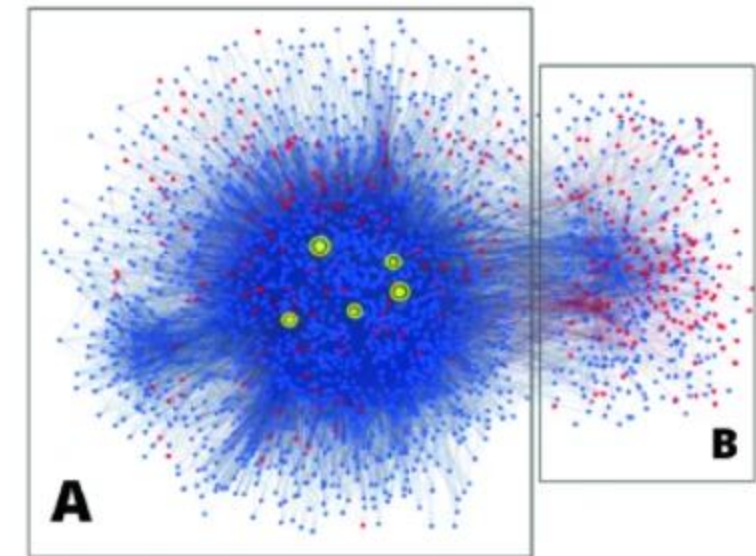
# Information Maneuvers



Bot discovery through  
hashtag hijacking



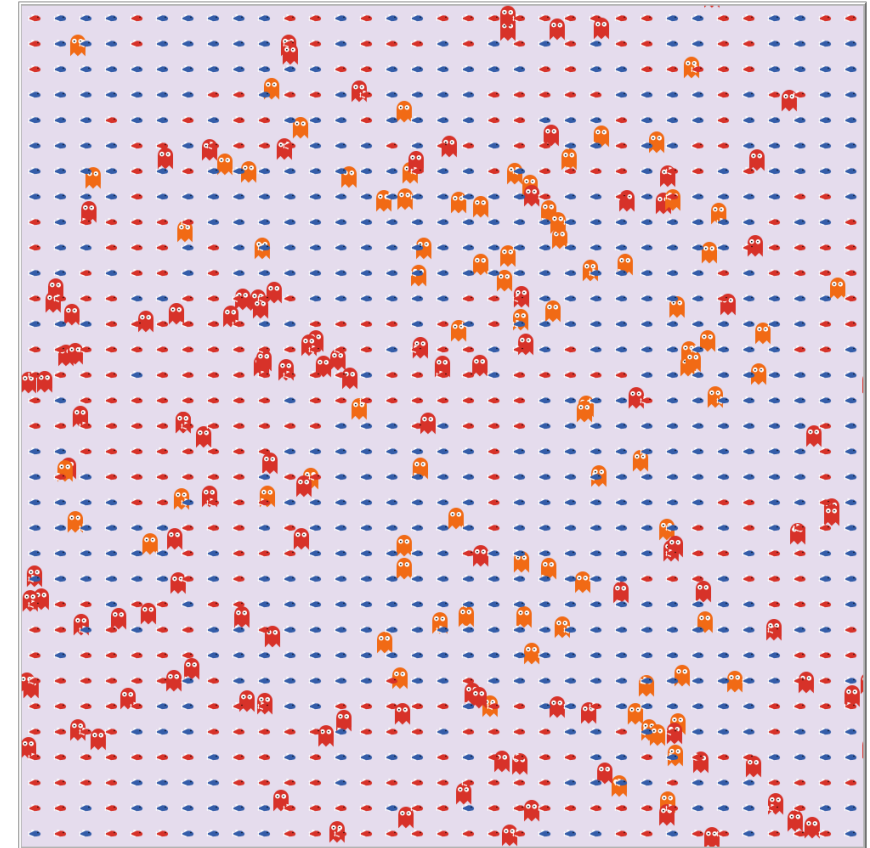
Bots use much more  
information maneuvers



Results in network  
polarisation

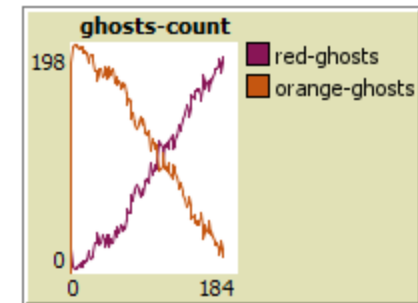
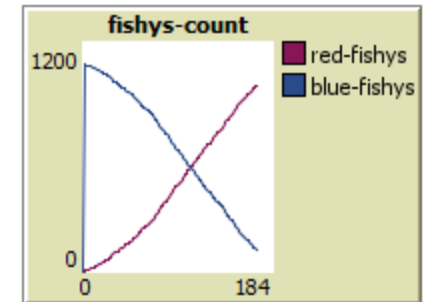
# Simulating bad bot behavior

- ❑ Agent-Based Simulation Model
  - ❑ Orange/ Red Ghosts: Humans/ Bots
  - ❑ Blue / Red Fishys: Information/ Disinformation
- ❑ Research Questions:
  - ❑ What proportions of bots make a conspiracy society?
  - ❑ [Inoculation] What probability of resistant to fishy color change makes a society resistant to conspiracies?
  - ❑ [Fact-Checkers] What proportion of fact checkers required to keep a society from being a conspiracy society?
  - ❑ [Flood with good info] What proportion of good humans required to keep society from being conspiracy society?



# Simulating bad bot behavior

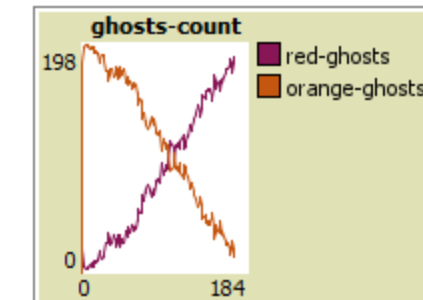
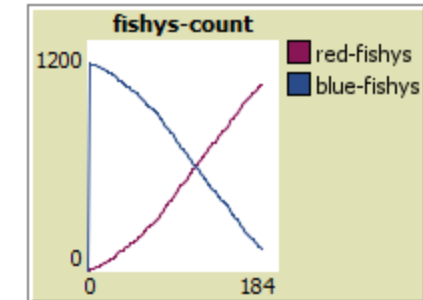
- ❑ What proportions of bots make a conspiracy society?
- ❑ Without any interventions, everyone eventually becomes converted





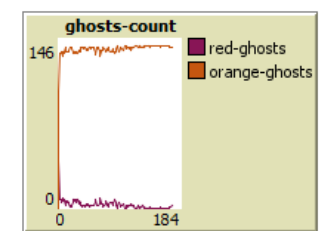
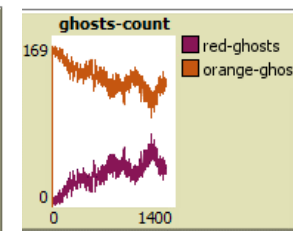
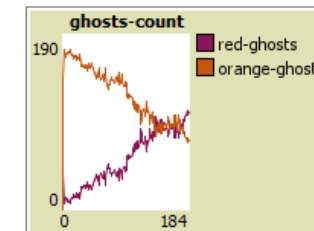
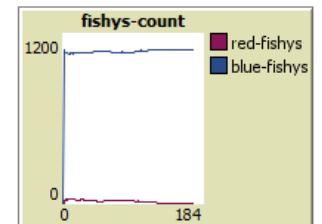
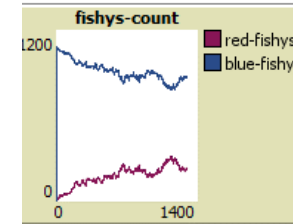
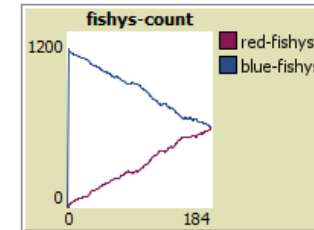
# Simulating bad bot behavior

- ❑ [Inoculation] What probability of resistant to fishy color change makes a society resistant to conspiracies?
- ❑ Regardless of how resistant you are, you will get converted if the bad guys are persistent enough



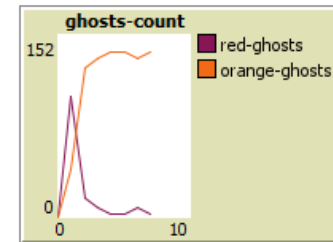
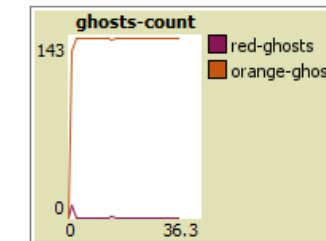
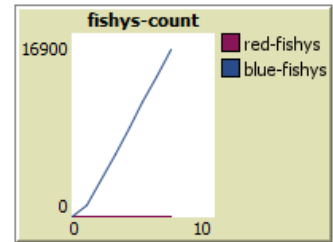
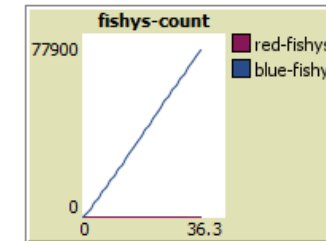
# Simulating bad bot behavior

- [Fact-Checkers] What proportion of fact checkers required to keep a society from being a conspiracy society?
  - Fact-Checkers will flip red fishys to blue fishys if they land on it
- If the ratio of fact-checkers:bad guys =  $2/3$ , a conspiracy society will not form



# Simulating bad bot behavior

- ❑ [Flood with good info] What proportion of good humans required to keep society from being conspiracy society?
  - ❑ Good people will add more good information to the space.
  - ❑ If the square is overwhelmed by good info, it is considered a good info square
- ❑ Flooding with good info can reverse a conspiracy society



# Applications of Social Media Bot Detection

- ❑ Bot Activity in Elections
- ❑ Bot Activity in Digital Diplomacy

# Bot Activity in Elections

- ❑ Are the bot activity in both elections the same?
  - ❑ Proportion of Bots
  - ❑ Narratives of Bots

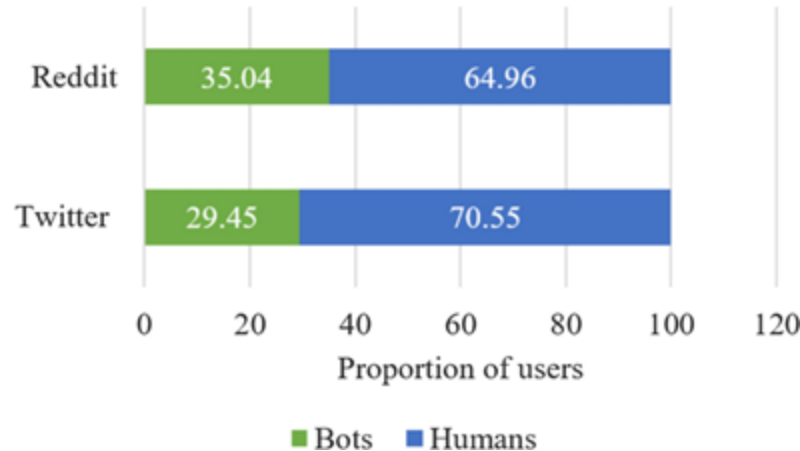
US Elections	Singapore Elections
3 Nov 2020	10 July 2020

**Ng, Lynnette Hui Xian**, and Kathleen M. Carley. "Assembling a multi-platform ensemble social bot detector with applications to US 2020 elections." *Social Network Analysis and Mining* 14, no. 1 (2024): 45.

Uyheng, Joshua, **Lynnette Hui Xian Ng**, and Kathleen M. Carley. "Active, aggressive, but to little avail: characterizing bot activity during the 2020 Singaporean elections." *Computational and Mathematical Organization Theory* 27, no. 3 (2021): 324-342.

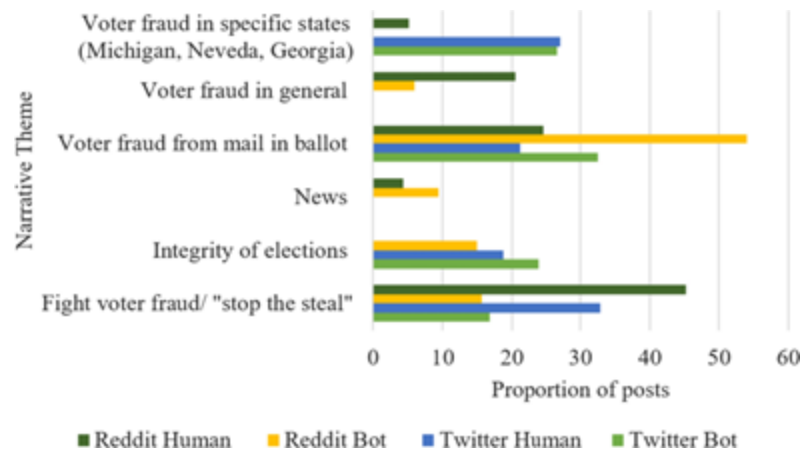
# Bot Activity in Elections – United States

## Proportion



~30% of bot users across  
Reddit & Twitter

## Narratives

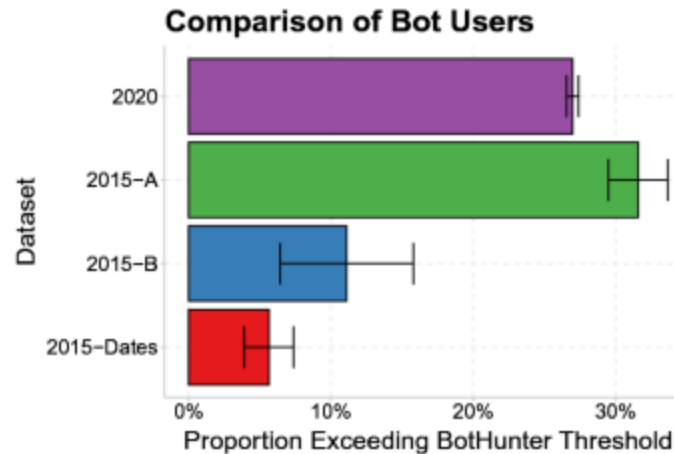


Echoed dis/misinformation  
narratives of voter fraud

Bots disseminate  
disinformation, humans  
advocate for action

# Bot Activity in Elections – Singapore

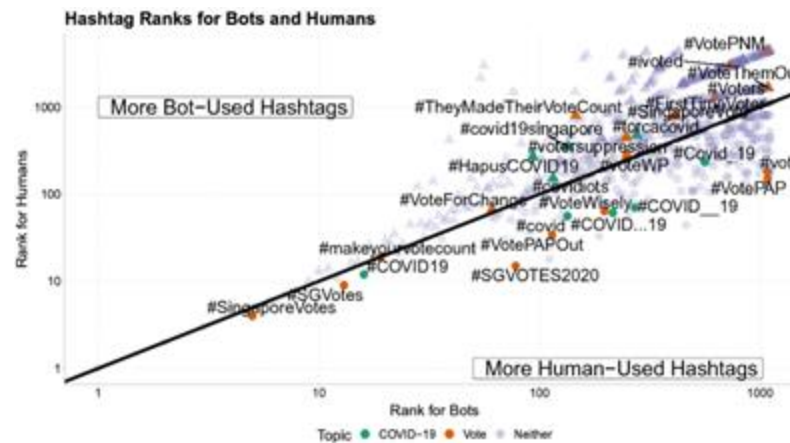
## Proportion



~30% of bot users across Reddit & Twitter

Bot activity about the same in 2015 and 2020

## Narratives



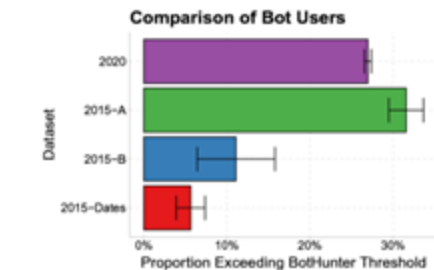
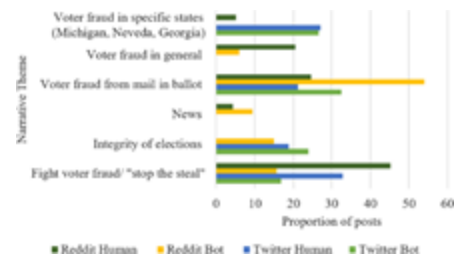
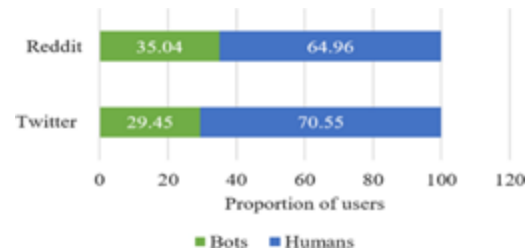
Common hashtags used between bots & humans but

Different specialized hashtags

- Bots: un-trustworthiness of elections
- Humans: mainstream topics

# Bot Activity in Elections

United States	Singapore
Active bot activity	Active bot activity
~30% users are bots	~30% users are bots
Bots disseminate disinformation on voter fraud	Bots share messages on the un-trustworthiness of elections
Humans advocate for action	Humans talk about mainstream topics





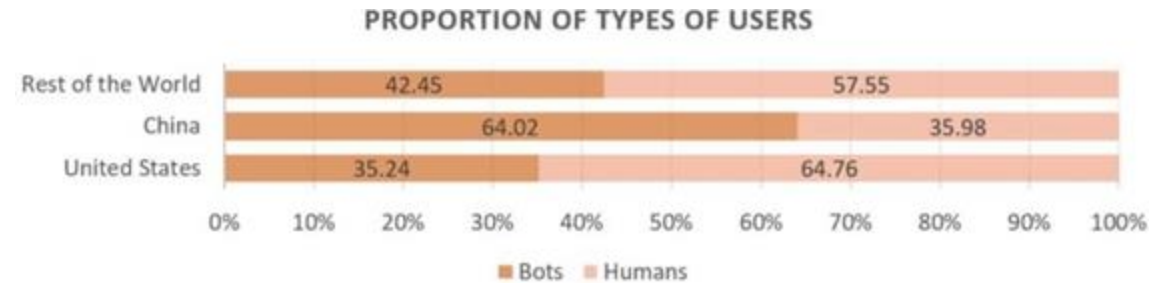
# Digital Diplomacy

- ❑ The story of the 2023 US-China balloon incident
- ❑ Is the Bot activity similar with the US-affiliated and China-affiliated bots?



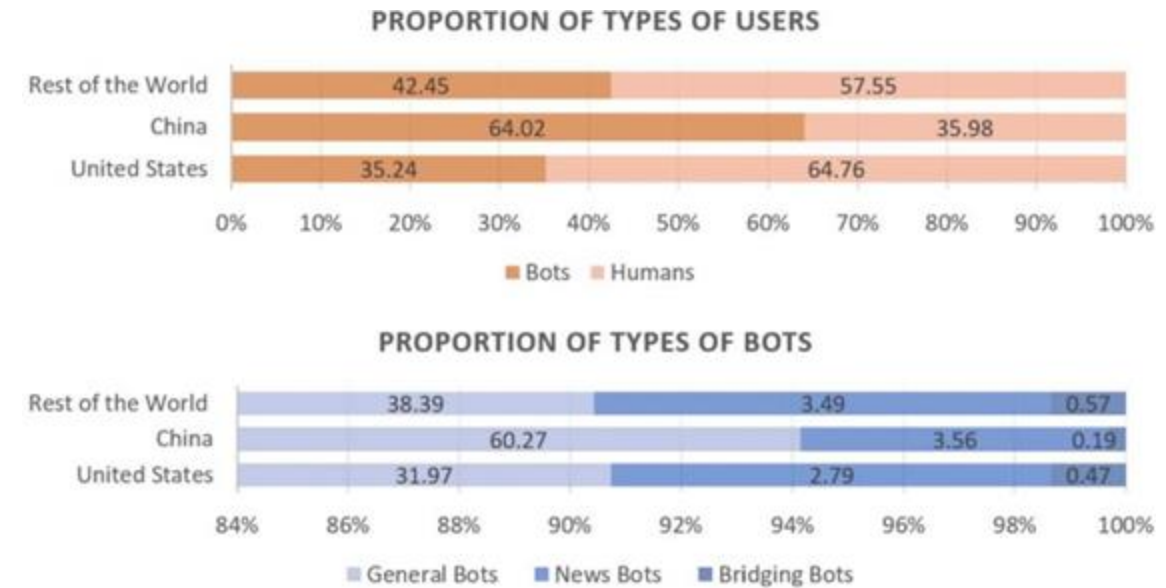
# Proportion of Bots

- ❑ Proportion of users affiliated with each region on Twitter
- ❑ Proportion of bots are higher compared to the other events
  - ❑ Bots are active, leverage on high-profile events



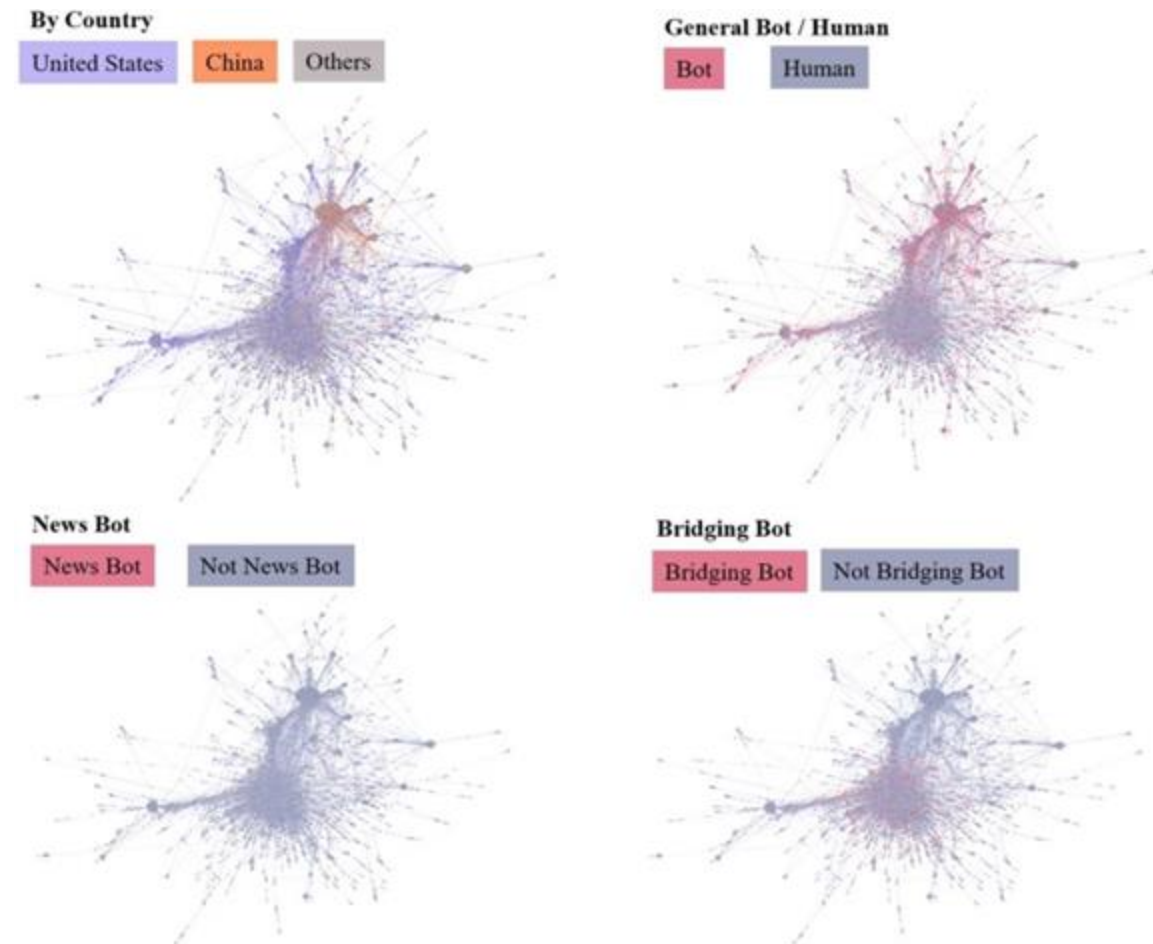
# Proportion of Bots

- Types of Bots – are present in all regions
  - General bots
  - News Bots
  - Bridging Bots



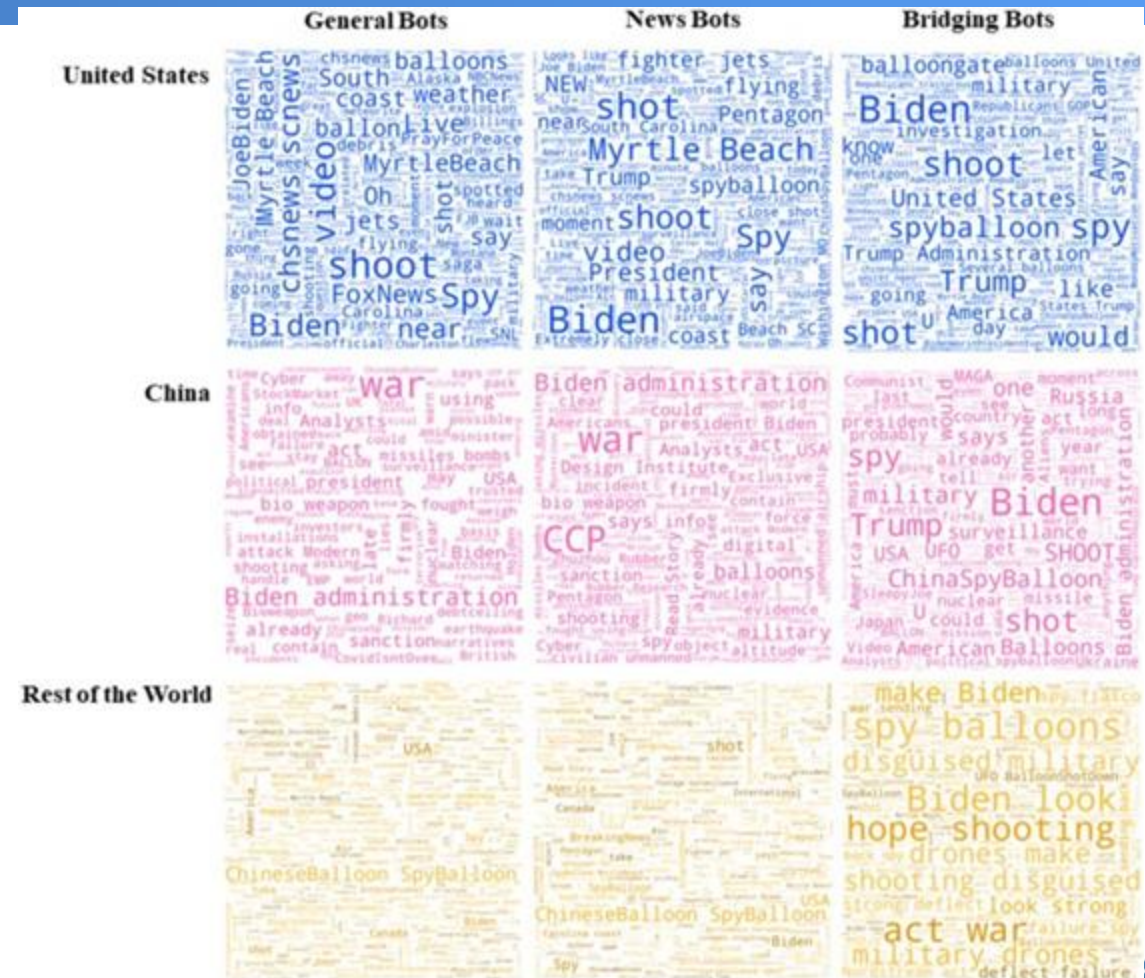
# Proportion of Bots

- ❑ Users affiliated with US & China in different clusters
- ❑ General bots cluster together
- ❑ News bots dispersed
- ❑ Bridging bots straddle US & China-affiliated clusters



# Proportion of Bots

- ❑ Themes are region-specific, not type specific
- ❑ US-affiliated: spatial location of balloon, possible surveillance properties
- ❑ China-affiliated: worry of escalation of war and sanctions
- ❑ Rest of world-affiliated: aliens and UFOs





←  r/science • 2 mo. ago  
Wagamaga


Armies of bots battled on Twitter over Chinese spy balloon incident. Around 35 per cent of users geotagged as located in the US exhibited bot-like behaviour, while 65 per cent were believed to be human. In China, the proportions were reversed: 64 per cent were bots and 36 per cent were humans.

Computer Science



newscientist.com

Open

 b2q • 2mo ago

So instead of world wars we have this invisible cultural war fought by bots that most people don't know about (tiktok and instagram) resulting in weird polarizations.

⬆ 243 ⬇ Reply ⬆ Share ...



Bakkster • 2mo ago

It's the new cold war. Though it doesn't necessarily stay on social media, the goal is always to push it mainstream.

⬆ 100 ⬇ Reply ⬆ Share ...



Annoverus • 2mo ago

90% of Instagram comments I see are by bots and I can't be proven wrong.

⬆ 1 ⬇ Reply ⬆ Share ...



kingbane2 • 2mo ago

someone should do a study on the comments section on most canadian news outlets. like ctv etc. i swear there must be a ton of bots on there. i looked through a few of the commenters accounts and they're all empty youtube channels or they follow the same channels, or there are reposted videos. tons of them very similar to each other.

⬆ 1 ⬇ Reply ⬆ Share ...

←  r/science • 2 mo. ago  
Wagamaga

Armies of bots battled on Twitter over Chinese spy balloon incident. Around 35 per cent of users geotagged as located in the US exhibited bot-like behaviour, while 65 per cent were believed to be human. In China, the proportions were reversed: 64 per cent were bots and 36 per cent were humans.

Computer Science





newscientist.com

Open



• 2mo ago

Per cent 🤔

↑ 5 ↓  Reply  Share ...



• 2mo ago

per cent

↑ 5 ↓  Reply  Share ...



• 2mo ago

Percent.

↑ 1 ↓  Reply  Share ...

• 2mo ago

I a bot

↑ 1 ↓  Reply  Share ...



# Hands on Section

Lynnette Hui Xian Ng, Jeffery Reminga

The CMU centers for:

Informed DEMocracy And Social cyber-security

Computational Analysis of Social and Organizational Systems



**Carnegie Mellon University**



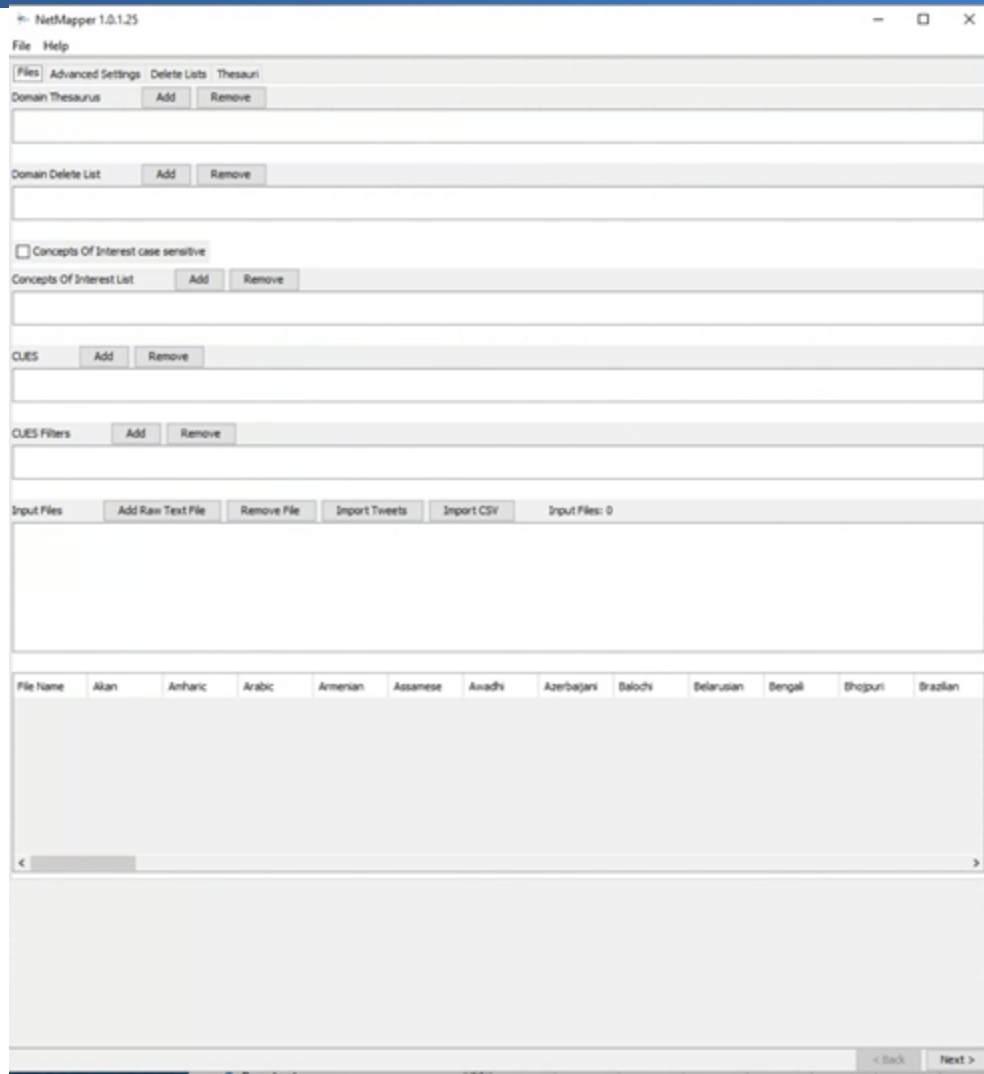
# Objectives

- ❑ Understand the differences between network positions, coordination activity and information maneuvers of bot and human accounts
- ❑ Use ORA to perform these investigations
- ❑ We are using Twitter data about COVID from Jan 2021

# NetMapper & BotBuster UI

- ❑ Running NetMapper to extract CUES attributes from tweet texts
  - ❑ <explain cues>
- ❑ Running BotBuster UI to identify if the user is likely to be a bot or not
  - ❑ BotBuster uses a mixture-of-experts machine learning model to identify the user based on data pillars like username, screenname etc
- ❑ Import information into ORA

# NetMapper & BotBuster UI



NetMapper 1.0.1.25

File Help

Files Advanced Settings Delete Lists Thesauri

Domain Thesaurus Add Remove

Domain Delete List Add Remove

☐ Concepts Of Interest case sensitive

Concepts Of Interest List Add Remove

CUES Add Remove

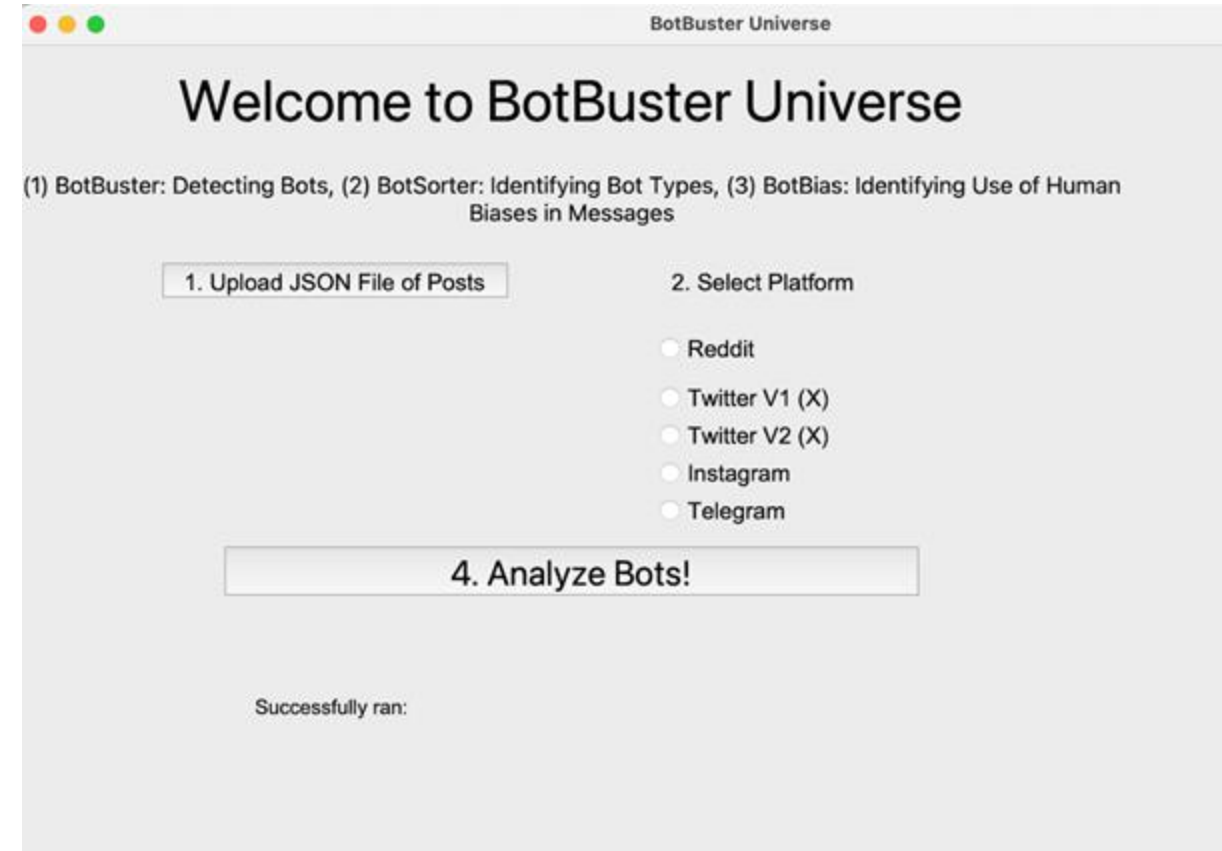
CUES Filters Add Remove

Input Files Add Raw Text File Remove File Import Tweets Import CSV Input Files: 0

File Name	Akan	Anharic	Arabic	Armenian	Assamese	Awadhi	Azerbaijani	Balochi	Belarusian	Bengali	Bhojpuri	Brazilian

< >

< Back Next >



BotBuster Universe

## Welcome to BotBuster Universe

(1) BotBuster: Detecting Bots, (2) BotSorter: Identifying Bot Types, (3) BotBias: Identifying Use of Human Biases in Messages

1. Upload JSON File of Posts

2. Select Platform

4. Analyze Bots!

- ☐ Reddit
- ☐ Twitter V1 (X)
- ☐ Twitter V2 (X)
- ☐ Instagram
- ☐ Telegram

Successfully ran:

# Step 1: Import DyXML data

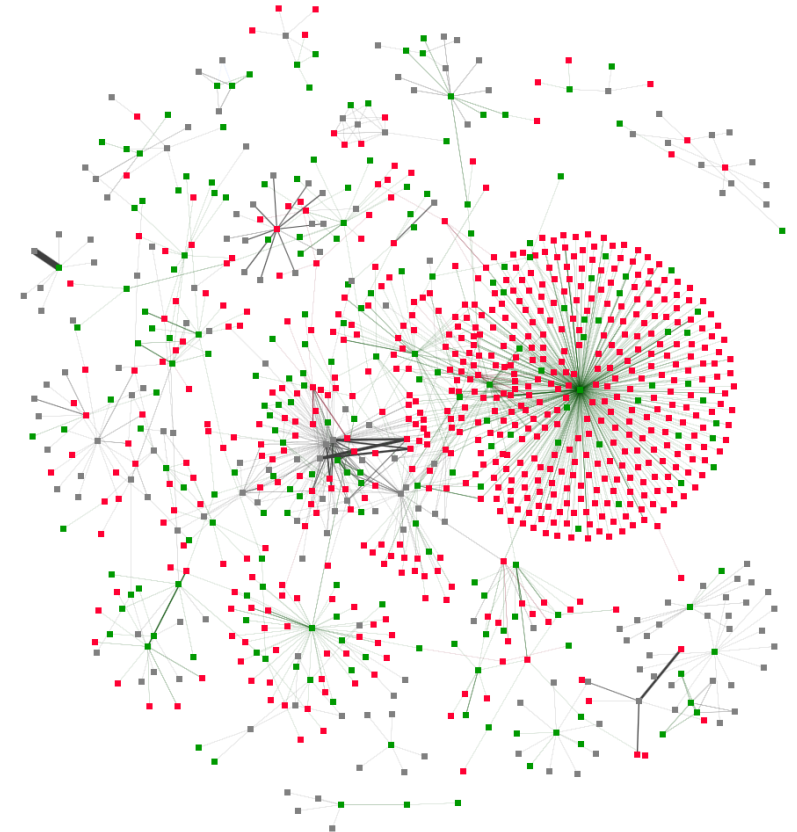
- ☐ Take DyXML data and drag into ORA
- ☐ How many Agents and Tweets do you get?

# Step 2: Transform Bot Probability Score

- ☐ Find the probability attribute
- ☐ Transform the attribute
  - ☐ Transform attribute > Recode
  - ☐ Attribute: probability
  - ☐ >0.7 => True
  - ☐ <0.7 => False
  - ☐ Create new attribute > is\_bot
- ☐ How many bots are there?
- ☐ How many not-bots are there?

# Step 3: Visualize the All-Communication Network

- ☐ Click on Visualize > Agent x Agent – All Communication
- ☐ Color nodes by bot or not
- ☐ How many clusters do you see?
- ☐ What patterns do you see?
- ☐ Do you see similar clusters / patterns in other networks?





# Step 4: Locate Groups Report

- ☐ Run the Locate Groups report
- ☐ How many groups do you get?
- ☐ Are there any characteristics to the groups?

# Step 5: Run Coordination Analysis Report

- ☐ Which users coordinate the most?
- ☐ What dimensions are they coordinating the most?

# Step 6: Run BEND Report

- ☐ Run BEND report > Split by group is\_bot
- ☐ Which BEND maneuver is most used?
- ☐ Do bots or humans use more BEND maneuvers?
- ☐ Are there any patterns in the usage of BEND Maneuvers?

# For More Information

- ❑ Lynnette Ng & [lynnetteng@cmu.edu](mailto:lynnetteng@cmu.edu)
- ❑ Director Kathleen M. Carley - [kathleen.carley@cs.cmu.edu](mailto:kathleen.carley@cs.cmu.edu)
- ❑ IDeaS website - <https://www.cmu.edu/ideas-social-cybersecurity/>
- ❑ CASOS website – <http://www.casos.cs.cmu.edu/>
- ❑ Social Cybersecurity Working Group - <http://social-cybersecurity.org>
- ❑ Facebook: [@IDEasCMU](https://www.facebook.com/IDEasCMU)
- ❑ Twitter: [@IDEaSCMU](https://twitter.com/IDEaSCMU)
- ❑ YouTube: [IDEaS Center](https://www.youtube.com/IDEaS%20Center)
- ❑ Email-Distro Lists