

# Identifying Bot Networks through Synchronized Network Structures

Lynnette Ng

The CMU centers for:

Informed DEMocracy And Social cyber-security

Computational Analysis of Social and Organizational Systems



**Carnegie Mellon University**



# Social media bots

A **social media robot**, or “**bot**” is a

- ❑ programmable social media account that
- ❑ automatically carries out a series of social media mechanics on social media platforms
  - ❑ le. Automatic retweeting of specific tweets;
  - ❑ Automatic posting of generated and pre-canned texts

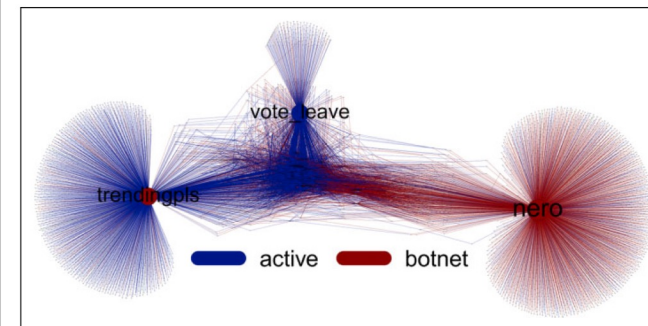
# Why do we care about bots?

- ❑ There are good bots, benign bots and bad bots
  - ❑ Good bots: aid human users and their communities, e.g. help social media managers automatically post at specified time; chat bots for emotional support
  - ❑ Benign bots: put forth information without interacting with people, e.g. posting the time every hour
  - ❑ Bad bots: disrupt the social space
- ❑ Researchers mostly care about bad bots because they can:
  - ❑ Spread dis/misinformation
  - ❑ Amplify harmful narratives
  - ❑ Create calls for action for violence, e.g. protests, riots

# Bot Networks

- ❑ In 2017, 13k bot-like Twitter accounts were found to be part of a bot network
  - ❑ Expressed ideological positions about the United Kingdom Brexit debate
  - ❑ Actively supporting the Leave campaign during the Brexit referendum, disappearing shortly after the debate
- ❑ Discovered by inspecting URLs within Twitter posts
  - ❑ Identify users that post URLs of domain names that source information to bots
- ❑ Botnets had seed users that posted content
  - ❑ Other bots performed content and influence amplification, to aggregate and share content posted by seed users
  - ❑ Accounts in bot net exhibited clear specialization: sharing posts, echo campaign slogans, trigger discussion cascades

<sup>1</sup> Bastos, M.T., Mercea, D.: The brexit botnet and user-generated hyperpartisan news. Social science computer review 37(1), 38–54 (2019)



# Bot Networks

Bot Networks, or BotNets, are

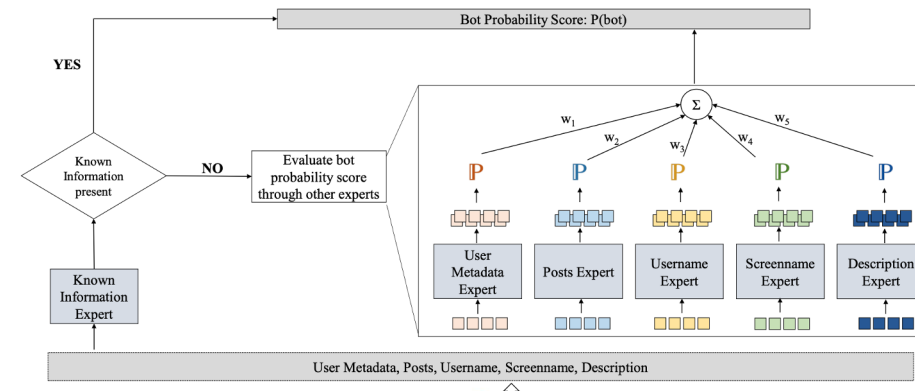
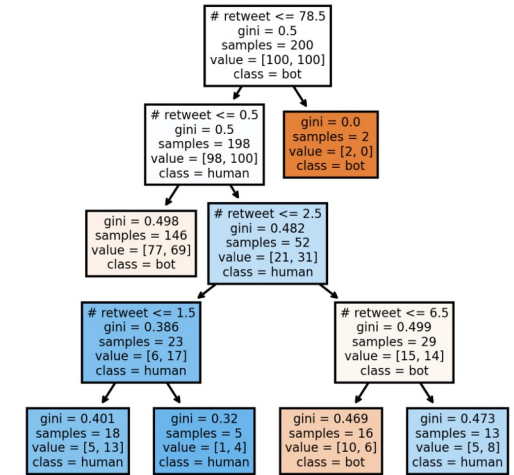
- ☐ clusters of bots that
- ☐ work together to achieve a common goal
  - ☐ Disseminate a single narrative
  - ☐ Prop up the network influence of a certain account
- ☐ Some social media actions they use: follow each other, tag each other in a post, share each other's post

# Bot Detection Algorithms

- ❑ Identifies whether a social media user is a bot or not
- ❑ Typically provides a probability from 0 to 1 that the social media user is a bot
- ❑ Uses features from the account like post texts, post metadata, user metadata (e.g. number of followers, number of likes), network graph features etc

# Bot Detection Algorithms

- ❑ Random Forest algorithms
- ❑ Support Vector Machine algorithms
- ❑ Logistic Regression algorithms
- ❑ Deep-learning based algorithms: convolutional neural networks, recurrent neural networks, long short-term memory networks
- ❑ Temporal methods
- ❑ Graph-based methods



# Detecting Bot Networks with Bot Detection Algorithms

- ❑ Can be difficult because not all users in a bot network are labelled as a bot
- ❑ Hence, we need other ways to do detection
- ❑ Follow the idea of:
  - ❑ No bot is an island
  - ❑ Find one, find all



# Synchronization in social media

- ❑ Phenomenon where many users simultaneously perform the same action
- ❑ Synchronized groups on social media can pose a threat to the social fabric
  - ❑ Organization of campaigns and protests
  - ❑ Analysis of 16 countries revolving around the 2011 Arab Spring protests show correlation between online synchronization and offline protests <sup>1</sup>
  - ❑ Analysis of similar texts in the 2011 United States Capitol Riots reveal groups of user clusters supporting disinformation narratives and themes, and we did observe an actual riot echoing some themes present on social media <sup>2</sup>

<sup>1</sup> Steinert-Threlkeld, Z.C., Mocanu, D., Vespignani, A., Fowler, J.: Online social networks and offline protest. *EPJ Data Science* **4**(1), 1–9 (2015)

<sup>2</sup> Ng, L. H. X., Cruickshank, I. J., & Carley, K. M. (2022). Coordinating Narratives Framework for cross-platform analysis in the 2021 US Capitol riots. *Computational and Mathematical Organization Theory*, 1-17.

# Measuring synchronization

## ❑ Temporal synchronization <sup>1</sup>

- ❑ Identifies users that employ a defined social media mechanic (e.g. post same hashtag/@mention) within a short time window. Users that employ a high frequency of this combination are termed to be synchronized

## ❑ Narrative synchronization <sup>2</sup>

- ❑ Identifies users that post similar texts within the conversation as synchronized users
- ❑ Texts are represented in a vector form (e.g. using BERT) and compared via a distance metric (e.g. Euclidean)

## ❑ Image synchronization <sup>3</sup>

- ❑ Identifies users that post similar images within the conversation as synchronized users
- ❑ Images are represented in a vector form (e.g. using ResNet50) and compared a distance metric (e.g. Euclidean)

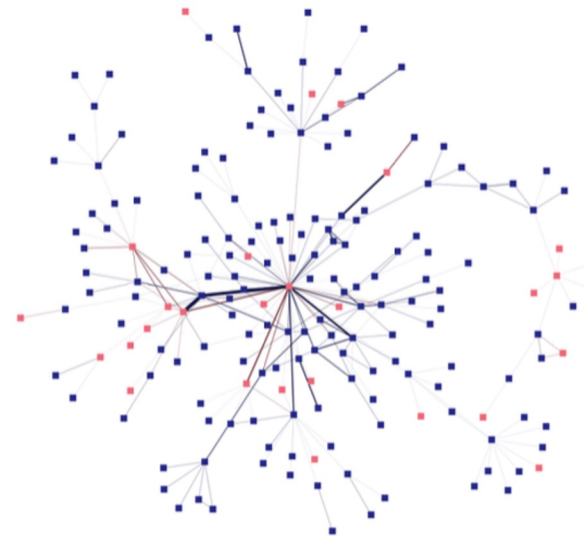
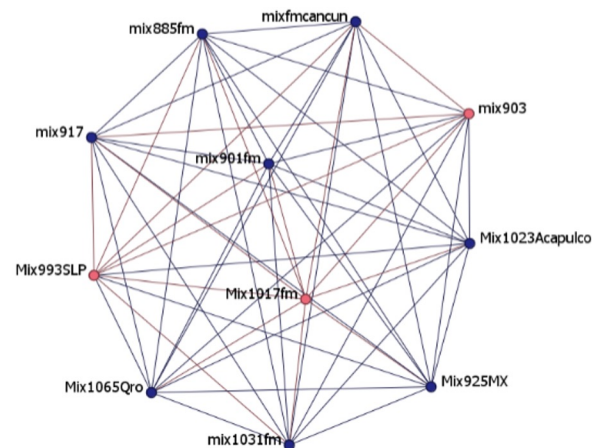
<sup>1</sup> Magelinski, T., Ng, L., & Carley, K. (2022). A synchronized action framework for detection of coordination on social media. *Journal of Online Trust and Safety*, 1(2).

<sup>2</sup> Ng, L. H. X., Cruickshank, I. J., & Carley, K. M. (2022). Coordinating Narratives Framework for cross-platform analysis in the 2021 US Capitol riots. *Computational and Mathematical Organization Theory*, 1-17.

<sup>3</sup> Ng, L. H. X., Moffitt, J. D., & Carley, K. M. (2022). Coordinated through aWeb of Images: Analysis of Image-based Influence Operations from China, Iran, Russia, and Venezuela. *arXiv preprint arXiv:2206.03576*.

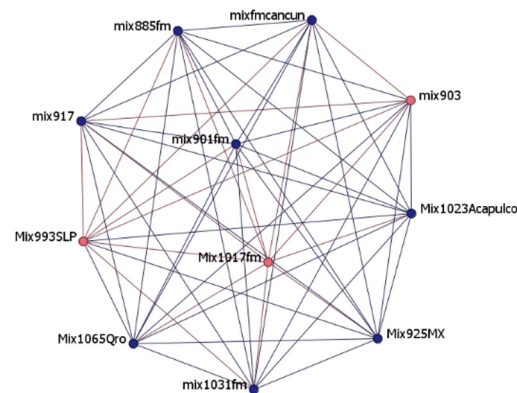
# Bot Network through temporal synchronization

- ❑ Black Panther movie release dataset
- ❑ Users (nodes) tweeting the same hashtag within 5 min time window have an edge in the network graph
- ❑ Filtered for links of high frequency: edges that are above the (mean + std dev) are kept

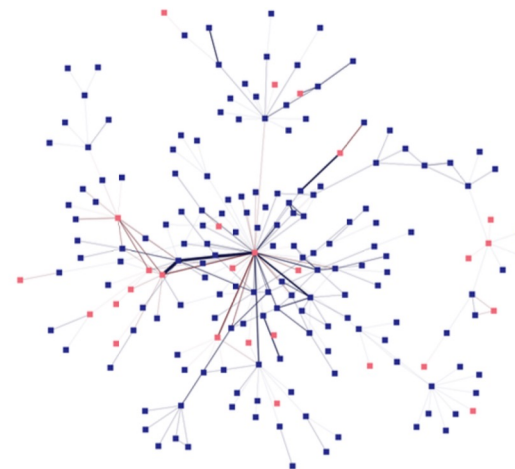


# Bot Network through temporal synchronization

- ❑ Two bot networks extracted
- ❑ Accounts identified as bots through a bot detection algorithm are marked as red; the rest as blue
- ❑ Not all accounts are identified as bots: thus, the need to use synchronization methods to identify botnets rather than individually identify bots



(a) Bot Network 1. All accounts were suspended a year later.



(b) Bot Network 2. 70% of the accounts were suspended a year later and some still-remaining accounts are self-declared bot accounts.

# Bot network through narrative synchronization

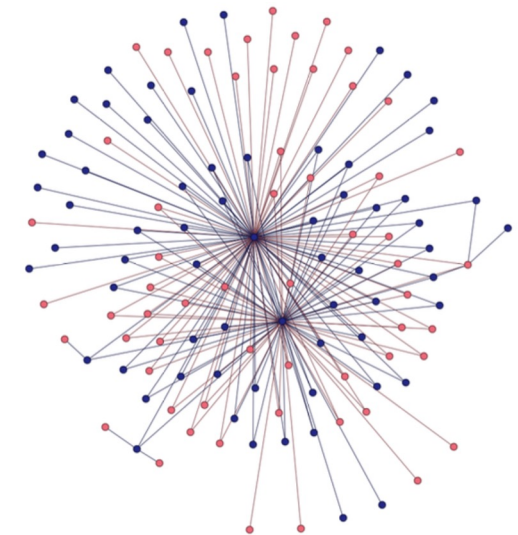
- ❑ Black Panther movie release dataset
- ❑ Every tweet is represented via a BERT vector
- ❑ All-pairs tweet comparison is performed
  - ❑ Posts are deemed similar if they are at least 70% similar in terms of the BERT vector representation of the words in the posts.
- ❑ Users (nodes) have edges representing that number of similar tweets
- ❑ Filtered for links of high frequency: edges that are above the (mean + std dev) are kept

# Bot Network through narrative synchronization

- ❑ Two bot networks extracted; not all users are identified as bots (red nodes)
- ❑ Bot Network 1: Spanish news/announcement botnet that retweets
  - ❑ Prefixes usernames with “milliena”
  - ❑ Each of them focus on a different speciality:
    - @MilenioTech focuses on tech news and
    - @mileniotv focuses on updates on television shows
- ❑ Bot Network 2: Two key coordinating nodes with peripheral nodes
  - ❑ Bot like behavior of consistent retweeting



(a) Bot Network 1. These are announcement bots with the prefix “millenia”. All accounts were suspended a year later.

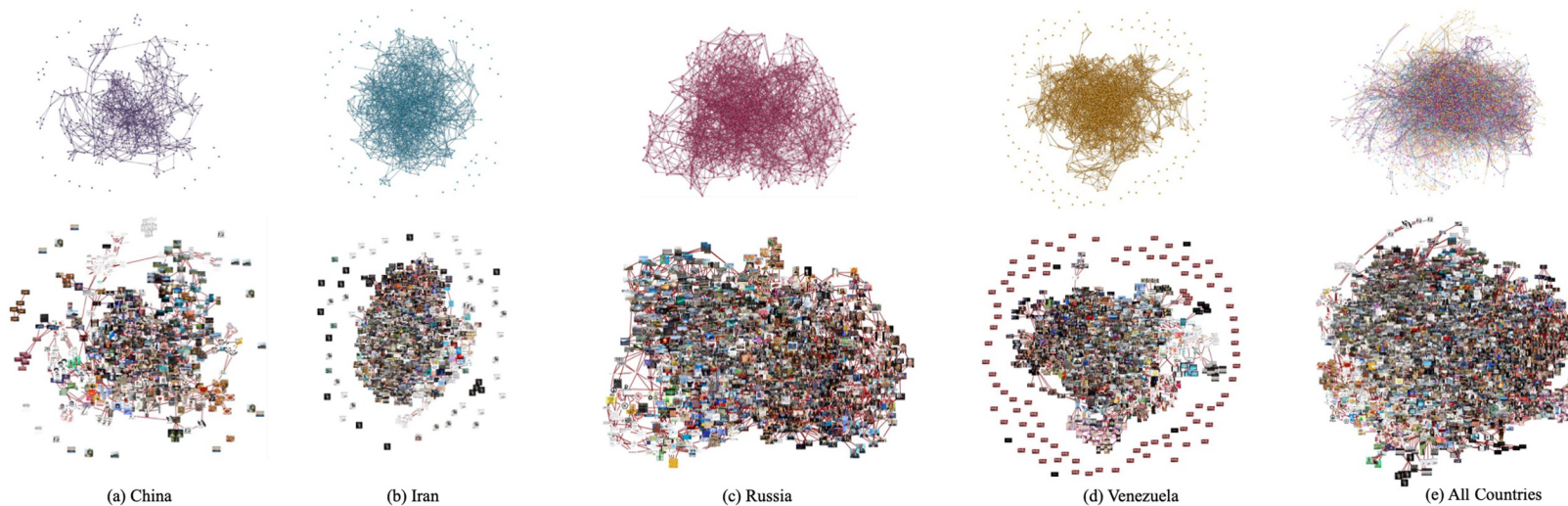


(b) Bot Network 2. 57% of the accounts were suspended a year later.



# Bot Network through image synchronization

- ❑ Inauthentic accounts released by Twitter that are operated by different countries
- ❑ Images are represented by ResNet-50 vectors
- ❑ All-pairs image comparison performed
  - ❑ Images are linked together if they are similar (i.e. more than 70% of their vectors are similar)

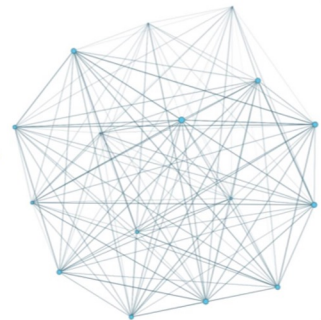


# Bot Network through image synchronization

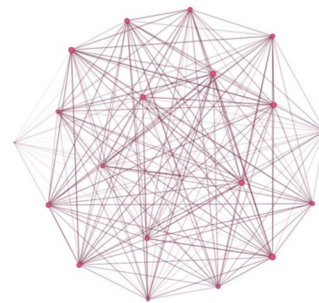
- Accounts are linked together if they have similar images
- Networks are densely connected, consistent with the data (classified under coordinated inauthentic behavior by Twitter)
- Varying levels of density and number of agents taking part in image synchronization
  - Provides visual cues towards identifying the main coordinators of influence campaigns



(a) China



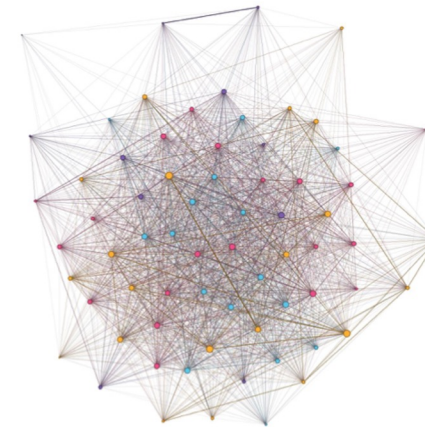
(b) Iran



(c) Russia



(d) Venezuela



(e) All Countries



# Limitations

- ❑ Temporal synchronization dependent on time windows
  - ❑ Too wide a time window, too many users synchronize
  - ❑ Too narrow a time window, too little users synchronize
- ❑ Narrative and image synchronization dependent on the robustness of vector representation used
  - ❑ Image vector representation must be robust enough to handle skews, stretches and changes in perspective
  - ❑ Narrative vector representation must be robust enough to handle misspellings, different spellings, different sentence structures

# Future Work

- ❑ Investigate the users and user types involved of bot networks identified through synchronization methods
- ❑ Investigate the actions of the users involved in the bot networks, e.g. amplification of narratives etc

# Thank you!

- Lynnette Ng
- [lynnetteng@cmu.edu](mailto:lynnetteng@cmu.edu)
- Twitter: @quarbby