Thesis Proposal
The Bot-Human Nexus in Social Media

Lynnette Hui Xian Ng

28 September, 2023

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Kathleen M. Carley, Chair, Carnegie Mellon University
L. Richard Carley, Carnegie Mellon University
Nicolas Christin, Carnegie Mellon University
Melissa Chua, Defense Science and Technology Agency

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

Abstract

Social media platforms are digital environments in which two key species inhabit: bots and humans. Bots are automated accounts that have been observed to affect businesses, country-wide elections, healthcare discourse and even the entertainment sphere. A long string of efforts have been dedicated to detection social media bots. However, common approaches are constrained by their binary classification of bot/human users. This binary classification scheme collapses the diverse variations of bots into a single class, and collectively group bots as malicious digital actors. These assessments focus on a one-dimensional view of bots and thus lose the nuance of the variations within the bot species, and the dynamic intra- and inter-species interactions.

This thesis approaches the bot-human nexus as two unique and dynamic organisms interacting within the social media space. The heart of this thesis examines two questions: How do bots and humans co-exist in the social media landscape? Are there different types of bots in social media and do they have unique characteristics? Within this thesis, I leverage on computational social science and network science methods to characterize the bot species and its interaction with humans. At the ecosystem level, I use linguistic and network methods to differentiate between a bot and a human, and characterize the similarities and differences between the two species in terms of the type of language used, the expression of social identities and emotions, and their network communication structures. At the habitat level, I analyze bot detection models and showcase the diverse bot types through a typology of bots. This expands the detection of social media bots to provide details into bot detection algorithms and the types and mechanics of the bot species. At the community level, I build on current research on synchronization on social media to analyze intra-species interactions through three dimensions: temporal, narrative and image coordination, which results in groups of bots deliberately spreading a specific message. At the ecosystem interaction level, I analyze how bots capture the hearts of humans through cognitive biases. Through empirical observations, I profile the tactics, techniques and procedures. Finally, I observe ecosystem changes by connecting both bot and human activity through a social influence model. This model simulates the changes in the ecosystem and towards each species in terms of their expressed opinion towards a topic, investigating whether the ecosystem can eventually find a balance. Collectively, these contributions enhance our social scientific understanding of the nature, interactions and impact of social media bots, and underscore the importance of theoretically informed computational methods to observe and engage this unique species.

Acknowledgments

For my family who has always been rooting for me.
For my advisor and my friends who have believed in me.
And for my heart transplant donor and the family who gave me the breath of life to tell this story.

Contents

Chapter 1

Introduction

1.1 Overarching Thesis Goals

Social media platforms are internet-based applications used for social networking, which includes creating user-generated content and making connections with other people. Social media bots, or bots for short, are automated accounts controlled by software algorithms instead of human users [56]. These creatures have gained the attention of social cybersecurity researchers because they have been observed to amplify disinformation [20], manipulate opinions [86] and disseminate propaganda [53]. Some of these online conversations eventually spillover into the offline world, resulting in threats to public safety like protests [62, 75].

Past researchers built numerous computational methods to differentiate between bots and humans in the social media space. However, these approaches typically consider bot detection in isolation from its social context through binary labels of bots and humans. Despite there not being a good/bad bot label resulting from these bot algorithms, a large proportion of studies focus on bots for malicious uses [4, 9, 30, 67]. Similarly, a large proportion of people believe that bots are used for malicious purposes [5]. But there are good bots too. Bots are not homogeneous and are used for a variety of purposes [4], from content moderation to chat bots for social good to organizing volunteers in times of crisis [40, 83, 93].

Fundamentally, this thesis examines the following questions: Can bots and humans coexist in the social media habitat? Are there different types of bots in social media and do they have unique characteristics? With these research questions, this thesis argues that bots and humans are unique and dynamic species within the social media ecosystem: the bot species need to be understood in terms of its differences from the human species, its variations, its communities and finally its relationship with humans. I propose methods to examine the bot-human nexus via an ecosystem perspective on social media in terms of social and computational aspects, and make sense of the massive bot-human communication patterns and interpretations of automated bot messages and interactions. Throughout this work, I contribute to interdisciplinary theory around social media bots, a typology for the different archetypes of bots, new tools for identification and characterization of bots and the bot-human interaction, and a wide range of empirical insights across global and regional events and several social media platforms.

1.2 Literature Review

Social media bots have been of great interest of the computational social science world because of their increasing population in our online ecosystem. Studies estimate that more than half of the traffic on the Internet is generated by bots [57], a quarter of the tweets on Twitter is created by bots [25], and two-thirds of the links posted on Twitter are generated by bots [104]. The use of software automation, whether through official Application Programming Interfaces (APIs) provided by the social media platforms, or by unofficial means such as web scraping methods, means that these accounts can disseminate information very quickly and to a wide range of users [78]. Automation provides a low-cost and low-effort solution for large social media reach compared to using human resources [4], coupled with the poor ability of humans to differentiate bots from human users [33], bots can be used as an amplification device in disseiminating information campaigns through a social network [105]. These creatures are of note to researchers because they have significantly impacted public opinions, both for better and for worse. Bots can stoke anger and fuel protests [30, 67], as well as spread positive messages [63] and chatbots for medical triage and mental health support [40].

Bots are created for many different goals, ranging from marketing to politics, and their behaviors vary based on those distinct goals [4]. In literature, bots are mostly observed for their malicious uses. For example, bots have been observed in the 2011 Egyptian uprising to be distributing manipulative and potentially disruptive political discussions, such as supporting violence and human rights abuses [30]. Another example includes Russian troll bots have been observed to generate over 10 million tweets as part of their influence campaign during the 2016 US presidential elections, participating in activities such as stoking political fears, influencing political debates and circulating memes [4, 9]. These online conversations are concerning because human users can be influenced by the messages and ideologies, and eventually the disruption stoked online can spillover to the offline world and result in protests, riots or manipulated voting.

However, there are good bots too. Chatbots for social good provide low-cost conversational interfaces to support medical triage, mental health support and education [40]. Amplifier bots can be used to support positive behaviors such as encouraging vaccinations [63]. During the coronavirus pandemic, many public health agencies and pro-vaccination groups worked through the Twitter social media platform to disseminate good health habits and encourage and enhance the uptake of vaccination [109].

To identify bots on social media quickly and at scale, a series of past work constructed a string of bot detection algorithms [79]. These algorithms make use of a large range of input, from a social media user's text posts, to account features like screen names [15]and account description [70], to account metadata like the number of followers and number of likes [46], to temporal information [26], to network information [37, 38]. There are also a wide range of algorithm architectures implemented in bot detection algorithms: random forests [14], logistic regression [49], convolutional neural networks [34], long short-term memory [102] and other neural network methods [79], deep learning methods [49, 56], temporal-based methods [26] and graph-based methods [37, 38]. Usages of these bot detection algorithms in online social media networks include influence campaign detection [17] and identifying the impact of bot activity on information propagation [67].

Unfortunately, these algorithms are with bias, for many of the datasets used for training these

bot-detection algorithms are comprised of a singular instance of bots (e.g., political, financial), thus training the classifier to differentiate bot/human users results in classifier overfitting, reducing generalizability [85, 99]. The training data are typically manually annotated, whereby experts identify the bot nature of a series of accounts. However, this identification can be subjective, which then result in downstream classifier errors [12].

Bot detection algorithms can also be evaded, just like most feature-based machine learning algorithms. Bot features are constantly evolving over the years, therefore making a robust bot classifier can be challenging [70]. Social bot detection research is currently working on a cat-and-mouse-game model [28], hence we approach this problem through identifying bots by behavioral features, which are harder to evade [107].

To persuade humans, bots use a variety of persuasion techniques. Humans on social media can be swayed by information cascades or social contagion, resulting in an adaptation of their behavior based on their perception of the behavior or thinking of other users [13]. Anti-vaccine bots during the 2020 coronavirus pandemic attempt to persuade people against the the vaccine, increase fear and vaccine hesitancy by: constructing anecdotal evidence or stories; using humor and sarcasm; participating in conversations; and questioning the information [84]. During their attempt to persuade and influence users, bots probe the socio-cognitive biases of individual human users, including seeking belief-consistent information and homophily with connected neighbors [27]. Exposure to social bots amplifies perceptual cognitive biases [106], and manipulation of these perceptions can drive behavioral changes such as the adoption of different policies or opinions [90, 103].

Bots also exhibit signs of collaboration among a community of accounts, where they form cliques [61]. This concept is measured via synchronization of social media mechanics. That is indicative from a high frequency of same sets of hashtags [100], or same sets of URLs [68] or similar sentences [75]. The deliberate synchronization between users is termed as coordination. Coordinated groups of bot accounts working together can manipulate the online discourse [75]. Multiple bot accounts can coordinate the posting of a message simultaneously or at staggered intervals to achieve mass dissemination of the message across a wide network. Such bot networks have been observed in the 2017 French presidential elections, sowing discord against the president Emmanuel Macron with the amplification of the MacronLeaks campaign [39]. Coordinated groups have also been observed to artificially manipulate information about elections online and boost political identities, such as during the 2018-2019 Italian elections [42].

Coordination by bots in the online space is a multi-dimensional problem. Current techniques to uncover coordinating users fundamentally make use of discovering anomalously high levels of synchronized actions within a specific time window. This involves identifying users through the use of a defined behavior trace that links two users within the same space [80]. Common social media behavior that are used include: same retweets [100, 101], same user @-mentions [62], same URLs [22, 42, 64] or similar texts [75]. These techniques are useful to preempt offline protests, such as in the case where hashtag coordination had captured changes between online coordination and offline protests in countries affected by the 2011 Arab Spring protests [87].

Finally, the coordination of pressures from social media bots can influence users to change their opinions. Users have been observed to change their stance towards vaccination (from pro-vaccine to anti-vaccine and vice versa) if their communication network consists of neighbors that are of the opposite stance and are coordinating together [69]. This indicates that it is possible to

manufacture artificial consensus through automated bot users in order to manipulate or influence online opinion [81]. Bots in a bot network can work together to provide humans with multiple exposures to an intervention (i.e., encouraging positive human interactions) [63]. This can be done by exposing the human user to their own content as well as content from other bots within the same network, all sharing the same message.

Chapter 2

Data and Tools

2.1 Data

This thesis leverages on several large-scale datasets to understand bot characteristics in social media. This thesis makes use of a hybrid data collection, where some datasets were obtained from a repository, others were self-collected, and still others had a mix of obtaining certain portions from a repository and supplementing that with additional data collection. The datasets are summarized in Figure 2.1. In total, this thesis makes use of data from 4 social media platforms (Twitter, Facebook, Instagram and Reddit), covering 10 different events across the world (i.e., Canada, France, Kashmir, United States). These datasets consists of over 200 million social media users and over 5 billion social media posts.

2.1.1 Twitter[1] Datasets

OSOME Bot Dataset (hybrid collection) is a series of datasets hosted on `https://botometer.osome.iu.edu/bot-repository/datasets.html`, which consists of expert annotated data of bot and human accounts in domains like political, entertainment and financial bots. Due to Twitter's Terms-Of-Service, only the account ID was shared on the OSOME website. To form the dataset, complete with the user's tweets and metadata, we rehydrated the datasets in June 2021, collecting 40 tweets per account using the Twitter V1 API for data collection. I chose 40 tweets because a prior study performed a systematic analysis on the stability of bot classification showed that 40 tweets is a reasonable collection size for a consistent bot probability score [77].

Asian Elections (obtained from repository) follows the elections in Philippines, Indonesia, Taiwan and Singapore that occurred during 2019 and 2020 [95, 96].

2018 Black Panther Movie (obtained from repository) was Marvel Studio's first superhero film with a strong female lead. This dataset follows the online discussion surrounding gender diversity and misinformation about views from the actors [8].

[1] Although Twitter is recently renamed to X, I will still refer to the social media platform as Twitter throughout this thesis, because the developer API is still named as Twitter Developer API

Twitter	**Dataset Name**	**Details**	**Ch 1**	**Ch 2**	**Ch 3**	**Ch 4**	**Ch 5**
	OSOME Bot Dataset *^	Users: 86k, Posts: 3.4mil	√				
	2018 Black Panther Movie *	Users: 1.6mil, Posts: 17.7mil	√	√	√		
	Asian Elections *	Users: 951k, Posts: 4.1mil	√	√			
	2019 Canadian Elections *	Users: 1.9mil, Posts: 18mil	√	√			
	2019-2020 US Elections *	Users: 1.6mil, Posts: 55mil	√	√	√	√	
	2020-2021 Coronavirus *^	Users: 208mil, Posts: 4.2mil	√	√	√	√	√
	2020 ReOpen America *	Users: 201k, Posts: 4.4mil	√	√	√		
	2021/ 2023 French Protests *^	Users: 343k, Posts: 644k			√		
	2023 Chinese Balloon ^	Users: 121k, Posts:1.2mil		√		√	
Reddit	2022 Reddit ^	Users: 667, Posts: 13k	√	√			
Instagram	2022 Instagram ^	Users: 1935	√	√			
Facebook	2022 US Elections *				√		

Figure 2.1: Summary of Datasets used in this thesis. $*$ indicates the dataset was collected from a repository. $\wedge$ indicates the dataset was self-collected. $*\wedge$ means part of the dataset was obtained from the repository but collection had to be done to obtain the full data.

2019 Canadian Elections (obtained from repository) took place on 21 October 2019. The Liberal party won the vote and Justin Trudeau become the Prime Minister. The dataset follows around six months of online campaigning and discussion about the election [55].

2020 US Elections (obtained from repository) dataset follows the United States elections from the Primaries to the aftermath of the voting [71]. The Democratic party won the election and Joe Biden was named the 46th President of the United States.

2020-2021 Coronavirus (hybrid collection) is a collection of tweets that stemmed from the coronavirus pandemic during 2020-2021. This dataset follows one year of discourse on the health pandemic. Most of the dataset is contained obtained from the lab's central data repository, but I had collected some portions related to conspiracy theories and the vaccine to supplement it.

2020 ReOpen America (obtained from repository) protests were launched across the United States against the government lockdown response to the coronavirus pandemic. The dataset follows three months of Twitter discourse during the heightened protests emotions [62, 71].

2021/2023 French Protests (hybrid collection) dataset followed the protests in 2020 to 2021 that revolved around the vow from French President Emmanuel Macron to protect the right to caricature the Islamic prophet Muhammad as a cartoon, and the protests in 2023 revolving around the pension reformed signed by the French President Macron. The 2021 dataset was obtained from a repository, while the 2023 dataset was self-collected.

2023 Chinese Balloon (self-collected) dataset followed the online conversations on Twitter about the Chinese balloon spotted over the US airspace in January 2023. The US announced it was a surveillance balloon while China maintained it was a weather balloon.

2.1.2 Reddit Dataset (self-collected)

The Reddit dataset was curated in 2022. For bot accounts, we downloaded the 500 highest ranked "bad bot" in BotRank [2], a crowdsourced list of bot ranking [92]. For the humans, we collected users from 5 subreddits that generally require conscious writing and manually verified that the users are likely to he humans. We use the PushShift API [11] to collect data for this dataset.

2.1.3 Instagram Dataset (self-collected)

The Instagram dataset was curated in 2022 through a manual collection, from an observation of a group of accounts that followed an Instagram account within a few hours of the same day.

2.1.4 Facebook Dataset (self-collected)

was curated using CrowdTangle search tool for the 2022 United States midterm elections. We use the Python Crowdtangle API [3] to collect this dataset.

2.2 Tools Used

This section describes computational tools are used throughout this thesis to identify bots, characterize their activity and their interactions.

ORA is a dynamic network analysis and visualization tool with capabilities to import data from several social media sites [23]. It is used in this thesis to handle calculations of social network metrics such as centrality calculations, community detection and visualizations.

Psycholinguistic Analysis is done using the NetMapper software [23] and the LIWC software [91]. These softwares produce lexical counts of well-studied psycholinguistic features like pronouns, emotion words, identity terms and so forth. These tools make use of dictionary and supervised machine-learning based methods to extract the relevant features. The tools were constructed for short texts like Tweets by training on the corresponding data types, and thus should be generalizable across the social media texts studied in this thesis. Features returned from these software are used for understanding linguistic properties of bots and humans and building machine learning and simulation models used in this thesis [69, 70].

[2] `http://botrank.pastimes.eu`
[3] `https://github.com/UPB-SS1/PyCrowdTangle`

Bot Detection Algorithms are crucial to segregating the social media system into bots and humans. This thesis employs several bot detection algorithms to generate probabilities of an account showing automated activity. The probabilities range from 0 to 1, where a probability closer to 1 represents a higher likelihood the account is a bot; and a probability closer to 0 represents a higher likelihood the account is a human. The primary tools used are: BotHunter [14], BotBuster [70] and Botometer [98].

Construct is a framework for implementing agent-based modeling [58]. It facilitates reading in network data from ORA and the simulation of a social influence model.

Chapter 3

Research Plan

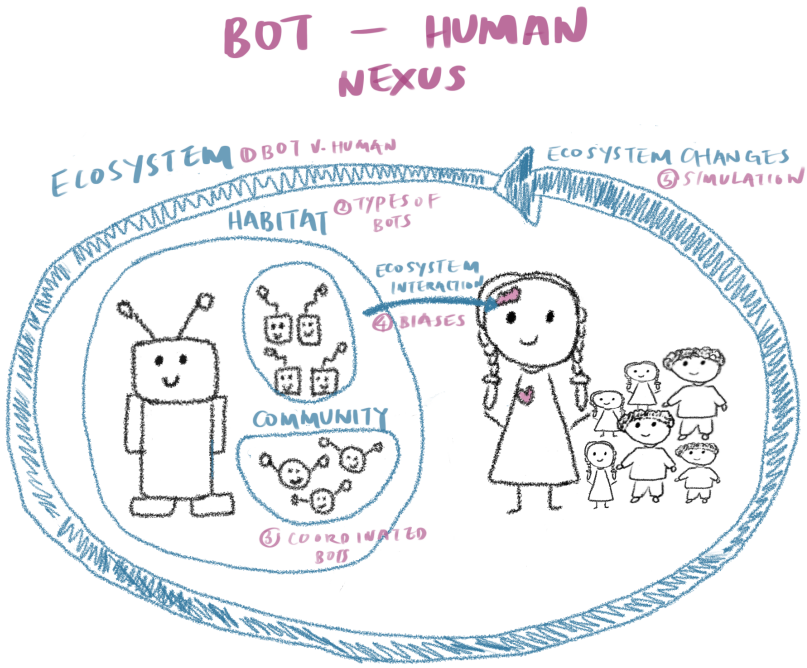

Figure 3.1: Illustration of Thesis and Research Plan

This thesis is organized into five chapters. They present a cumulative view of the bot-human nexus in social media via an ecosystem point of view adapted from natural habitat. Bots and humans are two dynamic creatures that live in the digital social media space. This social media space is an ecosystem (Ch 1), which can be partitioned into habitats of bots and humans (Ch 2), within which we analyze communities of bots (Ch 3). A key ecosystem interaction is how bots target human cognitive biases (Ch 4). Finally, we seek to understand ecosystem changes through simulation (Ch 5). Figure 3.1 illustrates the big picture that this thesis tackles.

3.1 Ecosystem: Bots vs Humans (Ch. 1)

3.1.1 Research Questions

This chapter provides the foundational empirical basis and theoretical understanding of the understanding of the two key species in social media spaces: bots and humans. This chapter works at defining the core of this thesis: the bot. It builds on past research to develop a robust methodology to identify the bot. Then, this chapter works at profiling the similarities and differences of the two species in terms of linguistic cues, self-presented identities and emotional values, and communication network properties. The key research question for this chapter is:

- Can we systematically and efficiently differentiate a bot account from a human account?
- What are the similarities and differences between bots and humans?

3.1.2 Completed Work

Bot detection classifiers While there is a long string of bot detection classifiers [14, 26, 98], these classifiers have two main flaws: dealing with incomplete data and a multi-platform bot detector. Bot detection algorithms take in an account's features and metadata as an input to perform a prediction of whether the account is a bot or human. However, data collection on accounts is usually incomplete, which occurs in fast moving events, in historical data that has already been collected with a smaller set of features or the difficulty of collecting the required set of features for the algorithm (e.g. rate limits). The second issue of a multi-platform bot detector occurs because most bot detection datasets and by extension algorithms, are built for the Twitter platform, resulting in algorithms for other platforms being sparse.

To bridge this research gap, I constructed a multi-platform bot detection classifier that can handle incomplete input data. This falls on a mixture-of-experts concept [70, 82]. To handle incomplete data, each input type (e.g., username, post text, account metadata) is handled by a separate expert. Each expert will then be separately trained on their corresponding data to provide a preliminary prediction for the account based on the specialized subset of data. The predictions for each expert will then be aggregated together to provide a final bot prediction. Thus, if the data is not present for an account, the expert will not be activated, and bot prediction relies on the rest of the experts, thus accounting for incomplete data. The experts should take in data input in a input-agnostic fashion, thus accounting for multi-platform functionality of the data stemming from different social media platforms.

The work on a mixture-of-experts based bot detection algorithm that can be used on Twitter, Reddit and Instagram is published as BotBuster. It is being packaged as a Docker container for portability so other researchers can make use of it.
Lynnette Hui Xian Ng 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*, volume 17, pages 686–697, 2023.

I further test my methods with both deep learning architectures (e.g., LSTMs, CNN) and traditional machine learning algorithms (e.g. random forests, logistic regression) to train the experts. While the hypothesis is that a deep learning algorithm can extract more feature nuances

and therefore have better performance, a traditional machine learning algorithm will be useful for general and fast runs without access to a GPU. My results show that bot detection algorithms do not perform significantly better with deep learning architectures, therefore we can make these algorithms more accessible by using random forests for each data model.

The work on comparing deep learning and traditional machine learning based models is submitted for review at:
Lynnette Hui Xian Ng and Kathleen M. Carley. Assembling an ensemble for bot detection with applications in US 2020 elections. Social Network Analysis and Mining, 2023.

Lastly, I performed empirical analyses across large scale datasets, i.e. the coronavirus and U.S. elections datasets, to evaluate what is a best threshold value to set for bot detection algorithms. Bot detection algorithms typically need a threshold value, where if the probability of a bot of an account is above the value, the account is deemed as a bot. In literature, a wide range of values from 0.2, 0.5 to 0.7 have been used [18, 67, 98]. However, there should be a threshold value defined that will provide the most stability in a user's bot score across time. To determine these threshold values, we evaluate the change in bot probability scores across an increasing number of posts, finding the values at which there is least random variation of bot probability scores.

The work on a systematic analysis of bot classification scores is published at:
Lynnette Hui Xian Ng, 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*, 28:100198, 2022

Differences between bots and humans I annotated the bot probability score of each account within the Twitter-based data in our datasets. An average of 20% of the user population of the dataset are classified as bots by the BotHunter detection algorithm. A preliminary manual investigation on the differences in linguistic cues, social identities and network communication structures was done. This work has yet to be published. It will be combined with the proposed work on this section.

Expression of emotions is a social process – emotions tend to be elicited by and expressed towards others, and regulated to influence other people or comply with societal norms [97]. I examined the differences in emotions expressed by bots and humans within a protest against a curfew in the Kashmir region. The results show that bots express a subset of emotions compared to humans, and are extremely prolific in expressing simpler emotions like sadness and shy away from more complex emotions like disgust and anticipation [67].

This work has been published in:
Lynnette Hui Xian Ng and Kathleen M Carley. Bot-based emotion behavior differences in images during kashmir black day event. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pages 184–194. Springer, 2021

I also investigated the topic differences between bots and humans within two different events.

I do this using the Latent Dirichlet Allocation method of identifying topics [16], which uses a Gaussian model to decompose a series of sentences into topical themes. In terms of textual narratives, I did not observe a huge difference in the discourse perpetuated by bots and humans. The two account types generally used similar phrases, hence expressing the same ideas. However, the bots in different regions, as measured by the user's self-expressed location, express different ideas. This indicates that topic expression is not by species type, but rather by geographical boundaries.

One part of this work has been published in:
Lynnette Hui Xian Ng and Kathleen M Carley. Popping the hood on chinese balloons: Examining the discourse between us and china-geotagged accounts. *First Monday*, 2023

3.1.3 Proposed Work

Bot Detection Classifiers I propose to extend the bot detection classifier an additional social media platforms, Telegram. Since there is currently no annotated dataset for Telegram, I propose to build our own. The data will be obtained from an existing repository of Telegram messages from Russian war related channels. Then, myself and two other students will annotate a subset of the data that will be sampled through a stratified sampling method. The bot/human classification will be the majority class of annotation. Following this, we will use the Mixture of Experts model as in the BotBuster model to construct a Telegram bot detector.

A bot detection algorithm is also not generalizable if it is can only be used on posts on a single language. BotBuster, the bot detection algorithm that I developed, can only be used on the English language. This is not unique as most bot detection algorithms are trained on data from a single language. Therefore, I propose to build a multi-lingual bot detector. To do so, I will make use of language-agnostic language models from the HuggingFace repository[1] to represent text posts as vectors. With vectors generated from different languages, we can then train a bot detection algorithm to interpret different language texts. The text data will come from two sources: annotated non-English bot data [2], and translation of current bot repository texts to other languages.

Differences between bots and humans Bots have been observed to be differentiated by the linguistic cues within their language in their posts [1]. A preliminary investigation shows that bots and humans use different sets of linguistic cues [96]; bots and humans present themselves on social media with different sets of identities and emotional values; and bots have a flower-burst communication network structure, while humans communicate mostly with immediate network before extending outwards. We propose to extract these bots and analyze their bot characteristics through social identities, emotional cues and network properties.

Social identities are terms like doctor, lawyer that represents the affiliation of the user to a social group [54]. On social media platforms, identity presentation and content propagation are related. Tumblr users often reblog content by other users that present similar identities [108] and the self-presentation of identities by Facebook users correlate with post popularity [10]. Studies

[1] `https://huggingface.co`

that investigate harmful information on social media observe that different social groups spread different messages [75]. We propose to extract social identity terms from the user description, assuming the identity that the user presents is the identity he affiliates with. To do so, we use social identity terms from the NetMapper software, and profile the top 50 commonly used identities of bots and humans within each event. Then, we perform statistical analyses to identify whether the differences in the top 50 terms are significant between the two species across events.

Bot accounts are observed to convey feelings through their texts and images [35, 67]. To analyze emotional cues, we use the NetMapper software which extracts the number of words that correspond to each emotion in the text, e.g. anger, joy. We then profile the top 50 commonly used emotional cues of bots and humans within each event, and perform statistical analyses to identify whether the distribution of emotions are significantly different across the events.

Lastly, we investigate the differences between bots and humans in terms of their network communication structure. We construct all-communication networks of the events and analyze key network metrics of bots and humans. These metrics include: betweenness centrality, eigenvector centrality and total degree centrality. Using both quantitative and visual analysis, we aim to profile the differences in social interaction between bots and humans. A preliminary analysis reveals that bots have a more flower-burst shape in communication structure, while humans interact with others in a hierarchical fashion, exhibiting several radii of friendship closeness. Analysis of these metrics will identify whether bots are playing influential roles in disseminating information, or key roles in joining information paths, or have a huge reach.

3.1.4 Challenges and Limitations

A key challenge with bot detection classifiers is that it must continually evolve and improve by training on newer datasets to retain the accuracy. The characteristics of bot users have shifted over time, which calls for continual development and refinement of the algorithms [70]. Also, bot detection classifiers suffer from the lack of expertly annotated data, especially in platforms other than Twitter, for which bot detection classifiers are sparse.

In terms of analyzing characteristics of bots and humans, our methods and analysis presents several limitations. The analysis of linguistic cues in this chapter is based on curated dictionary lists of the NetMapper and LIWC tools. While these lists are based on psycholinguistic research, the digital space is dynamic. As new terms spring up and enter commonspeak, these lists need to be updated frequently and newer sets of data need to be run against the updated lists to capture the language of the time. While self-presented identities is a way of identifying social groups on social media, not all users present their affiliations online. Thus, these users are not accounted for in the analysis. Lastly, network analysis is useful for understanding social communication structures and dynamics, but it does not fully capture all the relevant factors and interactions within a social network. These characterization of the bots and humans are based on simplified structures and further observation are required to construct a comprehensive view of the bot-human differences.

3.2 Habitat: Types of Bots (Ch. 2)

3.2.1 Research Questions

After analyzing the bot-human ecosystem as a whole, this chapter examines the bot habitat in terms of different types of bots. Bots are not homogeneous and are used for different functions in the digital space [4]. Groups of bots are used for different purposes: health communication [4], marketing and advertising [89], and malicious means such as exaggerating social situations to cause panic in emergencies [81].

The key research questions for this chapter is:

- What are the types of bots that live in the social media space?
- What is the habitat that different types of bots live in? That is, what type of actors use each type of bots, and what are their network interactions like.
- What are the characteristics of the different types of bots?

In this chapter, I propose to break down the generic Bot type into several commonly occurring archetypes based on their behavioral characteristics. This expands and harmonizes the taxonomies constructed in previous work which defines subtypes of bots as crawlers, chat bots, spam bots, social bots, sock puppets and cyborgs [43, 88]. I opt for the use of behavioral characteristics as definitions in order to ensure that the bot type definitions can be timeless. I then propose computational methods to automatically identify these bots, which aids in increasing the throughput that bots can be classified into archetypes.

3.2.2 Completed Work

Types of Bots Completed work includes a literature review to determine the different types of bots that have been previously observed, and synthesize definitions in terms of their function and activity. Figure 3.2 shows the current typology of bots and examples of their uses in the social media space.

After synthesizing the definitions, I developed methods to systematically detect these types of bots and profile their activity on social media. From first identifying which users are bots using a bot detection algorithm, I then breakdown the bot users into several commonly occurring types. I developed a classification that characterizes each type of bot in terms of a broad definition, detection methods, and their linguistic and network properties, providing finer-grained information towards that works across social media platforms.

My methodologies for identifying each type of bot are as follows:

- Self-Declared Bot: Parse the usernames, screenanmes, description or other metadata of the bot to identify the term "bot"
- Cyborgs: Identify bots through frequent changes of bot classification, i.e. change of class from bot to human and vice versa.
- News Bots: Since most social media posts about news are usually a news headline with a referral link to the actual news site, I construct a machine learning model that is trained on a dataset of news headlines. This dataset contains news headlines across politics, sports, science, business and so forth [45]. For each bot, run the news model through its posts to

Type of Bot	Definition	Usages
Self-Declared Bot	• Users that outwardly declare themselves as bots	• Pull data from websites (e.g. weather, moon phases) • Announcements (e.g. health directives)
Cyborgs	• Accounts that exhibit both human and bot-like activity	• Used by politicians, CEOs, activists etc to provide periodic announcements and personal touch • Mix up account behavior to prevent suspensions
News Bots	• Automated accounts that post news updates	• Originate news (e.g. news channels) • Aggregate news • Provide news to oneself • Disseminate (fake) news
Announcer Bots	• Automated accounts that announce certain information	• Trigger-based bots for alerts (e.g. price drops) • Content Moderation (e.g. subreddit threads) • Periodic announcement to disseminate news
Amplifier Bots	• Bots that programmatically boost narrative themes and manufacture support	• Buy influence for own account • Create and maintain influence for accounts of leaders • Amplify content with specific narratives • Amplify influence of specific user or groups of users
Repeater Bots	• Bots that excessively repeat posts and keep bulk of the content the same	• Personal reminders • Promotions, advertisements, propaganda
Bridging Bots	• Bots that connect groups together	• Connect groups of users to take notice of their narratives
Content Generation Bots	• Automated users that create content for the online ecosystem	• Integration bots to connect multiple platforms with personal accounts • Chat bots (e.g. customer service, telemedicine) • Mass content generation to push pre-defined narratives

Figure 3.2: Types of Bots and examples of uses

identify whether each post is likely to be a news headline or not. I deem a news bot as a bot that has 80% of its posts being news headlines.

- Announcer Bots: I propose to identify periodic announcer bots as bots that post a message at every time interval. The message pattern is typically templated. I identify these bots by constructing a frequency-domain graph representing the number of posts across time for each bot. Then I transform this into a time-domain graph using the Fast Fourier Transform (FFT) technique. If the FFT graph results in clear peaks, there is a periodic time interval of posts and the bot is thus an announcer bot.
- Amplifier Bots: Construct a share/retweet or mention network where nodes are users and links between users represent that the users share/retweet/mention each other a lot. Amplifier bots are bots that perform these mechanics very frequently, and can be discovered by finding the central users of the network.
- Repeater Bots: Embed the text of tweets into vectorized forms, by using multilingual language models from the Hugging Face repository (`https://huggingface.co/docs/transformers/multilingual`), then performing a all-pairs comparison between all the vectors to identify same/similar vectors. Bots that frequency post similar vectors are deemed to be constantly repeating the same messages.
- Bridging Bots: Split an all-communication network into groups by the Louvain clustering algorithm [32], which finds cliques of users based on the strength of their ties in the network. Identify bridging bots as bots that communicate between two groups.
- Content Generation Bots: Parse the ratio of original content to shared contend (shares/retweet) a bot posts. A content generation bot posts a large proportion of original content.

Work on identifying and analyzing Cyborgs is under review at:
Lynnette Hui Xian Ng, Dawn C Robertson, and Kathleen M Carley. Cyborgs for strategic communication on social media *Big Data and Society*, 2023

Work on Repeater Bots is to appear at:
Charity S Jacobs, Lynnette Hui Xian Ng and Kathleen M Carley. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation,* 2023 To appear.

3.2.3 Proposed Work

Types of Bots Currently, the types of bots are characterized by separate identification algorithms. It will be useful to consolidate all these algorithms into a single tool that begins with the determination of bot/human of an account. I propose to develop an all-in-one bot type detection tool that combines the detection methodologies of all the separate detectors for each bot type. This tool should take in information about a social media user and return the type(s) of bot the user is. The tool should be platform agnostic and be able to take in data from all the platforms studied in this thesis.

Habitat and Characteristics of Types of Bots After profiling each type of bot, I will construct case studies to showcase the key properties of the bot type using our collected data. This step also validates the ability of our methodology in detection of each type of bot. I will analyze characteristics via the comparison of linguistic cues and the patterns of network interaction between the bots and other users (both bots and humans). This involves investigating and identifying the different types of bots throughout the separate events and across the many social media platforms in terms of their linguistic and network differences.

3.2.4 Challenges and Limitations

One key limitation of this typology is that the archetypes of bots and their profiles can evolve along with the evolution of social media platforms and the digital era. The typology hence needs to be constantly scrutinized and updated in order to keep up with the times. Similarly, because many of the methodologies rely on observation and training machine learning models on collected datasets, these algorithms need to be retrained alongside the evolving definitions.

While efforts are made to construct as comprehensive a typology as possible, there are likely to be bot types that we have not observed and thus are unable to profile. This could be because of the scope of data collection efforts, resulting in the restriction of observed bot types, or it could be a lack of awareness and observation on our part.

3.3 Community: Coordinated Bots (Ch. 3)

3.3.1 Research Questions

No bot is an island. This chapter examines the clique formation of bots in coordinating the information and influence spread. Coordination is the deliberate synchronization of users across time, space and narratives. Coordinated groups on social media can pose a threat to the social fabric through the organization of campaigns and protests [87]. Analysis of 16 countries revolving around the 2011 Arab Spring protests show there is a correlation between online synchronization and offline protests [87], and an analysis of similar texts in the 2011 United States Capitol Riots reveal groups of bot user clusters supporting disinformation narratives and themes, alongside an actual riot echoing some of the themes present on social media [75].

The key research questions for this chapter is:

- How do bots synchronize with each other to disseminate information
- How do bots coordinate together to increase influence?

This chapter proposes to develop methods to analyze coordination across three dimensions (time, space and narratives), cumulating in a Combined Synchronization Index to measure the extent which an individual user coordinates among other users.

3.3.2 Completed Work

Coordination across time Several studies analyzed coordination between users across time [62, 80, 101]. This involves setting a specified time window and identifying users that perform a certain social media action (e.g., post with the same hashtag, post with the same @mention) frequently within the same time window as synchronizing users.

I adapted this methodology across multiple events in the United States to identify coordinating users. I then investigated further into the users and profile the types of users that coordinate across multiple social media actions [68].

These work on coordination across time has been published in the following:
Thomas Magelinski, Lynnette Ng, and Kathleen Carley. A synchronized action framework for detection of coordination on social media. *Journal of Online Trust and Safety*, 1(2), 2022
Lynnette Hui Xian Ng and Kathleen M Carley. Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In *Proceedings of the 14th ACM Web Science Conference 2022*, pages 12–21, 2022

Coordination across space Coordination among users can take place across space, namely across social media platforms. This cross-platform coordination can be analyzed using similar texts or similar URLs [75]. There are a variety of users that spread discussion across platforms about the 2020 US election fraud and incite protests: bidirectional introducers, repeat introducers and cross-platform linkers [64].

I analyzed coordination by bots across space by first extracting bot users and the texts they post. Then, I converted the text into a contextual vector representation. The vector representa-

tion facilitates the next step, all-pairs vector comparison. With this vector data, I can identify similar texts, and by extension the users that spread those similar texts. In order to process the large amount of vector data, I used FAISS, a library for efficient similarity search developed by Facebook [52]. From identifying similar texts, I constructed a network where the nodes are texts and the links are weighted by the similarity of texts. I then deduce the user-user network, where the nodes are users, and the links represent the total similarity weight that the texts between two users have. I then analyzed the network structure and identified important coordinating users and clusters within the network.

We performed this analysis of coordination across space with Twitter and Parler data surrounding the 2021 US Capitol Riots.

These work on coordination across space has been published in the following:
Lynnette Hui Xian Ng, Iain J Cruickshank, and Kathleen M Carley. Cross-platform information spread during the january 6th capitol riots. *Social Network Analysis and Mining*, 12(1):133, 2022

Coordination across narratives One dimension of narrative coordination is coordination through the use of similar texts. Narrative coordination identifies users that post similar texts within the scope of conversation as synchronized users. To analyze coordinating users, I first represent texts a vectorized form, e.g. using BERT vectorization techniques [31]. With all the texts vectorized, I compare them to each other in a pairwise comparison and rank the similarity between the pairs using the Euclidean distance measure [74]. Then, I employ network science techniques to construct a text-text network graph, where nodes represent texts and two texts are linked together if they are at least 70% similar. From this, I can derive user-user network graphs, where nodes are users, and two nodes are linked together if they have texts that are at least 70% similar, providing avenues for further insights on the influential coordinating users within the network.

Another dimension of narrative coordination is coordination through the use of images. Image coordination identifies users that post similar images within the conversation as synchronized users [76]. Work in analyzing themes of similar images include finding image clusters in a activist event [67], and the analysis of groups of images shared by state-sponsored Russian bots during an influence campaign [110].

I represent images in a vector form using image representation techniques like ResNet50 [47]. With each image represented as a vector, I find similar images by performing an all-pairs comparison, calculating the Euclidean distances between the image vectors. I incorporate network techniques to form a image-image network graph where nodes are images and two images have a link with each other if they are at least 70% similar in terms of their image vector representation. From this image-image network graph, I form a user-user network graph, where nodes are users and links between two users represent that they have very similar images. This adapts the methodology from past work which constructs network graphs representing the similarity of political images disseminated by Russian bots, whereby these bots are highly effective in sowing discord using image coordination techniques [110].

I analyzed coordination across narratives with datasets such as the ReOpen America dataset, the U.S. elections datasets and the Coronavirus datasets. I further analyzed the properties of the users that coordinate using images, investigating the common countries, languages and so forth

of these users.

Work completed on coordination across narratives has been published in:
Lynnette Hui Xian Ng, Iain J Cruickshank, and Kathleen M Carley. Coordinating narratives framework for cross-platform analysis in the 2021 us capitol riots. *Computational and Mathematical Organization Theory*, pages 1–17, 2022
Lynnette Hui Xian Ng, JD Moffitt, and Kathleen M Carley. Coordinated through aweb of images: Analysis of image-based influence operations from china, iran, russia, and venezuela. *arXiv preprint arXiv:2206.03576*, 2022
Lynnette Hui Xian Ng and Kathleen M Carley. Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In *Proceedings of the 14th ACM Web Science Conference 2022*, pages 12–21, 2022

Combined Synchronization Index Finally, this chapter seeks to develop a Combined Synchronization Index which serves to measure a user coordination across the different coordination dimensions. This index provides an overall quantification of coordination across dimensions within an event, which allows for ranking of users. I performed a study across six Twitter datasets that show that bot-bot pair exhibit the most synchrony [71]. The harmony and dissonance of the index with network centrality values can be used to observe the presence of organic and inorganic coordination, providing insights into the species that are actively amplifying messages surrounding the event.

The formulation of this index has been integrated into the ORA software under "Coordination Analysis" report. It has been published at:
Lynnette Hui Xian Ng and Kathleen M Carley. A combined synchronization index for evaluating collective action social media. *Applied network science*, 8(1):1, 2023

3.3.3 Proposed Work

Coordination across time Much of the coordination work across time involves defining a time window, whereby users that perform the same action within that time window are regarded as coordinating. However, the definition of this time window has been fluid, ranging from 5 minutes [62, 101] to 30 minutes [68]. I propose to investigate and define an appropriate time window that should be used to bound the extraction of coordinating users. Past work have demonstrated that too small a time window results in very little coordinating users, while too large a time window results in a large number of noise [68, 100]. Therefore, the qualitative definition of "coordinating users" and the empirical definition of "time window" needs to be properly defined prior to user analysis in order to accurately extract the bot users that are coordinating for downstream analysis.

Combined Synchronization Index With the formulation of the Combined Synchronized Index, it is possible to compare the extent of coordination of bots, and even different types of bots in different events. I propose to calculate the Index of users in the suite of event datasets that I have collected. I would then profile the type of bot per user and perform statistical tests that

provides understanding towards the nature of synchronization of different types of bots. For example, I would expect that Announcer Bots do not synchronize much as their primary function is to broadcast events; while News Bots will have a mixture of high synchronization where they will mention other related news outlets or aggregate posts from a certain group of news outlets, and low synchronization of original news bots that post news fresh off the press. I would also identify bot users that appear in more than one event, and analyze their synchronization index, thereby providing insight to whether users in different events coordinate differently, and the patterns of coordination in different event types.

3.3.4 Challenges and Limitations

One limitation with regard to constructing bot detection algorithms is with respect to the training dataset. The algorithms are trained on the OSOME bot repository datasets. While this large training dataset lends weight to the generalizability of the algorithm, the algorithms might not be entirely effective in identifying bots in other unseen domains, such as the protests and coronavirus events that were analyzed in this thesis.

Another limitation is that much of this coordination is calculated in terms of similar high frequency activity within a specified time window. There are temporal nuances to time window specification in detection of coordination: too small a time window results in high coordination [100], while too loose a time window results in sparse coordination [68].

3.4 Ecosystem Interaction: Biases (Ch. 4)

3.4.1 Research Questions

After studying bots as an individual species, this chapter studies the interaction between both bot and human species. Social media bots have been known to persuade humans, for example, convince humans to join an activist cause [83], or sway humans on political stances [60]. To understand what makes their persuasion effective, I propose to study their persuasive techniques in terms of the biases they employ. Biases are systematic inaccuracies, that therefore leads to certain behavior. This includes structural, social-cognitive and cognitive biases. Structural biases is often also described as population biases, where different demographic slices of the population (e.g. gender, race, income, education etc.) react differently [51]. Social-cognitive biases refer to how different social groups interpret the same information differently [21]. In the 1970s, the term "cognitive bias" was coined to describe the human systematically flawed patterns of responses to judgment [94], thereby creating their own versions of social reality based on their sensory input [44]. Cognitive biases can be manipulated which exacerbate their negative effects, and have been shown to be influenced in information seeking behaviors and outcomes [7].

The key research questions for this chapter is:

- What are the human biases that Bots target? This includes structural, social-cognitive and cognitive biases.
- Do different types of bots target different biases?

3.4.2 Completed Work

This chapter proposes to profile the cognitive biases that bots leverage on to spread their messages and influence humans. I propose to profile the biases in terms of the TTP framework: the employment of general intent (Tactics), the methods employed (Techniques) and the step-by-step implementation process (Procedures) [48].

The completed work within this chapter involves some literature review.

3.4.3 Proposed Work

To analyze the methodology that bots use to spread their messages, I propose to profile biases in terms of their TTPs. This will be done by analyzing multiple datasets to identify the TTPs that bots used to target human cognitive biases in order to disseminate their messages. In order to make my work generalizable, I will perform this analysis across several social media platforms: Twitter, Facebook and Instagram. This will demonstrate that regardless of platform structure and mechanics, the types of biases that are typically used for information propagation are similar.

We will conduct initial empirical analysis to systematically uncover these biases across social media platforms. To do so, we propose using a series of techniques that range from natural language processing methods of similar text matching [75] to network science methods of detecting amplification through excessive retweet/sharing. These methods have been established in the previous chapter which was used to identify different archetypes of bots through their social media mechanics. We then propose to use qualitative analysis to group and identify the TTPs that the bots use in harnessing the cognitive biases in their bid to disseminate their messages.

Structural Biases We plan to study structural biases by studying two slices of society: gender and identities. We would infer the identities users affiliate themselves with based on identifying self-presented identities, which are identities that social media users write about within their profile information. These identities include information about gender (i.e., male, female, mom, dad) and social identities (i.e., artist, writer, scientist). These identities are derived from a lexicon curated from a population census [73].

Past work has shown that people of different ages and genders tweet differently, and understanding of such structural biases can lead to better understanding of algorithmic biases [51]. Bots are also observed to exhibit some form of biases, e.g. chat bots do exhibit gender biases [36]. We thus build on this literature to examine the presence of biases exhibited through the language uses in bots and profile them in terms of TTPs.

Cognitive Biases We propose to study 6 cognitive biases within two large categories: Information Overload and Societal Bias. These cognitive biases are adapted from past work on combatting fake news [59]. Figure 3.3 shows the cognitive biases and the TTPs that we propose to be studied within this thesis.

Social-Cognitive Biases In terms of social-cognitive biases, we plan to study how different social groups interpret information differently, and therefore how bots change their language

	Cognitive Bias	Definition	Type of Bot	Tactic	Technique	Procedure
Information Overload	Illusory Truth Effect	Prioritization of familiarity over fact	Repeater Bots	Create an illusion that the message spread is truth	Excessively repeating the message to flood the landing pages of human users	Excessive posting of the same message; work together as a group to post a variant of same message
	Multiple Source Effect	Presenting same content from multiple sources	Amplifier Bots	Create illusion that message is credible because many sources share the same content	Share content from multiple sources with same messaging/ stance	Share the same content from established outlets
	Motivated Reasoning	Subscribing to the same view as the others around to reduce cognitive dissonance	General Bots	Create illusion that there is consensus	Identify prevailing stance that a bot is central to and change stance	Change presented stance through message posted
Societal Bias	Authority Bias	Look towards trusted sources or depend on amount of support an opinion has to determine credibility	Aggregator Bots	Establish credibility through using authority sources	Shares information from established sources and authority figures	Quoting authorities or established sources within posts
	Homophily	Associate oneself with similar others	Content Generation Bots	Present an identity that matches the group of users	Identify the dominant identity of group and create content that conforms	Generating content that matches a particular identifier of community of users
	Availability Cascade	Collective belief gains more plausibility through increased repetition	Persuasive Bots	Persuade others through memorable texts	Increase ease of recollection of fake news	Using narratives and memorable quotes to increase memorability of text

Figure 3.3: Cognitive Bias and TTPs that Bots use

and interaction based their targeted social groups. The social groups we plan to look at include groups that present different types of ideologies, e.g. pro- and anti-vaccination groups.

3.4.4 Challenges and Limitations

One key limitation is that in our review of biases, it is impossible to discuss all the different techniques that emerge from bot automation. Therefore, we pick the most salient biases that are common across multiple social media platforms. We hope that our discussion points help the field to critically reflect upon the techniques of bots and provide a starting point for future research in this area.

For each type of bias, the TTPs used by different types of bots may vary and the framework needs to be adapted or modified to capture the unique characteristics for employed by different community of bots.

3.5 Ecosystem Changes: Simulation as a test bed (Ch. 5)

3.5.1 Research Questions

The final chapter of the thesis integrates work from earlier chapters to synthetically generate the activity and interactions within the social media ecosystem. This chapter contributes to ongoing efforts of projecting effects of bot activity in the social media space [6, 63].

The key research question in this chapter is:

- How much pressure from Bots is required to change the ecosystem?

3.5.2 Completed Work

Completed work observed that there are users within the twitter space that change stances with respect to their opinion towards the 2020 Coronavirus vaccine (pro-vaccine, anti-vaccine) [69]. This work observed that users are more likely to change stances if they are surrounded by co-ordinating bots of opposite stances. This work built a social influence model to profile users in terms of their intrinsic properties (measured by linguistic properties of the texts in previous posts) and extrinsic environment (network connections with other social media users) and predict the probability the users will change stances.

This model has been published in:
Lynnette Hui Xian Ng and Kathleen M Carley. Pro or anti? a social influence model of online stance flipping. *IEEE Transactions on Network Science and Engineering*, 10(1):3–19, 2022

Another work completed analyzed the circumstances which social influence operations are likely to succeed. To do so, we intentionally provoke the stances on a simulated network and analyze the trade off between perturbing stances and maintaining influence. The results show that influential agents are the best types of agents to provoke to cause changes (similar to how Taylor Swift can get people to vote), and the most effective and widespread change happens with cascading of local ego networks.

This work is to appear in:
Peter Carragher, Lynnette Hui Xian Ng and Kathleen M Carley. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 2023. To appear.

3.5.3 Proposed Work

I propose extending current research to on flipping stances and stance perturbation strategies to build a bot simulation model [69]. Past work on stance perturbation strategies investigated that influential agents are the best at causing an overall stance change throughout a scale-free network, and about 20-25% of agents need to perturb the network to provide sufficient pressure for a stance change.

I propose to build a social influence model [41] to characterize the users. This model will contain of two parts: (1) intrinsic variables which will be obtained by linguistic values of posts that the users put forth, i.e. number of pronouns, number of angry words; and (2) extrinsic variables which will be obtained by centrality measures of an users within an all-communication network. The initial base networks will be formed through the suite of datasets I have collected, thus relying on real-world data as an input. As such, the input data will mimic real-world data and the simulation will be more realistic.

For this simulation,Iwe will focus on the change in stance towards a key topic of the dataset, e.g. for the US elections dataset, the key topic will be support for democrat/republican. At each

time step of the simulation, the user stances will be re-evaluated in accordance to their intrinsic and extrinsic values. Their stances will be changed in accordance to the algorithms if need be. I will run the simulation to analyze if stances in a social network environment will converge to a single stance, and investigate the properties of key users that will put high pressures on flipping stances.

Another scenario to run is to manipulate the values of bot accounts. For example, artificially increasing or decreasing some intrinsic values to simulate the bots posting excessively and expressing those particular values (e.g. amplifier bots). Similarly, a third scenario manipulates the values of extrinsic variables to simulate the increase or decrease of influence of bot accounts. This simulation scenario will tell us the effect of bot accounts on the ecosystem.

To evaluate the accuracy of the simulation model, I propose to validate with real-world data from our collected dataset. Thus, I propose to process our collected datasets at several timesteps, e.g. every month. Then, I will run our simulation model till the same time step to obtain simulated user data. I will then compare the differences in the number of users that express each type of stance between the simulated and the real-world dataset.

3.5.4 Challenges and Limitations

The primary challenge of building a social influence model to simulate the pressure from bots on opinion change is the lack of data. In my past study on flipping stances towards the Coronavirus vaccine, only 1% of the users in the dataset are observed to flip stances [69]. Therefore, the dataset of users that change opinions on social media can be rather sparse, which makes validation of the simulation against real-world data difficult.

A key limitation of observation of ecosystem changes is that users with extreme opinions are typically more vocal on social media. Therefore, the model and simulation are more likely to be favor the stances of the vocal group and will not capture the opinions of the silent majority [65].

Another limitation is the validation of synthetic scenarios and the applicability to real-life. While the construction of simulation parameters and initial social network is based on the studies in the preceding chapters, these synthetic scenarios may not fully capture the complexity and nuances of real-world persuasion from bots and opinion changes. There may be confounding variables not discovered or accounted for within this thesis. It is also unethical to probe a real network to validate our hypotheses, thus we need to rely on corresponding our results with collected data from different timesteps.

Chapter 4

Contributions

4.1 Theoretical Contributions

This thesis presents several interdisciplinary theoretical contribution to understanding social media bots. Fundamentally, we reframe the study of bots in the online space in terms of an ecosystem-habitat-community point of view. This showcases bots as a species in the information environment from a macro perspective to the micro communities. This thesis further recognizes the two interconnected species (bots and humans) in social media platforms, and provides a wide lens to understand the characteristics of each species, alongside their interactions within the ecosystem.

The coverage of this thesis across global and regional events and four social media platforms acknowledges the dynamic nature of the information environment and the interaction between bots and humans. By connecting several social media platforms together, this thesis builds the foundational mechanics and properties of social media bots. After establishing this mechanics and properties of bots, this thesis proposes to develop a simulation model to project the effect of bots on the digital social space [69], benefiting existing work with insights beyond the static time period.

From a social cybersecurity standpoint, I strengthen connections between rich literature of bot detection with cross-platform bot detection algorithms [70] and studies of bot detection algorithms [77]. I build on past research of bot identification to synthesize a bot typology to characterize bots by their online behavior and actions, providing further granularity into the variations and archetypes of bots. I expand the scope of studying coordination in social media beyond looking at temporal means [71, 74] to identifying narrative [74, 75] and image coordination [76].

Overall, this thesis offers new ways of thinking about the relationship of bots and humans within the social media space which contributes to the nuanced understanding of the personality of bots within the social media ecosystem. The proposed range of approaches and data within this thesis contributes to a comprehensive understanding of automated bot users in the information environment, providing theoretical insights that underscores the unique roles and interactions different types of bots play in the information society.

4.2 Methodological Contributions

This thesis contributes a series of interoperable methods for characterizing bots in social media platforms. These tools range from bot detection tools (BotBuster, BotBuster For Everyone) that have been implemented in the CASOS servers to tools that characterize types of bots (News Bot detector, Cyborg Hunter), to tools that analyze the degree of user synchronization (Synchronized Action Framework, Combined Synchronized Index) which have been integrated in the ORA software. Figure 4.1 summarizes the completed and planned tools in this thesis. These tools have been used across a wide range of events and social media data within this thesis, showcasing their versatility and interoperability.

In terms of detection of bots, this thesis goes beyond the binary bot classification which differentiates bots vs humans. Besides constructing a novel bot detection algorithm that can provide multi-platform classification [70], it further provides finer-grained differentiation of the type of bots. This novel classification provides insights into the personality and tactics of the bot user.

Furthermore, this thesis integrates empirical findings into a simulation package to project the influence and effect bot activity have on population dynamics. This practical contribution provides a valuable resource for researchers and analysts in the field, offering a method of studying long term effects of the social media mechanics that bots employ.

Tool	Function	Ch 1	Ch 2	Ch 3	Ch 4	Ch 5
BotBuster	Bot Detection by Classification (neural network-based, GPU)	√				
BotBuster For Everyone	Bot Detection by Classification (fast, CPU-based)	√				
News bot detection	Detection of bots that post news		√			
Amplifier bot detection	Detection of bots that amplify information		√			
Cyborg Hunter	Detection of agents that are sometimes-bots, sometimes-human		√			
Repeater Bot Hunter	Detection of bots that repeatedly post messages		√			
Type-Of-Bot tool *(planned)*	All-in-one types of bot detection tool		√			
Synchronized Action Framework/ Coordinating Narratives Framework/ Similar Images Framework	Detection of Synchronization through temporal/ narrative/ image means			√		
Combined Synchronized Index	Identification and ranking of synchronized users			√		
Bot Simulation Tool *(planned)*	Detection of changes in a variable as social network changes due to bot activity across time					√

Figure 4.1: Summary of Tools developed in this thesis

This thesis also offers several dataset contributions. Curating bot datasets for this thesis did not consist only of passive collection. The datasets were enriched with labels representing bot/human probability and linguistic cues [8, 55, 62, 70, 71, 75, 95, 96, 98]. These datasets are also readily available for future research beyond the scope of online hate.

Figure 4.2 shows a summary of the completed and planned contributions for this thesis.

Chapter	Completed Work	Planned Work
Ch 1: Bots vs Humans	• Development of **bot detection by classification** model • **Empirical analysis** over large corpus of dataset	• Refinement and write up of **empirical analysis**
Ch 2: Types of Bots	• **Baseline typology** established • Development of **types-of-bot identification** technologies across large corpus	• Development of **integrated Types-Of-Bots detection model** • Development of **case studies** for each type of bot
Ch 3: Coordinated Bots	• Development of **Framework for identification** and analysis of synchronization • Empirical analysis of **temporal and narrative synchronization**	• Development of framework of synchronization and coordination types
Ch 4: Biases	• **Literature review**	• Development of framework that **harmonizes** Type of Bots with Biases
Ch 5: Simulation as a Test Bed	• Initial empirical analysis of opinion change patterns of bots and humans	• Development of **bot simulation model** • **Execution and analysis** of model

Figure 4.2: Summary of Completed and Planned Work

4.3 Academic Contributions

Work for this thesis has already resulted in a series of publications at journals and conferences. These have presented an analysis of the differences and similarities of bots and humans (The Big Book of Bots), bot detection algorithms and typology of bots (Online Social Networks and Media [77], ICWSM [70]), coordinated bot analysis (Applied Network Science [71], Social Network Analysis & Mining [75]) and simulations of bot activity (IEEE Transactions of Network Science [69]). In addition, a book titled The Big Book of Bots is being prepared for publication with a university press. Further academic submissions will also be prepared based on planned work. Figure 4.3 shows a summary of the published and in-progress academic contributions.

The list of published academic contributions of this thesis are as follows:

Chapter 1: Bots and Humans

- Lynnette Hui Xian Ng, 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*, 28:100198, 2022
- Lynnette Hui Xian Ng 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*, volume 17, pages 686–697, 2023
- Lynnette Hui Xian Ng and Kathleen M Carley. Bot-based emotion behavior differences in images during kashmir black day event. In *International Conference on Social Computing,*

Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pages 184–194. Springer, 2021

- Lynnette Hui Xian Ng and Kathleen M Carley. Popping the hood on chinese balloons: Examining the discourse between us and china-geotagged accounts. *First Monday*, 2023

Chapter 2: Types of Bots

- Charity S. Jacobs, Lynnette Hui Xian Ng, and Kathleen M. Carley. Tracking china's cross-strait bot networks against taiwan. In Robert Thomson, Samer Al-khateeb, Annetta Burger, Patrick Park, and Aryn A. Pyke, editors, *Social, Cultural, and Behavioral Modeling*, pages 115–125, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43129-6

Chapter 3: Coordinated Bots

- Lynnette Hui Xian Ng and Kathleen M Carley. Do you hear the people sing? comparison of synchronized url and narrative themes in 2020 and 2023 french protests. *Frontiers in Big Data*, 6:1221744
- Lynnette Hui Xian Ng, Iain J Cruickshank, and Kathleen M Carley. Coordinating narratives framework for cross-platform analysis in the 2021 us capitol riots. *Computational and Mathematical Organization Theory*, pages 1–17, 2022
- Adya Danaditya, Lynnette Hui Xian Ng, and Kathleen M Carley. From curious hashtags to polarized effect: profiling coordinated actions in indonesian twitter discourse. *Social Network Analysis and Mining*, 12(1):105, 2022
- Lynnette Hui Xian Ng and Kathleen M Carley. A combined synchronization index for evaluating collective action social media. *Applied network science*, 8(1):1, 2023
- Lynnette Hui Xian Ng and Kathleen M Carley. Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In *Proceedings of the 14th ACM Web Science Conference 2022*, pages 12–21, 2022
- Thomas Magelinski, Lynnette Ng, and Kathleen Carley. A synchronized action framework for detection of coordination on social media. *Journal of Online Trust and Safety*, 1(2), 2022

Chapter 4: Biases There is no published academic contributions for this chapter.

Chapter 5: Simulation as a Test Bed

- Lynnette Hui Xian Ng and Kathleen M Carley. Pro or anti? a social influence model of online stance flipping. *IEEE Transactions on Network Science and Engineering*, 10(1): 3–19, 2022
- Peter Carragher, Lynnette Hui Xian Ng, and Kathleen M. Carley. Simulation of stance perturbations. In Robert Thomson, Samer Al-khateeb, Annetta Burger, Patrick Park, and Aryn A. Pyke, editors, *Social, Cultural, and Behavioral Modeling*, pages 159–168, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43129-6

Chapter	Completed Publications	In-Progress Publications
Ch 1: Bots vs Humans	• The Big Book of Bots • Online Social Networks and Media • International AAAI Conference for Web and Social Media • SBP-Brims • First Monday	• The Big Book of Bots
Ch 2: Types of Bots	• The Big Book of Bots • SBP-Brims	• The Big Book of Bots • Big Data & Society (under review) • EPJ Data Science (under review)
Ch 3: Coordinated Bots	• The Big Book of Bots • Journal of Online Trust & Safety • Social Network Analysis & Mining • Computational & Mathematical Organizational Theory • Applied Network Science • ACM Web Science Conference • Social Media + Society	• Frontiers in Big Data, Misinformation and Misbehavior Mining on the Web (under review)
Ch 4: Biases		• 1 journal paper
Ch 5: Simulation as a Test Bed	• The Big Book of Bots • IEEE Transactions on Network Science and Engineering • SBP-Brims	• 1 journal paper

Figure 4.3: Summary of Published and In-Progress Publications

4.4 Limitations

There are several limitations within the scope of this thesis. The first limitation is that the analyses are largely Twitter based due to the availability of annotated data and the ease of large data collection at the time of the studies. While we have also evaluated the presence of bots on other platforms, i.e., Twitter, Reddit, Instagram and Facebook, these studies are much smaller scale than the Twitter studies. With the constant change in API structures and limits of social media platforms, we must constantly shift strategies to continue to study the online space. New platforms and mediums of information communication are discussed and the study of bots must evolve as the digital ecosystem evolves.

Second, the studies are largely English-based, and most datasets are filtered to contain only social media posts in English. This is unfortunately a language barrier. Complementary studies of bots in other languages [3, 19, 29] are suggested by literature, but are beyond the scope of this thesis. Developed tools may be extended to contain multi-lingual properties, and adapted to address the transfer of models across languages.

Lastly, bots are dynamic creatures in the digital social space. They will constantly change and adapt to the evolving social environment. In doing so, they will evade machine learning bot classifiers that are based on identifying common patterns from past input. While their fundamental behavioral characterizations should remain unchanged, their methodologies will evolve alongside new social media platforms and mediums arise, and also when social media platforms place restrictions or leave loopholes.

Chapter 5

Timeline

Figure 5.1 shows the timeline of completed work prior to this proposal.

	Spring 2021	Fall 2021	Spring 2022	Fall 2022	Spring 2023	Fall 2023
Ch 1: Bots vs Humans						
Development of bot classification model		√				
Establishment of bot vs human analysis methodology		√	√			
Empirical analysis over large dataset	√	√	√			
Ch 2: Types of Bots						
Development of baseline typology through literature synthesis			√	√		
Development of types-of-bots identification technologies and applications to large dataset			√	√	√	
Ch 3: Coordinated Bots						
Development of frameworks for identification and analysis of synchronization		√	√	√	√	
Empirical analysis of temporal and narrative synchronization	√	√	√		√	
Ch 4: Cognitive Biases						
Literature Review					√	
Ch 5: Simulation as a Test Bed						
Development of social influence model of opinion change	√	√	√			

Figure 5.1: Completed Work

Figure 5.2 shows my proposed timeline from Fall 2023 through my projected thesis defense in April 2026.

	Fall 2023	Spring 2024	Fall 2024	Spring 2025	Fall 2025	2026
Ch 1: Bots vs Humans						
Refinement and write up of empirical analysis	√					
Ch 2: Types of Bots						
Development of integrated Types-Of-Bots detection model	√	√	√	√		
Development of case studies of Types of Bots	√	√	√	√		
Ch 3: Coordinated Bots						
Development of framework of coordination types		√	√			
Ch 4: Biases						
Harmonization of types of bots with biases		√	√	√		
Ch 5: Simulation as a Test Bed						
Development of bot simulation model				√	√	
Execution and analysis of model				√	√	
Finalize Thesis Document						√
Thesis Defense						√

Figure 5.2: Proposed Timeline

Bibliography

[1] Aseel Addawood, Adam Badawy, Kristina Lerman, and Emilio Ferrara. Linguistic cues to deception: Identifying political trolls on social media. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 15–25, 2019. 3.1.3

[2] Nuha Albadi, Maram Kurdi, and Shivakant Mishra. Hateful people or hateful bots? detection and characterization of bots spreading religious hatred in arabic social media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–25, 2019. 3.1.3

[3] Iuliia Alieva and Kathleen M Carley. Internet trolls against russian opposition: A case study analysis of twitter disinformation campaigns against alexei navalny. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 2461–2469. IEEE, 2021. 4.4

[4] Izzat Alsmadi and Michael J O'Brien. How many bots in russian troll tweets? *Information Processing & Management*, 57(6):102303, 2020. 1.1, 1.2, 3.2.1

[5] Sara Atske. 1. Most Americans have heard about social media bots; many think they are malicious and hard to identify — pewresearch.org. https://www.pewresearch.org/journalism/2018/10/15/most-americans-have-heard-about-social-media-bots-many-think-they-are- 2013. [Accessed 16-Jul-2023]. 1.1

[6] Aldo Averza, Khaled Slhoub, and Siddhartha Bhattacharyya. Evaluating the influence of twitter bots via agent-based social simulation. *IEEE Access*, 10:129394–129407, 2022. 3.5.1

[7] Leif Azzopardi. Cognitive biases in search: A review and reflection of cognitive biases in information retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, CHIIR '21, page 27–37, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380553. doi: 10.1145/3406522.3446023. URL https://doi.org/10.1145/3406522.3446023. 3.4.1

[8] Matthew Babcock, David M Beskow, and Kathleen M Carley. Beaten up on twitter? exploring fake news and satirical responses during the black panther movie event. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pages 97–103. Springer, 2018. 2.1.1, 4.2

[9] Adam Badawy, Aseel Addawood, Kristina Lerman, and Emilio Ferrara. Characterizing the 2016 russian ira influence campaign. *Social Network Analysis and Mining*, 9:1–11, 2019. 1.1, 1.2

[10] Liad Bareket-Bojmel, Simone Moran, and Golan Shahar. Strategic self-presentation on facebook: Personal motives and audience response to online behavior. *Computers in Human Behavior*, 55:788–795, 2016. 3.1.3

[11] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 830–839, 2020. 2.1.2

[12] Victor Benjamin and TS Raghu. Augmenting social bot detection with crowd-generated labels. *Information Systems Research*, 34(2):487–507, 2023. 1.2

[13] George Berry, Christopher J Cameron, Patrick Park, and Michael Macy. The opacity problem in social contagion. *Social Networks*, 56:93–101, 2019. 1.2

[14] David M Beskow and Kathleen M Carley. Bot-hunter: a tiered approach to detecting & characterizing automated activity on twitter. In *Conference paper. SBP-BRiMS: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation*, volume 3, page 3, 2018. 1.2, 2.2, 3.1.2

[15] David M Beskow and Kathleen M Carley. Its all in a name: detecting and labeling bots by their name. *Computational and mathematical organization theory*, 25:24–35, 2019. 1.2

[16] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003. 3.1.2

[17] Olga Boichak, Sam Jackson, Jeff Hemsley, and Sikana Tanupabrungsun. Automated diffusion? bots and their influence during the 2016 u.s. presidential election. In Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett, editors, *Transforming Digital Worlds*, pages 17–26, Cham, 2018. Springer International Publishing. ISBN 978-3-319-78105-1. 1.2

[18] Olga Boichak, Sam Jackson, Jeff Hemsley, and Sikana Tanupabrungsun. Automated diffusion? bots and their influence during the 2016 us presidential election. In *Transforming Digital Worlds: 13th International Conference, iConference 2018, Sheffield, UK, March 25-28, 2018, Proceedings 13*, pages 17–26. Springer, 2018. 3.1.2

[19] Gillian Bolsover and Philip Howard. Chinese computational propaganda: Automation, algorithms and the manipulation of information about chinese politics on twitter and weibo. *Information, communication & society*, 22(14):2063–2080, 2019. 4.4

[20] David A Broniatowski, Amelia M Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C Quinn, and Mark Dredze. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American journal of public health*, 108(10):1378–1384, 2018. 1.1

[21] Laura A Brown and Alex S Cohen. Facial emotion recognition in schizotypy: the role of accuracy and social cognitive bias. *Journal of the International Neuropsychological Society*, 16(3):474–483, 2010. 3.4.1

[22] Cheng Cao, James Caverlee, Kyumin Lee, Hancheng Ge, and Jinwook Chung. Organic or organized? exploring url sharing behavior. In *Proceedings of the 24th ACM International*

on Conference on Information and Knowledge Management, pages 513–522, 2015. 1.2

[23] L Richard Carley, Jeff Reminga, and Kathleen M Carley. Ora & netmapper. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer*, volume 3, page 7, 2018. 2.2, 2.2

[24] Peter Carragher, Lynnette Hui Xian Ng, and Kathleen M. Carley. Simulation of stance perturbations. In Robert Thomson, Samer Al-khateeb, Annetta Burger, Patrick Park, and Aryn A. Pyke, editors, *Social, Cultural, and Behavioral Modeling*, pages 159–168, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43129-6.

[25] Pete Cashmore. Twitter Zombies: 24mashable.com. `https://mashable.com/archive/twitter-bots`, 2009. [Accessed 17-Jul-2023]. 1.2

[26] Nikan Chavoshi, Hossein Hamooni, and Abdullah Mueen. Debot: Twitter bot detection via warped correlation. In *Icdm*, volume 18, pages 28–65, 2016. 1.2, 3.1.2

[27] Wen Chen, Diogo Pacheco, Kai-Cheng Yang, and Filippo Menczer. Neutral bots probe political bias on social media. *Nature communications*, 12(1):5580, 2021. 1.2

[28] Stefano Cresci, Marinella Petrocchi, Angelo Spognardi, and Stefano Tognazzi. The coming age of adversarial social bot detection. *First Monday*, 2021. 1.2

[29] Adya Danaditya, Lynnette Hui Xian Ng, and Kathleen M Carley. From curious hashtags to polarized effect: profiling coordinated actions in indonesian twitter discourse. *Social Network Analysis and Mining*, 12(1):105, 2022. 4.4

[30] Kareem Darwish, Dimitar Alexandrov, Preslav Nakov, and Yelena Mejova. Seminar users in the arabic twitter sphere. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I 9*, pages 91–108. Springer, 2017. 1.1, 1.2

[31] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 3.3.2

[32] Scott Emmons, Stephen Kobourov, Mike Gallant, and Katy Börner. Analysis of network clustering algorithms and cluster quality metrics at scale. *PloS one*, 11(7):e0159161, 2016. 3.2.2

[33] Richard M Everett, Jason RC Nurse, and Arnau Erola. The anatomy of online deception: What makes automated text convincing? In *Proceedings of the 31st Annual ACM symposium on applied computing*, pages 1115–1120, 2016. 1.2

[34] Michael Färber, Agon Qurdina, and Lule Ahmedi. Identifying twitter bots using a convolutional neural network. In *CLEF (Working Notes)*, 2019. 1.2

[35] Michela Fazzolari, Manuel Pratelli, Fabio Martinelli, and Marinella Petrocchi. Emotions and interests of evolving twitter bots. In *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pages 1–8. IEEE, 2020. 3.1.3

[36] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. Gender bias in chatbot design. In *Chatbot Research and Design: Third International Workshop, CON-*

VERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers 3, pages 79–93. Springer, 2020. 3.4.3

[37] Shangbin Feng, Herun Wan, Ningnan Wang, and Minnan Luo. Botrgcn: Twitter bot detection with relational graph convolutional networks. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 236–239, 2021. 1.2

[38] Shangbin Feng, Zhaoxuan Tan, Herun Wan, Ningnan Wang, Zilong Chen, Binchi Zhang, Qinghua Zheng, Wenqian Zhang, Zhenyu Lei, Shujie Yang, et al. Twibot-22: Towards graph-based twitter bot detection. *Advances in Neural Information Processing Systems*, 35:35254–35269, 2022. 1.2

[39] Emilio Ferrara. Disinformation and social bot operations in the run up to the 2017 french presidential election. *First Monday*, 2017. 1.2

[40] Asbjørn Følstad, Petter Bae Brandtzaeg, Tom Feltwell, Effie LC Law, Manfred Tscheligi, and Ewa A Luger. Sig: chatbots for social good. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2018. 1.1, 1.2

[41] Noah E Friedkin and Eugene C Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990. Publisher: Taylor & Francis. 3.5.3

[42] Fabio Giglietto, Nicola Righetti, Luca Rossi, and Giada Marino. It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 italian elections. *Information, Communication & Society*, 23(6):867–891, 2020. 1.2

[43] Robert Gorwa and Douglas Guilbeault. Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, 12(2):225–248, 2020. 3.2.1

[44] Rainer Greifeneder, Herbert Bless, and Klaus Fiedler. *Social cognition: How individuals construct social reality*. Psychology Press, 2017. 3.4.1

[45] Xiaotao Gu, Yuning Mao, Jiawei Han, Jialu Liu, Hongkun Yu, You Wu, Cong Yu, Daniel Finnie, Jiaqi Zhai, and Nicholas Zukoski. Generating Representative Headlines for News Stories. In *Proc. of the the Web Conf. 2020*, 2020. 3.2.2

[46] Kadhim Hayawi, Sujith Mathew, Neethu Venugopal, Mohammad M Masud, and Pin-Han Ho. Deeprobot: a hybrid deep neural network model for social bot detection based on user profile data. *Social Network Analysis and Mining*, 12(1):43, 2022. 1.2

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3.3.2

[48] Department of the Army Headquarters. *ADP 3-90, Offensive and Defensive*, 2019. 3.4.2

[49] Maryam Heidari, H James Jr, and Ozlem Uzuner. An empirical study of machine learning algorithms for social media bot detection. In *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, pages 1–5. IEEE, 2021. 1.2

[50] Charity S. Jacobs, Lynnette Hui Xian Ng, and Kathleen M. Carley. Tracking china's cross-strait bot networks against taiwan. In Robert Thomson, Samer Al-khateeb, Annetta Burger, Patrick Park, and Aryn A. Pyke, editors, *Social, Cultural, and Behavioral*

Modeling, pages 115–125, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-43129-6.

[51] Isaac Johnson, Connor McMahon, Johannes Schöning, and Brent Hecht. The effect of population and "structural" biases on social media-based algorithms: A case study in geolocation inference across the urban-rural spectrum. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 1167–1178, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346559. doi: 10.1145/3025453.3026015. URL https://doi.org/10.1145/3025453.3026015. 3.4.1, 3.4.3

[52] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547, 2019. 3.3.2

[53] Marc Owen Jones. The gulf information war— propaganda, fake news, and fake trends: The weaponization of twitter bots in the gulf crisis. *International journal of communication*, 13:27, 2019. 1.1

[54] Heikki Karjaluoto and Matti Leppäniemi. Social identity for teenagers: Understanding behavioral intention to participate in virtual world environment. *Journal of theoretical and applied electronic commerce research*, 8(1):1–16, 2013. 3.1.3

[55] Catherine King, Daniele Bellutta, and Kathleen M Carley. Lying about lying on social media: a case study of the 2019 canadian elections. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pages 75–85. Springer, 2020. 2.1.1, 4.2

[56] Sneha Kudugunta and Emilio Ferrara. Deep neural networks for bot detection. *Information Sciences*, 467:312–322, 2018. 1.1, 1.2

[57] Adrienne LaFrance. The Internet Is Mostly Bots — theatlantic.com. https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043/, 2017. [Accessed 16-Jul-2023]. 1.2

[58] Michael J Lanham, Geoffrey P Morgan, and Kathleen M Carley. Social network modeling and agent-based simulation in support of crisis de-escalation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(1):103–110, 2013. 2.2

[59] David Lazer, Matthew Baum, Nir Grinberg, Lisa Friedland, Kenneth Joseph, Will Hobbs, and Carolina Mattsson. Combating fake news: An agenda for research and action. 2017. 3.4.3

[60] Summer Lightfoot and Sean Jacobs. Political propaganda spread through social bots. *Media, Culture, & Global Politics*, pages 0–22, 2018. 3.4.1

[61] Darren L Linvill and Patrick L Warren. Troll factories: Manufacturing specialized disinformation on twitter. *Political Communication*, 37(4):447–467, 2020. 1.2

[62] Thomas Magelinski, Lynnette Ng, and Kathleen Carley. A synchronized action framework for detection of coordination on social media. *Journal of Online Trust and Safety*, 1(2), 2022. 1.1, 1.2, 2.1.1, 3.3.2, 3.3.3, 4.2

[63] Bjarke Mønsted, Piotr Sapieżyński, Emilio Ferrara, and Sune Lehmann. Evidence of

complex contagion of information in social media: An experiment using twitter bots. *PloS one*, 12(9):e0184148, 2017. 1.2, 3.5.1

[64] Isabel Murdock, Kathleen M Carley, and Osman Yağan. Identifying cross-platform user relationships in 2020 us election fraud and protest discussions. *Online Social Networks and Media*, 33:100245, 2023. 1.2, 3.3.2

[65] Eni Mustafaraj, Samantha Finn, Carolyn Whitlock, and Panagiotis T Metaxas. Vocal minority versus silent majority: Discovering the opionions of the long tail. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 103–110. IEEE, 2011. 3.5.4

[66] Lynnette Hui Xian Ng and Kathleen M Carley. Do you hear the people sing? comparison of synchronized url and narrative themes in 2020 and 2023 french protests. *Frontiers in Big Data*, 6:1221744.

[67] Lynnette Hui Xian Ng and Kathleen M Carley. Bot-based emotion behavior differences in images during kashmir black day event. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pages 184–194. Springer, 2021. 1.1, 1.2, 3.1.2, 3.1.2, 3.1.3, 3.3.2

[68] Lynnette Hui Xian Ng and Kathleen M Carley. Online coordination: methods and comparative case studies of coordinated groups across four events in the united states. In *Proceedings of the 14th ACM Web Science Conference 2022*, pages 12–21, 2022. 1.2, 3.3.2, 3.3.3, 3.3.4

[69] Lynnette Hui Xian Ng and Kathleen M Carley. Pro or anti? a social influence model of online stance flipping. *IEEE Transactions on Network Science and Engineering*, 10(1): 3–19, 2022. 1.2, 2.2, 3.5.2, 3.5.3, 3.5.4, 4.1, 4.3

[70] Lynnette Hui Xian Ng 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*, volume 17, pages 686–697, 2023. 1.2, 2.2, 2.2, 3.1.2, 3.1.4, 4.1, 4.2, 4.2, 4.3

[71] Lynnette Hui Xian Ng and Kathleen M Carley. A combined synchronization index for evaluating collective action social media. *Applied network science*, 8(1):1, 2023. 2.1.1, 2.1.1, 3.3.2, 4.1, 4.2, 4.3

[72] Lynnette Hui Xian Ng and Kathleen M Carley. Popping the hood on chinese balloons: Examining the discourse between us and china-geotagged accounts. *First Monday*, 2023.

[73] Lynnette Hui Xian Ng and Iain J Cruickshank. Recruitment promotion via twitter: A network-centric approach of analyzing community engagement using social identity. *Digital Government: Research and Practice*, 2023. 3.4.3

[74] Lynnette Hui Xian Ng, Iain J Cruickshank, and Kathleen M Carley. Coordinating narratives framework for cross-platform analysis in the 2021 us capitol riots. *Computational and Mathematical Organization Theory*, pages 1–17, 2022. 3.3.2, 4.1

[75] Lynnette Hui Xian Ng, Iain J Cruickshank, and Kathleen M Carley. Cross-platform information spread during the january 6th capitol riots. *Social Network Analysis and Mining*,

12(1):133, 2022. 1.1, 1.2, 3.1.3, 3.3.1, 3.3.2, 3.4.3, 4.1, 4.2, 4.3

[76] Lynnette Hui Xian Ng, JD Moffitt, and Kathleen M Carley. Coordinated through aweb of images: Analysis of image-based influence operations from china, iran, russia, and venezuela. *arXiv preprint arXiv:2206.03576*, 2022. 3.3.2, 4.1

[77] Lynnette Hui Xian Ng, 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*, 28:100198, 2022. 2.1.1, 4.1, 4.3

[78] Lynnette HX Ng and Araz Taeihagh. How does fake news spread? understanding pathways of disinformation spread through apis. *Policy & Internet*, 13(4):560–585, 2021. 1.2

[79] Mariam Orabi, Djedjiga Mouheb, Zaher Al Aghbari, and Ibrahim Kamel. Detection of bots in social media: a systematic review. *Information Processing & Management*, 57(4): 102250, 2020. 1.2

[80] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. Uncovering coordinated networks on social media: methods and case studies. In *Proceedings of the international AAAI conference on web and social media*, volume 15, pages 455–466, 2021. 1.2, 3.3.2

[81] Susannah BF Paletz, Brooke E Auxier, and Ewa M Golonka. *A multidisciplinary framework of information propagation online*. Springer, 2019. 1.2, 3.2.1

[82] Hao Peng, Roy Schwartz, Dianqi Li, and Noah A Smith. A mixture of h-1 heads is better than h heads. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6566–6577, 2020. 3.1.2

[83] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 813–822, 2016. 1.1, 3.4.1

[84] Denise Scannell, Linda Desens, Marie Guadagno, Yolande Tra, Emily Acker, Kate Sheridan, Margo Rosner, Jennifer Mathieu, and Mike Fulk. Covid-19 vaccine discourse on twitter: A content analysis of persuasion techniques, sentiment and mis/disinformation. *Journal of health communication*, 26(7):443–459, 2021. 1.2

[85] James Schnebly and Shamik Sengupta. Random forest twitter bot classifier. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0506–0512. IEEE, 2019. 1.2

[86] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. The spread of low-credibility content by social bots. *Nature communications*, 9(1):1–9, 2018. 1.1

[87] Zachary C Steinert-Threlkeld, Delia Mocanu, Alessandro Vespignani, and James Fowler. Online social networks and offline protest. *EPJ Data Science*, 4(1):1–9, 2015. 1.2, 3.3.1

[88] Stefan Stieglitz, Florian Brachten, Björn Ross, and Anna-Katharina Jung. Do social bots dream of electric sheep? a categorisation of social media bot accounts. *arXiv preprint arXiv:1710.04044*, 2017. 3.2.1

[89] Zoetanya Sujon. 11. cambridge analytica, facebook, and understanding social media beyond the screen. *Social Media in Higher Education: Case Studies, Reflections and Analysis*, page 11, 2019. 3.2.1

[90] Ye Sun, Lijiang Shen, and Zhongdang Pan. On the behavioral component of the third-person effect. *Communication Research*, 35(2):257–278, 2008. 1.2

[91] Yla R Tausczik and James W Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29 (1):24–54, 2010. 2.2

[92] Milo Trujillo, Sam Rosenblatt, Guillermo de Anda Jáuregui, Emily Moog, Briane Paul V. Samson, Laurent Hébert-Dufresne, and Allison M. Roth. When the echo chamber shatters: Examining the use of community-specific language post-subreddit ban. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pages 164–178, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.woah-1.18. URL https://aclanthology.org/2021.woah-1.18. 2.1.2

[93] Milena Tsvetkova, Ruth García-Gavilanes, Luciano Floridi, and Taha Yasseri. Even good bots fight: The case of wikipedia. *PloS one*, 12(2):e0171774, 2017. 1.1

[94] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185 (4157):1124–1131, 1974. 3.4.1

[95] Joshua Uyheng and Kathleen M Carley. Bot impacts on public sentiment and community structures: Comparative analysis of three elections in the asia-pacific. In *Social, Cultural, and Behavioral Modeling: 13th International Conference, SBP-BRiMS 2020, Washington, DC, USA, October 18–21, 2020, Proceedings 13*, pages 12–22. Springer, 2020. 2.1.1, 4.2

[96] Joshua Uyheng, 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(3):324–342, 2021. 2.1.1, 3.1.3, 4.2

[97] Gerben A Van Kleef, Arik Cheshin, Agneta H Fischer, and Iris K Schneider. The social nature of emotions. *Frontiers in psychology*, 7:896, 2016. 3.1.2

[98] Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer, and Alessandro Flammini. Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the international AAAI conference on web and social media*, volume 11, 2017. 2.2, 3.1.2, 4.2

[99] R Vinayakumar, KP Soman, Prabaharan Poornachandran, Mamoun Alazab, and Alireza Jolfaei. Dbd: Deep learning dga-based botnet detection. *Deep learning applications for cyber security*, pages 127–149, 2019. 1.2

[100] Derek Weber and Lucia Falzon. Temporal nuances of coordination networks. *arXiv e-prints*, pages arXiv–2107, 2021. 1.2, 3.3.3, 3.3.4

[101] Derek Weber and Frank Neumann. Who's in the gang? revealing coordinating communities in social media. In *2020 IEEE/ACM International Conference on Advances in Social*

Networks Analysis and Mining (ASONAM), pages 89–93. IEEE, 2020. 1.2, 3.3.2, 3.3.3

[102] Feng Wei and Uyen Trang Nguyen. Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings. In *2019 First IEEE International conference on trust, privacy and security in intelligent systems and applications (TPS-ISA)*, pages 101–109. IEEE, 2019. 1.2

[103] Kim Witte and Mike Allen. A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health education & behavior*, 27(5):591–615, 2000. 1.2

[104] Stefan Wojcik. Bots in the Twittersphere — pewresearch.org. `https://www.pewresearch.org/internet/2018/04/09/bots-in-the-twittersphere/`, 2018. [Accessed 17-Jul-2023]. 1.2

[105] Samuel C Woolley and Philip N Howard. *Computational propaganda: Political parties, politicians, and political manipulation on social media*. Oxford University Press, 2018. 1.2

[106] Harry Yaojun Yan, Kai-Cheng Yang, James Shanahan, and Filippo Menczer. Exposure to social bots amplifies perceptual biases and regulation propensity. 2022. 1.2

[107] Chao Yang, Robert Chandler Harkreader, and Guofei Gu. Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers. In *Recent Advances in Intrusion Detection: 14th International Symposium, RAID 2011, Menlo Park, CA, USA, September 20-21, 2011. Proceedings 14*, pages 318–337. Springer, 2011. 1.2

[108] Michael Miller Yoder, Qinlan Shen, Yansen Wang, Alex Coda, Yunseok Jang, Yale Song, Kapil Thadani, and Carolyn P Rosé. Phans, stans and cishets: Self-presentation effects on content propagation in tumblr. In *12th ACM Conference on Web Science*, pages 39–48, 2020. 3.1.3

[109] Samira Yousefinaghani, Rozita Dara, Samira Mubareka, Andrew Papadopoulos, and Shayan Sharif. An analysis of covid-19 vaccine sentiments and opinions on twitter. *International Journal of Infectious Diseases*, 108:256–262, 2021. 1.2

[110] Savvas Zannettou, Tristan Caulfield, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. Characterizing the use of images in state-sponsored information warfare operations by russian trolls on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 774–785, 2020. 3.3.2