Social-Cyber Maneuvers for Analyzing Online Influence Operations

Janice T. Blane

CMU-S3D-23-102

May 2023

Software and Societal Systems Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Kathleen M. Carley, Chair
L. Richard Carley
Hirokazu Shirado
David M. Beskow, United States Military Academy

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

This work was supported in part by the Office of Naval Research (ONR) under grants N00014182106 and N000142112229, the Knight Foundation, the Omar N. Bradley Foundation, the United States Army, and the Center for Informed Democracy and Social-cybersecurity (IDeaS). The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, the ONR, the United States Army, or the United States Government.

Keywords: influence operations, social-cyber maneuver, social media, Twitter, BEND framework, CUES, social network analysis, COVID-19, election, Ukraine, Russia

To my dearest family

Abstract

Social media platforms have quickly become a primary news source for many users because of their convenience and easy access to information. The increasing reliance on social media as a news source has created an environment where online influencers can manipulate narratives and social network structures, often leading to large-scale influence campaigns with real-world consequences. Previous frameworks developed to characterize these online influence operations provide only generic guidelines for analysis rather than a comprehensive and quantitative approach for assessing and addressing these campaigns. The BEND framework offers a more comprehensive approach through identification of social-cyber maneuvers, or online methods of influence and manipulation, to accurately characterize and address the operations that support broader influence campaign objectives.

My research goal is to develop a robust framework for identifying the occurrence of social-cyber maneuvers, operationalizing the framework, and creating an influence operations assessment to inform decision-makers and provide possible counter-maneuvers to deter influence operations. This thesis builds on the BEND framework as a comprehensive tool for analyzing influence campaigns and their associated online operations.

To accomplish this, first, I provide refined definitions for the social-cyber maneuvers to improve the descriptions of each maneuver and its application to influence campaigns. Then I refine the metrics for detecting and understanding these maneuvers by employing an iterative process using statistical analysis of real-world data. Afterward, I develop a method for implementing the social-cyber maneuver framework on online social networks and creating an assessment of influence campaigns. By applying this social-cyber maneuver framework on Twitter data sets related to the COVID-19 Vaccine, the 2022 US Elections, and the Russian Invasion of Ukraine 2022, I illustrate the BEND framework efficacy while providing insight into how the maneuvers are used in combination over time to support overarching influence campaign objectives.

Acknowledgments

There were many times during the PhD journey when I questioned whether or not this was something that I really wanted to do. It wasn't just because getting a PhD is hard (because it definitely was one of the hardest things that I have ever done in my life), but it was because, during this time, I felt like I couldn't fail on so many levels. I didn't want to fail as a student, as an Army officer, as a friend, as a wife, and especially, as a mom to my three wonderful children. However, I would not be where I am without the many people who have supported me along the way. So many amazing individuals throughout my career and my time here in Carnegie Mellon have lead me to and through this Phd journey making it the worthwhile and rewarding experience that it has become.

First and foremost, I would like to thank my advisor, Dr. Kathleen Carley for all of her mentorship and guidance in making me a better academic and a more skilled analyst. Without her help and support, my thesis would never have gotten done. Additionally, I would like to thank Drs. Rick Carley, Hiro Shirado, and Dave Beskow for agreeing to serve on my committee and overseeing my progress until the end.

I have been fortunate enough to have been surrounded by the most amazing mentors and officers in the Army and throughout my academic career, and they have given helpful advice, provided ongoing support, and shown genuine care. These include my peers, all of my fellow instructors and senior faculty members at the United States Military Academy, many leaders in the Army, and my thesis advisor, Dr. Kent Choquette, at the University of Illinois in Urbana-Champaign.

At Carnegie Mellon, the CASOS research group and all of my peers in S3D have been a collection of the most caring, intelligent, and fun group of individuals, and I am so appreciative of everything they have done from helping me figure out PhD life to motivating me across the finish line. Thank you, especially to Lynnette Ng, Charity King, Catherine King, Stephen Dipple, J.D. Moffitt, Daniele Bellutta, and Courtney Miller (and the accompanying Chanel) for all the extra that you gave me.

And finally, the most important people in my life - my family. Thank you to my husband for staying up late with me to bounce off ideas, proofreading my papers, and being there to watch the kids whenever he could. His words of encouragement kept me going during the many times I felt like I couldn't go anymore. My parents, brothers, and sisters have always supported me even if it's to just chat about life on the commute or send me funny memes and reels. And last, but not least, my kiddos, Braxton, Remy, and Bradley. Though they could be challenging at times, they are my inspiration and loved me every day during my studies even when I was stressed. I hope they learn they could do hard and amazing things, too.

Contents

1 Introduction and Background 1

1.1 Introduction . . . 1

1.1.1 What Are Online Information Operations? *Or Do You Mean Influence Campaigns?* . . . 2

1.1.2 Analyzing Influence Operations . . . 2

1.2 Background . . . 4

1.2.1 Twitter . . . 5

1.2.2 Social Network Analysis . . . 5

1.3 Online Social Network Analysis Tools . . . 7

1.3.1 Bot Detection . . . 7

1.3.2 Stance Detection . . . 8

1.3.3 Organization Risk Analyzer - PRO Software . . . 8

1.4 Data . . . 9

1.5 Advancing the Methods of Analyzing Online Information Operations . . . 11

2 What are People Doing Online? *BEND Maneuvers* 13

2.1 Introduction . . . 13

2.2 Community Maneuvers . . . 16

2.2.1 Back . . . 16

2.2.2 Build . . . 18

2.2.3 Bridge . . . 21

2.2.4 Boost . . . 22

2.2.5 Neutralize . . . 24

2.2.6 Nuke . . . 26

2.2.7 Narrow . . . 28

2.2.8 Neglect . . . 28

2.3 Narrative Maneuvers . . . 30

2.3.1 Engage . . . 31

2.3.2 Explain . . . 31

2.3.3 Excite . . . 33

2.3.4 Enhance . . . 35

2.3.5 Dismiss . . . 35

2.3.6 Distort . . . 37

2.3.7 Dismay 38
2.3.8 Distract 40
2.4 BEND Tactical Categories: Combinations of BEND Maneuvers for Effects . . . 41
2.4.1 Develop the Narrative: engage, explain, enhance 41
2.4.2 Counter the Narrative: dismiss, distract, distort 42
2.4.3 Emotionally Influence: excite, dismay 42
2.4.4 Grow Groups: build, boost, bridge 43
2.4.5 Reduce Groups: neglect, narrow, nuke 43
2.4.6 Affect Leaders: back, neutralize 43
2.5 Conclusion 44

3 BEND Maneuvers in the Wild.
The Numbers **45**
3.1 Introduction 45
3.2 Humans Detecting BEND Maneuvers 46
3.2.1 Using *Label Studio* for Annotations 46
3.2.2 Interrater Agreement 46
3.2.3 Highly-Trained Annotator Interrater Reliability 48
3.2.4 Conclusions and Limitations Regarding Human Annotation 49
3.3 Cognitive/Emotional CUES 50
3.4 Man vs Machine: Human-Labeled vs Automated Maneuver Detection 51
3.5 Refining Automated Detection 52
3.5.1 Methodology for Refinement 53
3.5.2 Evaluating CUES Weights 54
3.5.3 Evaluating Threshold Values 55
3.5.4 Additional Considerations for Detection 57
3.5.5 Look How Far We've Come 58
3.5.6 Limitations and Future Work for Framing Tweets for Detecting Maneuvers 58
3.5.7 When is it Good Enough? 60
3.6 General Recommendations for Improvement 60
3.7 Developing a Baseline 61
3.8 What Maneuvers are Commonly Seen Together, or Not? 62
3.9 Conclusion 63

4 How Do We Analyze Online Influence?
The BEND Framework **65**
4.1 Introduction 65
4.1.1 What Do We Want to Know? 65
4.1.2 Developing the Social-Cyber Maneuver Analysis Framework 66
4.2 Methodology Overview 66
4.2.1 The Analysis Pipeline 66
4.2.2 Social-Cyber Maneuver Analysis: BEND Framework Workflow 67
4.3 Analysis Focus 69
4.3.1 BEND Maneuvers 69

4.3.2 Key Actors 69
4.3.3 Communities 71
4.3.4 Topic or Narrative 73
4.3.5 Network Measures 74
4.3.6 Overtime/Impact Analysis 75
4.4 Assessment 76
4.4.1 Making BEND Recommendations 76
4.5 An Adaptable Framework 78
4.6 Conclusion 78

5 Tell Me More About This Influence Campaign.
Three Case Studies **81**
5.1 Introduction 81
5.2 COVID-19 Vaccine Rollout Case Study 81
5.2.1 Introduction 81
5.2.2 Related Work 83
5.2.3 Methods 84
5.2.4 Results 88
5.2.5 Discussion 95
5.2.6 Conclusion 97
5.3 2022 US Midterm Election Case Study 98
5.3.1 Introduction 98
5.3.2 Methodology 100
5.3.3 Results 104
5.3.4 Conclusion 114
5.4 Ukraine-Russia Case Study 115
5.4.1 Purpose 115
5.4.2 Background 115
5.4.3 Sources 116
5.4.4 Analysis 117
5.4.5 Overall assessment 122

6 Concluding Remarks **125**
6.1 Summary 125
6.2 Contributions 125
6.3 Limitations 127
6.4 Future Research 128

Bibliography **133**

A Data Set Information **145**
A.1 List of terms for Data Collection 145
A.2 List of Terms for Stance Detection 145

B List of CUES and Descriptions 149

C CUES Usage Correlation by Maneuver 155

D BEND Framework Tutorials 165
D.1 How to Import CUES 165
D.1.1 Select *Import Attributes* function (Figure D.1) 165
D.1.2 Import Attributes Parameters (Figure D.2) 165
D.1.3 Import Successful (Figure D.3) 165
D.2 How to Run a BEND Report 168
D.2.1 Generate Report (Figure D.4) 168
D.2.2 Select BEND & Community Assessment Report (Figure D.5) 168
D.2.3 Select Nodesets (Figure D.6) 168
D.2.4 Agent attributes (Figure D.7) 168
D.2.5 Document CUES (Figure D.8) 169
D.2.6 Agent Analysis (Figure D.9) 170
D.2.7 Document Analysis (Figure D.10) 171
D.2.8 Finish BEND report (Figure D.5) 172
D.3 How to Create Communities for BEND Analysis 173
D.3.1 Option 1: Agent Communities (attribute-based) Figure D.12) 173
D.3.2 Option 2: Agent Communities (topic-oriented) Figure D.13) 174
D.3.3 Community Analysis Figure D.14) 174

E BEND Maneuver Simulation 177
E.1 Introduction 177
E.2 Related Works 178
E.2.1 Modeling Information and Beliefs 178
E.2.2 BEND Framework 178
E.2.3 Modeling Emotions and Reason 179
E.3 Model Description 179
E.3.1 Review of `twitter_sim` Features 179
E.3.2 Model Changes Introduced for `twitter_sim2.0` 180
E.4 Experiments 181
E.4.1 Network Based Maneuvers 181
E.4.2 Information Based Maneuvers 182
E.5 Results and Discussion 183
E.5.1 Baseline 183
E.5.2 Network Maneuver 183
E.5.3 Information Maneuver 184
E.6 Validation 185
E.7 Conclusion 185

List of Figures

1.1 Example meta-network with nodes and derived networks 6

2.1 BEND Maneuver Definitions 15
2.2 Example of *back* maneuver. Support for Raphael Warnock. 17
2.3 Network Impact of *back* maneuver. Overtime, an actor will have more positive connections between other users and a topic. 19
2.4 Example of *build* maneuver. Building a group surrounding the topic of support for President Trump based on *follows* and *retweets*. 20
2.5 Network Impact of *build* maneuver. Overtime, a community is created based on a given topic or attribute. 21
2.6 Network Impact of *bridge* maneuver. Overtime, two communities are joined. . . 23
2.7 Network Impact of *boost* maneuver. Overtime, a community increases in size and/or connections. 24
2.8 Example of *neutralize* maneuver. Decreasing importance and credibility of Dr. Fauci, a COVID-19 policy advisor. 25
2.9 Network Impact of *neutralize* maneuver. Overtime, an actor as related to a topic loses connections to other actors or has more negative connections to actors resulting in a decrease of importance and/or reach. 26
2.10 Example of *nuke* maneuver. Twitter removes accounts and tweets based on policy against spreading COVID vaccine misinformation, dismantling the COVID-19 anti-vaccine community 27
2.11 Network Impact of *nuke* maneuver. Overtime, a community is dismantled. 27
2.12 Network Impact of *narrow* maneuver. Overtime, a community is fissioned into distinct groups. 29
2.13 Network Impact of *neglect* maneuver. Overtime, a community decreases in size and/or density. 30
2.14 Example of *engage* maneuver. Author is making the actions for slowing the spread of COVID-19 of personal importance to their audience, attempting to make them take ownership of the narrative. 32
2.15 Example of *explain* maneuver. Author uses statistics and instructional infograph and discussion to elaborate why he/she does not wear a mask. 33
2.16 Example of *excite* maneuver. Author expresses positivity to encourage users to get the vaccine. 34
2.17 Example of *enhance* maneuver. Author draws on the expertise and importance of former US presidents to encourage audience to get COVID-19 vaccine. 36

2.18 Example of *dismiss* maneuver. Author dismisses discussions of COVID as unimportant and uninteresting. 37
2.19 Example of *distort* maneuver. Author sends disinformation in the form of an altered interpretation of Dolly Parton's image. She does not have Bell's Palsy as the author suggests. 38
2.20 Example of *dismay* maneuver. Author attempts to incite fear about vaccines to their audience. 39
2.21 Example of *distract* maneuver. Replies to the original author are unrelated to the issue of COVID. The replier discusses student loan relief as an alternate topic. . . 41

3.1 Label Studio interface for annotators 47
3.2 Interrater Reliability for BEND maneuvers between low-moderately trained annotators 48
3.3 Interrater Reliability for BEND maneuvers between two highly trained annotators 49
3.4 Comparison of Interrater Reliability for BEND maneuvers between two highly-trained annotators and two low-moderately trained annotators. 50
3.5 Calculations for agreements between low-moderately-trained human annotators and automated detection of BEND maneuvers (NM 94, ORA-PRO 154) 52
3.6 Current method for detecting BEND maneuvers. Indicators are used as inputs for calculating degrees of indication. A message is classified as a maneuver if the DOI of the message is greater than the mean of the DOIs for a maneuver for all the messages within the given data set. 53
3.7 Correlations between *excite* CUES. A higher correlation indicates the possibility that pairs of CUE values should be weighted differently. 54
3.8 Distribution of All Maneuver Degree of Indication Values Across All Datasets . . 56
3.9 Calculations for agreements between the SME and an older and more recent version of ORA and Netmapper. 58
3.10 Calculations for agreements between the low-mod trained annotators and an older and more recent version of ORA and Netmapper. 59
3.11 BEND Maneuver Baselines. Shown are the percentage of the data sets that contain each of the maneuvers based on the use of mean or median degrees of indication values for cutoff threshold values. 62
3.12 BEND Maneuvers Correlations. Darker correlation pairs indicate which maneuvers commonly co-occur within the same message. 63

4.1 Social-Cyber Maneuver Analysis Pipeline. This pipeline includes data preparation using supplemental algorithms and software, key actors identification, and a social-cyber maneuver analysis. 67
4.2 BEND Analysis Workflow 68
4.3 Example Hashtag x Hashtag co-occurence network based on COVID-19 vaccine data during the week of the initial Pfizer/BioNTech rollout; blue is anti-vaccine stance, red is pro-vaccine stance, yellow is unassigned; sized by degree centrality 74
4.4 Diagram for achieving influence from Joint Publication 3-13 [75] 77

5.1 Percentage of bots by stance by time period 89
5.2 Percentage of bots among top 100 influencers by stance, time period, and type of influencer 90
5.3 Social-cyber maneuvers and narratives for top 100 most propagated pro-vaccine tweets. 94
5.4 Social-cyber maneuvers and narratives for top 100 most propagated anti-vaccine tweets. 95
5.5 Methodology for Analyzing Online Community Influence - US Election 2022 . . 100
5.6 Number of key actor types among the top 100 superspreaders for Nov 1-9, 2022 . 105
5.7 Agent x Agent Reciprocal Network for Nov 1-9, 2022. Sized by superspreader. Node color indicates stance. Blue is left-wing; red is right-wing; and gray is unassigned. Top 30 superspreader verified users labeled. 106
5.8 Word Cloud of Top 100 Hashtags by Each Stance. Sized by usage count. Blue is left-wing; red is right-wing. 107
5.9 Over time description of number of users, hashtags, and tweets by usage. 108
5.10 Over time network measures by stance. 109
5.11 BEND maneuver categories usage by stance. Number is based on the summation of the number of individual maneuvers used in each category. 111
5.12 Number of left-wing and right-wing maneuverers and targets for BEND maneuvers affecting leaders. 112
5.13 Ukraine Map 115
5.14 Key Actor BEND maneuvers 119
5.15 Bot BEND maneuvers in *KremlinRussia_E* ego network 122
5.16 Bot BEND maneuvers in *ZelenskyyUa* ego network 123

6.1 Diagram of overarching path to improve BEND detection. Each contributing step improves the detection algorithm or CUES in a ratcheting process. 127

C.1 Correlations between *back* CUES 155
C.2 Correlations between *boost* CUES 156
C.3 Correlations between *bridge* CUES 156
C.4 Correlations between *build* CUES 157
C.5 Correlations between *dismay* CUES 157
C.6 Correlations between *dismiss* CUES 158
C.7 Correlations between *distort* CUES 158
C.8 Correlations between *distract* CUES 159
C.9 Correlations between *engage* CUES 159
C.10 Correlations between *enhance* CUES 160
C.11 Correlations between *excite* CUES 160
C.12 Correlations between *explain* CUES 161
C.13 Correlations between *narrow* CUES 161
C.14 Correlations between *neglect* CUES 162
C.15 Correlations between *neutralize* CUES 162
C.16 Correlations between *nuke* CUES 163

D.1 Select Import Attributes 166
D.2 Import Attributes Parameters 166
D.3 Import Successful 167
D.4 Generate Report 168
D.5 Select BEND & Community Assessment report 169
D.6 Select Nodesets 169
D.7 Select Agent Attributes 170
D.8 Select DocumentCUES 170
D.9 Agent Analysis 171
D.10 Document Analysis 171
D.11 Finish BEND Report 172
D.12 Agent Communities (attribute-based) 173
D.13 Agent Communities (topic-oriented) 174
D.14 Community Analysis 175

List of Tables

1.1 Comparison of informations operations frameworks 3
1.2 Data Sets Used in Thesis 11

3.1 Summary of Statistics of BEND maneuver Degree of Indication values (N=38383667) 56
3.2 Threshold recommendations (Degree of Indication (DOI) values) based on the distribution of DOI values for each maneuver from a large and diverse data set . . 57

4.1 16 BEND Maneuvers. 69

5.1 Keywords used to collect COVID-19 vaccine-related tweets. 85
5.2 The number of users labeled as pro-vaccine and anti-vaccine, along with the number of tweets by users of each stance after running the stance detector. 86
5.3 BEND maneuvers organized into application categories. 87
5.4 Hashtag hijacking: usage count of anti-vaccine–related hashtags by pro-vaccine users. 91
5.5 16 BEND Maneuvers. 104
5.6 Chi-squared test for BEND maneuvers in top BEND categories (df=1, N=496048).110
5.7 Key actor network characteristics 118
5.8 Top Targeted Accounts by Key Actors 120

6.1 Popular social media platform network comparison. Interaction networks refer to networks created by replies, quotes, mentions, and retweets. Community interaction networks refer to networks of users based on specific topics interacting through posts and replies. 129
6.2 Popular social media platform modality comparison 130

A.1 2022 U.S. Midterm Election (Swing States) Collection Terms 146
A.2 2022 U.S. Midterm Election (Democrat/Republican) Agent and Hashtag Stance Detection Terms 147

B.1 CUES indicators descriptions and BEND maneuvers (*Indicators not CUES terms)149

E.1 Docking Lite Comparison of `twitter_sim2.0`, `twitter_sim`, and Construct.178
E.2 Experimental Design for twitter sim2.0 181
E.3 BEND Maneuver Results 183

Chapter 1

Introduction and Background

> "Cyberspace isn't just for geeks. It's for warriors now."
>
> — Deputy Secretary of Defense John J. Hamre [65]

1.1 Introduction

Everyone is on social media. And they're all leading fantastic lives (or so it seems). Twitter, Facebook, Reddit, Instagram, TikTok–on these online platforms, users are connected to expansive, global social networks of government organizations, businesses, schools, news agencies, politicians, celebrities, parents, children, automated accounts, and countless other online profiles. These diverse congregations of users gather to share their thoughts about any topic including their kids' football games, politics, apple pie recipes, stock tips, and, of course, cat videos. Public organizations inform the masses about safety, health, or other vital topics. Businesses sell their ideas as much as they sell their products. This expansive network includes messages from 368 million Twitter users (2022), 2.96 billion Facebook users (2022), and millions of others on various other social media platforms [122]. Shared messages come in many forms (posts, texts, images, memes, URLs, music, videos, etc.). Users share multitudes of messages through progressively more entertaining mediums that spread across the globe at the speed of light. Included in all the "noise" created by the speed and volume of these messages is information–and sometimes disinformation–that reaches other users. More interesting than the source and information itself (the "who" and "what") is the intended audience ("to whom"), the motivation behind the sharing ("why"), and how this information influences behavior, which in some cases manifests into "real-world" action ("with what impact"). But even more important than answering these questions: "What are we doing about it?"

Though many online posts appear as innocuous bits of facts and entertainment, all are motive-driven. This is because online social networks have become a perfect medium for actors to spread influence (through a social media "user" account). Underlying motivation for online influence is often predictable, such as campaigning for elections or persuading consumers to purchase a particular product. However, some actors use social media more surreptitiously, spreading disinformation to cause division and distrust [64] to achieve their objectives. Simply connecting to the network and being exposed to news feeds filled with posts and messages creates an environment where members of these online communities are vulnerable to changes in their beliefs, ideas, and behaviors from those they interact with on social media. From individual messages to large-scale online information operations, millions of people's decisions and behaviors are shaped daily. They may decide to change their minds about getting the COVID-19 vaccine, who to vote for governor, or which side of the Russia-Ukraine conflict they support. Due to the increasingly immersive nature of social media, initial changes may be subtle but with drastic results over time.

1.1.1 What Are Online Information Operations? *Or Do You Mean Influence Campaigns?*

Influence operations, influence campaigns, information operations. These phrases or terms are often spouted by officials, academics, and industry alike when describing the spread of information online. Organizations may choose one phrase over another or have established precise definitions for these operations. For example, the U.S. military's Joint Publication 3-13 explicitly defines influence operations "as the integrated military operations, of Information Related Capabilities (IRCs) in concert with other lines of operation to influence, disrupt, corrupt, or usurp the decision making of adversaries and potential adversaries while protecting our own. " [75] Simplified, these different terms are used to refer to the same idea, which is *a collective effort by actors using influence methods towards a target audience to achieve a desired end state*. This thesis will use these terms interchangeably as they refer to this general definition.

A key development in the fight against online influence campaigns has been the growth of the field of social cybersecurity. Social cybersecurity is a computational social science that aims to protect the security of democratic societies by studying how actors exercise manipulation on social media platforms [6]. Recognized by the National Academies as a new science [20], key areas of social cybersecurity research include the study of information maneuvers, motive identification, and information diffusion, as well as the evaluation of the effectiveness of information campaigns and mitigation strategies.

As online platforms evolved into prominent places for gathering and quickly communicating with large audiences, in turn, they became mediums for influence and manipulation.

1.1.2 Analyzing Influence Operations

To better understand, visualize, and appropriately address influence operations, agencies such as the US military, NATO, academia, and other public and private organizations have attempted to develop their own framework and guides.

The Department of Defense (DOD) and the U.S. military have produced several publications on information operations and online behavior. Joint Publication 3-13: Information Operations [75] provides the doctrine for information operations across military operations. This includes discussion regarding information operations efforts as related to areas such as public affairs, cyberspace operations, Military Information Support Operations (MISO), and intelligence. This publication describes the information environment as having a physical, informational, and cognitive dimension. Then by applying information-related capabilities upon these different dimen-

Table 1.1: Comparison of informations operations frameworks

Characteristics	4 D's	ABC(D)(E)	SCOTCH	DISARM	BEND
General high-level approach	x	x	x	x	x
Specific systematic methodology					x
Key Actor Analysis		x	x		x
Behavior Analysis	x	x		x	x
Quantitative Analysis					x
Limited to Disinformation Analysis	x	x		x	
Influence campaign analysis/assessment			x		x
Develops recommendations		x		x	x
References	[98]	[3, 57, 101]	[22]	[50]	[17, 38]

sions to a target audience, this application can lead to influence and, ultimately, a desired end state. The joint doctrine also outlines online behavior for public affairs activities. In Joint Publication 3-61: Public Affairs [76], the doctrine lists 'Social Media Tenets' and considerations for social media management and social media planning. Most recently in January 2023, the DOD Instruction 5400.17 [99] was published with changes on the "Official Use of Social Media for Public Affairs." Social media is a tool that supports the military public affairs mission as it improves military presence, improves communications and building reach, and is a medium for countering adversarial propaganda.

DOD policies and Joint doctrine provide general guidelines for service members to operate. However, specific steps and details of conducting analysis, assessments, and responses to online information are typically delegated down to subordinate units in the separate services and outlined in their standard operating procedures (SOPs), assuming that units have created them.

Other frameworks conduct analysis through broad analysis using acronyms to generally describe disinformation campaigns. In 2015, Ben Nimmo, Director of Investigations for network analysis and a former NATO press officer, introduced the 4D Model of Disinformation Campaigns [98]. The four D's are dismiss, distort, distract, and dismay. Dismiss is used to counter negative messaging by denying the allegations or denigrating the one making the comments. Distort is twisting or altering information to support a narrative. Distract is to turn attention away from a topic and direct focus elsewhere. Dismay is to scare or threaten opponents. Nimmo developed this model while evaluating Russian propaganda tactics and methods of weaponizing the narrative online. This method has since been proven adaptable for analyzing disinformation in other regions of the world.

Camille Francois developed a separate ABC framework to highlight three vectors of deception and online disinformation as a regulatory guide for industry responses [57]. 'A' refers to the manipulative actors; 'B' refers to the deceptive behaviors or the tactics and techniques for manipulation; and 'C' refers to the harmful content. This method was developed to combine the three vectors, often analyzed in isolation depending on the discipline or stakeholder addressing the problem. Later, Alexandre Alaphilippe extended it to ABCD to add the Distribution of content, and how the disinformation spreads across the online platforms [3]. In its most recent branch, Pamment [101] added E to the framework to represent the the Effect of the disinformation.

In 2021, Blazek [22] introduced a framework for quickly assessing influence operations using the acronym SCOTCH. S is the Source (or actors) conducting the operations, C is the Channel (or platform or forum used), O is the Objective, T is the target of the operation, C is the composition of the content or image source, and H is the Hook or tactics for persuasion and diffusion. The characterizations produced from this framework are succinct reports intended to support decision-makers not requiring in-depth analysis. These models are all useful in that they capture some of the major points of analyzing influence operations, but they lack quantitatively based results.

The DISARM Foundation developed the general-purpose DISARM framework in 2019 to describe disinformation incidents and the responses to those incidents [50]. The framework focuses on a set of tactics, techniques, and procedures (TTPs) and stems from MITRE's ATT&CK and Adversarial Misinformation [88] and Influence Tactics and Techniques (AMITT) [43] frameworks. It consists of a repository organized into individual adversarial listings of TTPs with short summaries. An analyst browses the TTP categories to determine the appropriate descriptions and actions based on the summary and recommendations of the specific tactic. This system, however, does not provide a method for combining these tactics as an overarching assessment of the campaign.

In my thesis, I expand on the 16 BEND maneuvers conceived by Beskow and Carley [17] at Carnegie Mellon University and build on the conceptual development of an analysis framework [35]. The framework, named after the first letters of each of the maneuvers, frames the analysis around BEND. These maneuvers, or more accurately termed - **social-cyber maneuvers**, are actions or discussions actors within the cyber domain use to manipulate the narrative and social media network (social, conversational, and information) to manipulate beliefs and ideas to achieve an operational end state. The collection of maneuvers is an expansion of Nimmo's 4 D's framework and the analysis process contains many of the same aspects as others in identifying who is doing what to whom and with what impact.

However, the current state of the BEND framework and these other frameworks are limited as shown in Table 1.1. Some are too broad to provide a sometimes necessary comprehensive report or too specific that they do not provide an adequate overview of the entire operations. All of the frameworks lack a solid methodology for analysts to follow, and none use quantitative measures as a fundamental element for analysis. This work demonstrates a framework with a method of implementation to create a comprehensive and computationally supported analysis and assessment. This upgraded version of the BEND framework is a viable process for analyzing influence operations backed by robust metrics and illustrated assessments of notable events.

1.2 Background

This section summarizes the key concepts and frequently used terminology in the thesis to provide context for understanding the remaining chapters and their research value.

1.2.1 Twitter

Many of the examples used in this thesis are from data and behaviors observed on Twitter. Twitter is a social networking website where users post messages, or tweets, on their profile page in 280 characters or less to their followers and others across the platform. These tweets can also include URLs, videos, and images to enhance the post and are accessed either through a follower's news feed or through direct access to the post from a URL or link. The news feeds are generated for each account profile based on the platform's algorithm which caters posts to a user's preferences.

Users can interact with other users using the following functions [130]:

- **Retweet** - Repost another user's tweet
- **Quote** - Retweet another user's tweet with a comment
- **Reply** - Reply directly to a user's tweet on their profile
- **Mention** - Connect a message to the user by referencing them in a tweet (use of '@' symbol followed by username)
- **Like** - Click the heart icon to indicate liking a tweet
- **Follow/Unfollow** - Befriend, unfriend

Another feature used on Twitter is the use of hashtags. These are words or phrases (with no spaces) preceded by the '#' symbol used to index or link all tweets that include it. Hashtags are helpful in creating searchable terms that tie messages together of similar topics, themes, or concepts. Additionally, they can indicate the theme of a message or something interesting about the message. For example, searching *#foodporn* within the platform may bring tweets that used the hashtags within their posts related to amazing photos of decadent meals or beautifully presented food.

Twitter also offers a verified status for user profiles represented by a blue checkmark [131]. This means that an account is active, notable, and an authentic account of public interest. Since the collection of the data in this thesis, Twitter has adjusted the criteria for verified users to include users who have an active subscription to the Twitter Blue subscription, or who were previously verified under the legacy verification criteria. The verified users in this work are based on the legacy verification.

Users can access Twitter through the website or on an application using their mobile devices. Developers, however, can access tweets using the Twitter Application Programming Interface (API), though the rate limits of the service result in only a sampling of the entire network.

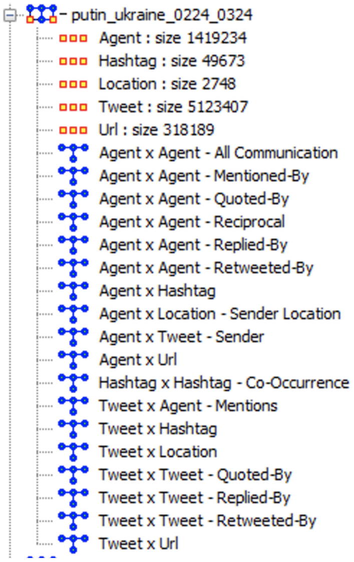
1.2.2 Social Network Analysis

The Twitter data, which comes in the form of JSON files, can be restructured as social networks, multivariate networks consisting of actors and information. The social networks used in this thesis are dynamic meta-networks. These meta-networks are the compilation of multiple networks with various types of nodes observed over time.

The meta-network has five types of nodes: agent, hashtag, tweet, location, and URL. Agent nodes represent the network's actors, users, accounts, or message authors. These agents have attributes that include the Node ID (the Twitter ID which uniquely identifies the users in the network), the Node Label (username), language, the number of followers, whether they are verified, and others. Tweet nodes are messages or posts tweeted by an author. This encompasses all types of tweets, including original tweets, retweets, quotes, and replies. Location nodes are the locations tweets were tweeted, and URL nodes are the URLs that appear in the tweets.

From the social networks exist subordinate networks that can be derived by selecting two node types and a rule that links them together. An agent-to-agent all-communication network is a bi-directional network where two agents are connected if they interact. This interaction is either a retweet, reply, quote, or mention; the weight of these links, or edges, is the number of interactions. If Agent B mentions Agent A, then in this network, the directional link points from Agent A to Agent B. The same concept applies to a simpler network, such as a retweet-by network. If Agent A is retweeted by Agent B, then the network link points from Agent A to Agent B. Other agent-to-agent networks include the quoted-by, replied-by, mentioned-by, and reciprocal networks. A reciprocal network is formed when two agents engage in two-way communication.

These derived networks also exist for other combinations of nodes. A hashtag co-occurrence network is a network of hashtags connected if they occur in the same tweet. An agent-to-tweet sender network ties a tweet to the agent that sent the tweet. A tweet-to-tweet quoted-by network is where quoted tweets are connected to the tweet it was quoted from. Many other types of networks can be derived from the meta-network. Any of these networks used in this thesis will be explained in the section they are analyzed. Figure 1.1 shows an example of the nodes and networks in one of the data sets.



- putin_ukraine_0224_0324
Agent : size 1419234
Hashtag : size 49673
Location : size 2748
Tweet : size 5123407
Url : size 318189
Agent x Agent - All Communication
Agent x Agent - Mentioned-By
Agent x Agent - Quoted-By
Agent x Agent - Reciprocal
Agent x Agent - Replied-By
Agent x Agent - Retweeted-By
Agent x Hashtag
Agent x Location - Sender Location
Agent x Tweet - Sender
Agent x Url
Hashtag x Hashtag - Co-Occurrence
Tweet x Agent - Mentions
Tweet x Hashtag
Tweet x Location
Tweet x Tweet - Quoted-By
Tweet x Tweet - Replied-By
Tweet x Tweet - Retweeted-By
Tweet x Url

Figure 1.1: Example meta-network with nodes and derived networks

Another relevant structure within the social network is the existence of communities. For social media platforms like Facebook and Reddit, communities exist as defined groups. However, Twitter has no formal groups, so identifying communities requires data processing. One type of group can be formed by combining users with similar attributes. A group of bots, or automated accounts, verified users, or news agencies can be communities in a data set. Agents can also be grouped by their locations, languages, or stance.

The term topic has a range of meanings in this thesis. They can be a **concept**, which represents a single ideation. For example, the terms ‘vaccines’ or ‘dog’ can be considered concepts, which are the topics that hashtags typically represent. More elaborately, topics can represent more focused ideas, themes, or overarching narratives. These are more complex versions of a concept, such as ‘get your vaccine’ or ‘dogs are cute’.

Once the topics are determined within the data set, topic-oriented communities can be detected. This can be achieved using clustering algorithms that calculate the groups based on user interaction and the various topics or concepts within the messages. An example method is to use the Leiden clustering algorithm [126] on the all-communication network and the agent-to-hashtag network, where hashtags represent the concepts.

1.3 Online Social Network Analysis Tools

This thesis uses several tools to develop the social-cyber maneuver framework and to create assessments on the various case studies. These include machine learning algorithms for deriving information from the data sets and commercial tools for conducting network calculations and visualizations.

1.3.1 Bot Detection

Of particular interest in social cybersecurity is the employment of bots on social media platforms. Bots are autonomous accounts that interact with users by tweeting, retweeting, or other platform actions. Bots can be valuable tools for disseminating pertinent information, such as those for safety and emergencies, to a large audience, but they can also be the sources of widespread malicious information, particularly for polarizing events [54]. Ill-intended actors have used bots to promote the vast and rapid spread of online disinformation to influence elections [54, 55]. Other actors have employed bots to manipulate public health discourse by propagating misinformation on topics such as e-cigarettes, diets, and medications [4]

Because of the influence of these bots on public opinion, several studies have been conducted on the use of bots for spreading vaccine information [144]. Before the COVID-19 pandemic, Broniatowski et al. [32] examined how bots spread anti-vaccine messages, showing the high rates of vaccine content they spread and comparing it with the effects of Russian trolls, whose messages primarily sought to increase discord online. Dyer [52] determined that after Russian trolls, bots were the most prolific vaccine-related tweeters. Huang and Carley [70] found that accounts linking to coronavirus information from less reliable sites were likelier to be bots. Hence, understanding the actions of automated accounts is a crucial part of online influence campaigns.

This work utilizes two bot detection platforms. The first is the Tier-1 BotHunter algorithm by Beskow and Carley [16], a random forest regression model trained on labeled Twitter data sets, that determines the probability a user is a bot. This machine learning model considers network-level features, user-level attributes, and tweet-level features in its calculations. Users with a bot score greater than 75% are labeled as a bot to reduce the chance of false positives [97]. The second bot detection algorithm used in this thesis is BotBuster by Ng and Carley [96]. The

BotBuster algorithm uses a Mixture of Experts method for bot detection. Each expert is trained to analyze a portion of account information, e.g., username, and then are combined to estimate the bot probability for an account. The experts are tuned to the types of data available from a single or multiple platforms and are constructed with neural network architectures corresponding to the data type.

1.3.2 Stance Detection

An important part of analyzing influence operations is identifying those on opposite sides of issues or differentiating the potential influencers from the targets by separating them by stances. Two stance detection algorithms are used in this thesis. The first by Kumar [80] uses a method of influence propagation through the all-communication network to detect the stances for agents and hashtags. A user labels a set of hashtags as pro or con for an issue, and the algorithm labels the rest of the hashtags and all of the agents. The agents and hashtags are separated into pro, con, and unassigned categories and given a confidence level. Sometimes it is necessary to repeat the process with more or less labeled hashtags to improve the confidence level.

The second stance algorithm, the Twitter Stance Propagation Algorithm (TSAP), by [140] is used only to identify agent stances. The algorithm propagates stance labels through user-similarity-weighted interaction networks [140]. The process is similar to the first method in that it takes a pre-labeled set of hashtags to seed the algorithm for propagation. The results also include a classification for pro, con, and unassigned and a confidence level for each category.

1.3.3 Organization Risk Analyzer - PRO Software

The Organization Risk Analyzer-PRO (ORA-PRO) software [39] is a dynamic meta-network analysis tool used extensively in this work to examine, characterize, and visualize online social network data over time.

ORA-PRO provides several useful functions. The most prominent feature of ORA-PRO is the ability to import data sets and visualize them as social networks. The software attaches attributes to the nodes, derives appropriate networks and meta-networks, and calculates network measures that facilitate the creation of informative social network graphs. The software also houses the stance detection algorithms used in this work, clustering algorithms, and tools for extracting the topic-oriented communities.

ORA-PRO also offers many reports for characterizing the data sets and providing network analyses. The Twitter analysis report calculates relevant statistics about the users, tweets, hashtags, and other content within the data set. The Twitter report also calculates information about superspreaders and superfriends, which allows an understanding of the influential users. Superspreaders are users that post messages that are spread often, indicating an account that efficiently disseminates information throughout the network. Superfriends are users that engage in numerous two-way communications or reciprocity with their interactions. Another report is the BEND and Community Assessments report. This report calculates information about the BEND maneuvers discussed in this thesis for the actors and their tweets. As part of the pipeline for using the BEND report, the NetMapper software [39], a text-mining tool, extracts the linguistic indicators, or CUES, for the BEND maneuvers from the messages within our data set. Then using ORA-PRO, the CUES are added as attributes. The BEND report function calculates and detects the occurrence of various BEND maneuvers within the messages and identifies the influencers using them and their targets. The report also has a function that creates topic-oriented communities or communities based on particular attributes and analyzes their maneuvers from the community perspective. Other reports include a key entities report for key nodes based on network position, a topic analysis report for detecting topics clusters, and an all-measures report, which calculates large amounts of standard network measures.

This thesis's contributions improve the ORA-PRO software's BEND detection function.

1.4 Data

This thesis uses the following eight data sets centered around notable events between 2020 and 2022. All of the data sets, except for the OMEN exercise data set, are used in Chapter 3 for conducting computational analysis on the BEND maneuvers and improving detection. The OMEN data is a semi-synthetic data set used in Chapter 4 for improving the methods for investigating online information operations. To illustrate the framework and understand the complex use of the maneuvers in Chapter 5, I used the COVID-19 vaccine rollout, 2022 US midterm elections, and 2022 Ukraine-Russia Conflict data as case studies.

The data was collected using the Twitter API.

Black Panther movie release

The Black Panther data set is a collection of tweets from February 8 to March 16, 2018, related to the opening weekend for Marvel's Black Panther movie, the first Marvel Cinematic Universe movie to have an African-American director and a predominately African and African-American cast. The data was initially collected to analyze disinformation campaigns on Twitter related to the movie [10].

Captain Marvel movie release

The Captain Marvel dataset is a collection of tweets related to Marvel's Captain Marvel from February 14 to March 15, 2019. This movie featured Marvel's first female-lead superhero movie and became a topic of misinformation and contention on Twitter. Users pushed narratives to boycott the movie based on claims about the lead actress and Marvel. The data was originally collected to compare misinformation campaigns that emerged due to the contention discussion. [11]

French attack in Nice 2020

This data set was collected for an event surrounding a terrorist act, where an Islamic extremist stabbed and killed three people in a Roman Catholic church in Nice, France on October 29, 2020. Within the collection for this event, a second event emerged as a major topic of discourse. A history teacher, Samuel Paty, was accused of showing his class a Charlie Hebdo cartoon depicting a controversial image of the Islamic Prophet Muhammad. On October 16, 2020, he was beheaded by an Islamic terrorist in Paris, France.

COVID-19 vaccine

After a long year of COVID-19 and related restrictions, Pfizer/BioNTech rolled out a vaccine that many hoped return the world back to normal. This data set is a collection of tweets related to the rollout. The data was collected using keywords related to COVID-19 that were then filtered using the vaccine-related terms. The data is separated into 3 time periods surrounding the vaccine's introduction. December 1-7, 2020 was the week before the rollout; December 8-10, 2020 was during the week of the rollout in the United States and the United Kingdom; and January 25-31, 2021 occurs 6 weeks after the rollout [21].

US Election 2020

The 2020 US Election Twitter data was collected using key terms and accounts relevant to the US Election in 2020. The dates range from October 31, 2020, to November 7, 2020. With the 2020 election taking place on November 3, 2020, the selected date range encompasses the set of days before the election (October 31 to November 2, 2020), election day itself (November 3, 2020), and the subsequent days after the election (November 4-7, 2020), when the delay in election results began to prompt discussions of a stolen election and voter fraud.

US midterm election 2022

The US Midterm Election 2022 data set collection focuses on the conversations surrounding the candidates for the Senate, House of Representatives, and governor seats in seven swing states. The swing states for this collection included Arizona, Georgia, Nevada, North Carolina, Ohio, Pennsylvania, and Wisconsin. The dates begin in the week leading up to election day, which fell on November 8, 2022, and extend through November 9.

Ukraine-Russia conflict 2022

This data set was collected from conversations related to President Volodymyr Zelensky of Ukraine and President Vladimir Putin of Russia during the Russian invasion of Ukraine in January 2022. The data was collected separately for each key actor to make comparisons between their behaviors and impacts on social media during the time.

OMEN exercise (semi-synthetic)

The purpose of the OMEN exercise is to train analysts in analyzing social networks using social network analysis techniques and tools. Part of this exercise is the use of data collected in support of the NATO exercise Trident Juncture in 2018. The data was altered into a form of semi-synthetic data to occur within a different time period, with different actors, and infused with additional events. The exercise occurs over the course of several days, and analysts choose specific courses of action to address different issues within the scenario. The data set may appear different to the playing analysts depending on which scenario they select. These courses of action incorporate various applications of the BEND maneuvers used within this thesis.

Table 1.2: Data Sets Used in Thesis

Dataset Name	Dates Collected	Num Tweets
Black Panther Movie Release	8 Feb to 16 Mar 2018	6,177,644
Captain Marvel Movie Release	14 Feb to 15 Mar 2019	5,592,170
Nice, France Terrorist Attack 2020	15 Oct 2020 to 9 Nov 2020	643,956
COVID-19 Vaccine Rollout	1-10 Dec 2020 and 25-31 Jan 2021	4,375,917
US Elections 2020	31 Oct to 12 Nov 2020	11,632,194
US Midterm Elections 2022	1-9 Nov 2022	103,156
Ukraine-Russia Conflict 2022	24 Feb to 24 Mar 2022	9,757,443
OMEN Exercise (semi-synthetic)	5 days	56,169
	Total	38,338,649

1.5 Advancing the Methods of Analyzing Online Information Operations

This thesis aims to characterize, develop, and apply a social-cyber maneuver framework to improve methods for detecting and countering influence operations. While the previously mentioned frameworks address influence operations, they only provide generic guidance for analyzing these campaigns. I evolve the BEND framework into a quantitative and methodological approach for creating a comprehensive analysis and assessment of online influence. The ultimate goal of this framework is to derive actionable information based on the detected actions conducted by online actors and assist leaders, including corporate executives, government and military officials, or even directors of health organizations, in making decisions about social media.

In Chapter 2, I expand the definitions of the individual maneuvers to create a comprehensive understanding of the maneuvers beyond the original definitions to gain an understanding of the center of the BEND framework. In Chapter 3, I refine the methods for detection using an iterative process of computational analysis and provide an overview of how the maneuvers behave. Also, in this chapter, I compare how humans detect the maneuvers versus the results from the ORA-PRO software. Then I explain the framework methodology developed for analyzing and assessing influence operations in Chapter 4. Finally, in Chapter 5, I illustrate the application of the BEND framework focusing on the complex use of the social-cyber maneuvers in combination and over time using case studies related to the COVID-19 vaccine rollout, the 2022 US midterm elections, and the Russian invasion of Ukraine in 2022.

Chapter 2

What are People Doing Online? *BEND Maneuvers*

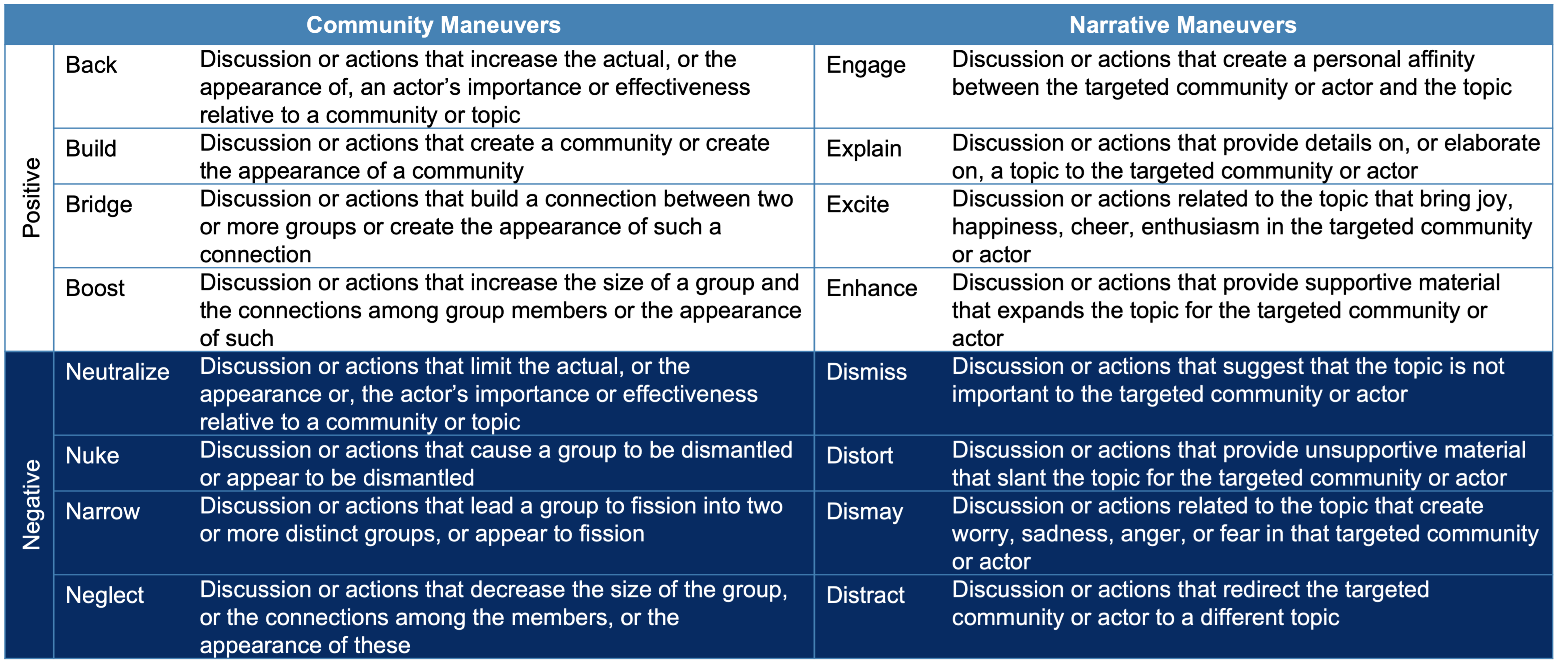
2.1 Introduction

The BEND maneuvers are social-cyber maneuvers where actors attempt to manipulate the narratives and online social networks to affect ideas and behaviors or achieve some other desired end state [17]. These maneuvers are building blocks of how influence campaigns operate and are the crux of the BEND framework [35], an online information operations analysis framework for understanding who is doing what to whom with what impact.

Influencers use these maneuvers through the messages they send on social media. This can be through the content of the message or how they use social platforms to engage with other users online, such as with mentions or replies. A message may contain one or more maneuvers or none at all. There are also some maneuvers that may not appear to be occurring unless placed in the context of other messages, such as when countering or replying to specific topics. In general, humans are unreliable detectors when it comes to identifying maneuvers. However, automated tools are useful in assisting with that detection [39]. Messages may contain CUES, described in detail in Chapter 3.3, that subconsciously elicit cognitive or emotional responses in their targets sometimes resulting in behavioral changes. These CUES can be seemingly innocuous, like exclamation points or capital letters, or unassuming such as terms that trigger happiness or anger. Automated detection has the ability to help overcome this human inability to detect the different indicators while analyzing millions of tweets within a reasonable amount of time.

There are 16 social-cyber maneuvers, the names of which create the BEND acronym. These maneuvers are divided into community and network maneuvers with a positive or negative aspect depending on if the maneuvers are focused on creating growth/increase or reduction/decrease (real or perceived) in a metric for a network or narrative.

In this chapter, I refine the definition for each of the maneuvers and provide a detailed description. The purpose is to improve clarity and understanding of the individual definitions, ultimately leading to better methods for detecting the maneuvers online and characterizing influence operations. The list of the refined abbreviated definitions of the maneuvers is shown in Figure 2.1, and throughout this chapter, each maneuver is described in greater detail to understand the definitions and how they are typically used in social media. They are then complemented with some of the indicators for identifying the maneuvers and illustrative potential impacts. Furthermore, while these are foundations for online influence, it is important to understand that many of these maneuvers harmlessly appear in everyday conversations. That makes it essential to discern how these maneuvers are used in the context of a campaign and how they are used in conjunction with other maneuvers to achieve a particular end state. Therefore, at the end of this chapter, I have also included a section on BEND tactical categories to discuss the use of the BEND maneuvers in combination.



	Community Maneuvers		Narrative Maneuvers	
Positive	Back	Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	Engage	Discussion or actions that create a personal affinity between the targeted community or actor and the topic
	Build	Discussion or actions that create a community or create the appearance of a community	Explain	Discussion or actions that provide details on, or elaborate on, a topic to the targeted community or actor
	Bridge	Discussion or actions that build a connection between two or more groups or create the appearance of such a connection	Excite	Discussion or actions related to the topic that bring joy, happiness, cheer, enthusiasm in the targeted community or actor
	Boost	Discussion or actions that increase the size of a group and the connections among group members or the appearance of such	Enhance	Discussion or actions that provide supportive material that expands the topic for the targeted community or actor
Negative	Neutralize	Discussion or actions that limit the actual, or the appearance or, the actor's importance or effectiveness relative to a community or topic	Dismiss	Discussion or actions that suggest that the topic is not important to the targeted community or actor
	Nuke	Discussion or actions that cause a group to be dismantled or appear to be dismantled	Distort	Discussion or actions that provide unsupportive material that slant the topic for the targeted community or actor
	Narrow	Discussion or actions that lead a group to fission into two or more distinct groups, or appear to fission	Dismay	Discussion or actions related to the topic that create worry, sadness, anger, or fear in that targeted community or actor
	Neglect	Discussion or actions that decrease the size of the group, or the connections among the members, or the appearance of these	Distract	Discussion or actions that redirect the targeted community or actor to a different topic

Figure 2.1: BEND Maneuver Definitions

2.2 Community Maneuvers

Where a social network is a connection of ideas, topics, and agents, or users, community maneuvers, sometimes referred to as network maneuvers, are discussions or actions taken to alter the connections within that network. The resulting impacts are structural changes in the relationships between actors and their discussion topics, which in turn, affects how information flows throughout the network.

Networks of communities are created, grow, evolve, shrink, or cease to exist as a result of the different actions of actors on social media. Different groups can also join together or become marginalized or broken into disjointed sub-communities. These communities come in varying sizes and are typically born out of their shared ideologies (e.g., flat-earthers), identities (e.g., bots), or topics of discussion (e.g., stuff to do in Pittsburgh). These groups of agents formed through shared discussion topics are also known as topic-oriented communities - clusters of agents who discuss the same topic at approximately the same time.

Network structure changes can also mean the changing support for opinion leaders as people connect or disconnect with these leaders. Agents can become more popular or more disliked in a network. This is observed by examining an agent's position and how central they are to positive or negative topics and the rest of the agents within the network.

The community maneuvers are divided into how they affect the network structure. The positive community maneuvers are the “B” maneuvers: back, build, bridge, and boost. These typically aim to create more connections and increase the reach of a message through a larger community and interaction network. The negative community maneuvers are the “N” maneuvers: neutralize, narrow, nuke, and neglect. When targeting specific actors or communities, these maneuvers alter the social network structure to impede the spread of narratives and beliefs and their subsequent impact. The “B” and “N” maneuvers' are in many ways complementary, though it is not uncommon to see both of these maneuvers in the same messages.

The implementation and impact of a community maneuver are not always apparent from the purview of a single message. For example, some indicators like mentions or conversations about a community may hint that a positive community maneuver is occurring, but deciding which of these individual maneuvers can be harder to deduce. Furthermore, just because a maneuver has a desired impact, say nuke, the resulting impact may be another maneuver, such as narrow.

Finally, community maneuvers affect the social network structure in ways that facilitates the movement of narrative maneuvers during a campaign. A perfectly crafted message with no one to receive it is not very influential. A good narrative is most effective with well-positioned leaders to convey the information to the masses and a network of actors situated to extend the reach to their target audience.

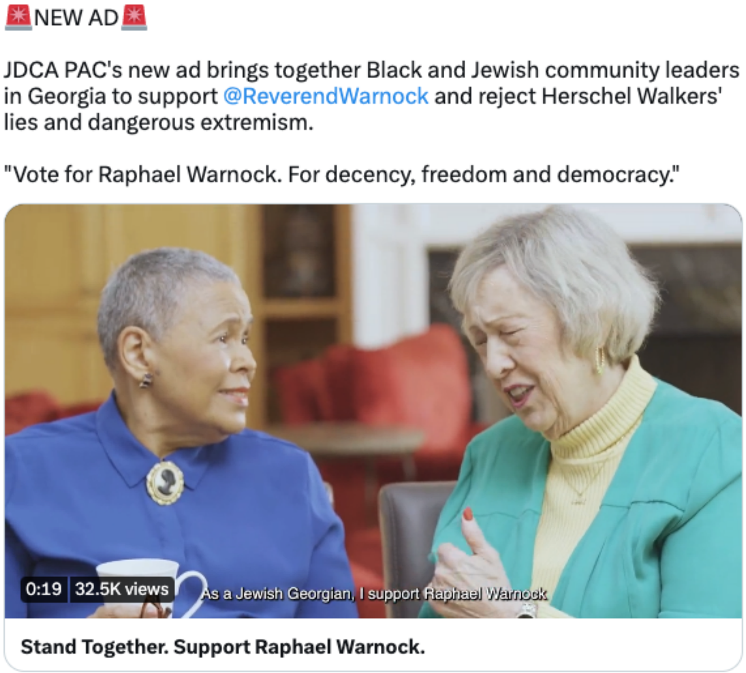
2.2.1 Back

> *Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic*

Description

The purpose of a back maneuver is to increase support for an actor to develop as an opinion leader within the network. An individual then becomes increasingly interconnected with positive messages and interacts with other actors resulting in an extended reach of the distribution of their beliefs, ideas, and ultimately, their influence.

Back maneuvers can be used to support causes or campaigns as shown in Figure 2.2. A politician or leader who requires a large following to push a particular agenda. The supporters would turn around and spread their messages and encourage more followers of this leader. The same follows for organizations or non-human entities. The World Health Organization (WHO), for example, has several campaigns for public health from immunizations to mental health. Their campaigns rest on a large following who support their cause, validate their importance as a reputable institution, and spread these messages to cascades of otherwise unreachable actors.



🚨NEW AD🚨

JDCA PAC's new ad brings together Black and Jewish community leaders in Georgia to support @ReverendWarnock and reject Herschel Walkers' lies and dangerous extremism.

"Vote for Raphael Warnock. For decency, freedom and democracy."

Stand Together. Support Raphael Warnock.

Figure 2.2: Example of *back* maneuver. Support for Raphael Warnock.

Indications

To identify a backing maneuver, we first identify the network structure of relationships between an actor and a topic of interest. In a multimodal network, an actor is connected to a topic. This may be a topic that they have an opinion about or is an idea that is commonly associated with this particular actor. As more people connect to this actor-topic relationship with positive sentiment, they become a leader or opinion leader within the network.

Mentioning other users is a method for building an opinion leader's ego network. This includes the use of the opinion leader's name in the body of the message, mentions of the opinion leader, mentions to other supporters, or mentions to potential supporters. When usernames are mentioned in a post, typically the user is notified of the mention and the post. The mentions take advantage of Twitter functions to draw people to specific messages and tie them to conversations that they may not otherwise have joined.

Since a backing maneuver is used to improve a leader's importance and image, indications for this maneuver may be positive terms, emojis, or words that indicate loyalty and encouragement. The use of these types of cues increases the likelihood that a message contains supportive messages. Also, commonly used with these types of maneuvers are the positive narrative maneuvers such as excite maneuver messages that exude happiness in reference to a topic related to a particular actor or explain maneuvers to detail the reasons why a person is positively associated with a topic.

Illustrative Potential Impact

The impact of a backing maneuver is the positive development of an opinion leader's ego network. A person with a large amount of backing has a dense network of actors positively interacting with them and with connections that reflect on them positively regarding a particular topic as graphically depicted in Figure 2.3 This can be measured by observing the change in degree centrality conducted with sentiment analysis or stance detection. With an extensive following, the leader has a larger reach and can communicate and influence their beliefs and ideas throughout that denser network of actors.

2.2.2 Build

> *Discussion or actions that create a community or create the appearance of a community*

Description

The purpose of a build maneuver is to form the interactions of a community based on similar topics or attributes. Building a community begins with actors and topics that have not been connected previously as a group as defined by their shared attributes or as a topic-oriented community. This maneuver encourages them to connect to form such groups. This maneuver involves a positive or enticing discussion about a particular topic that would bring people together to discuss various topics. The actors are then associated with these topics and with other actors discussing these same topics. Building user communities can be done simply in some social media platforms such as Facebook by creating a group and simply adding users to the group. Based on the topics of interest and other factors and the users involved, the Facebook algorithm would also make the group appear as a recommended group for users. However, on a site like Twitter, where there are no formal groups, communities are more likely to be groups of users with similar attributes (bots), similar ideologies (right-wing conservatives), or discussions of similar topics. In Figure 2.4, the author attempts to build a group surrounding support for President Trump. They discuss making connections and the strength of numbers. This particular post specifically refers

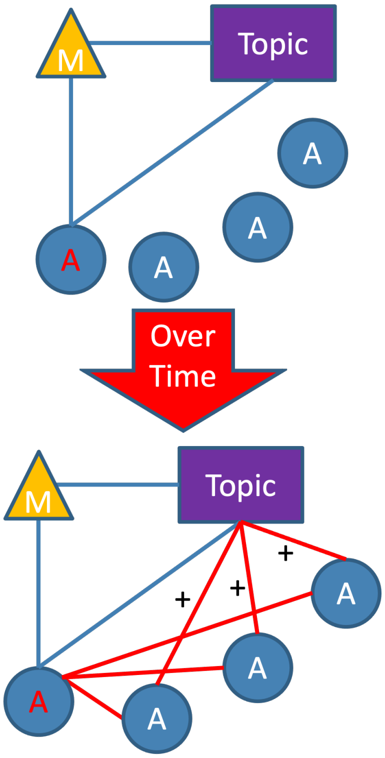


Figure 2.3: Network Impact of *back* maneuver. Overtime, an actor will have more positive connections between other users and a topic.

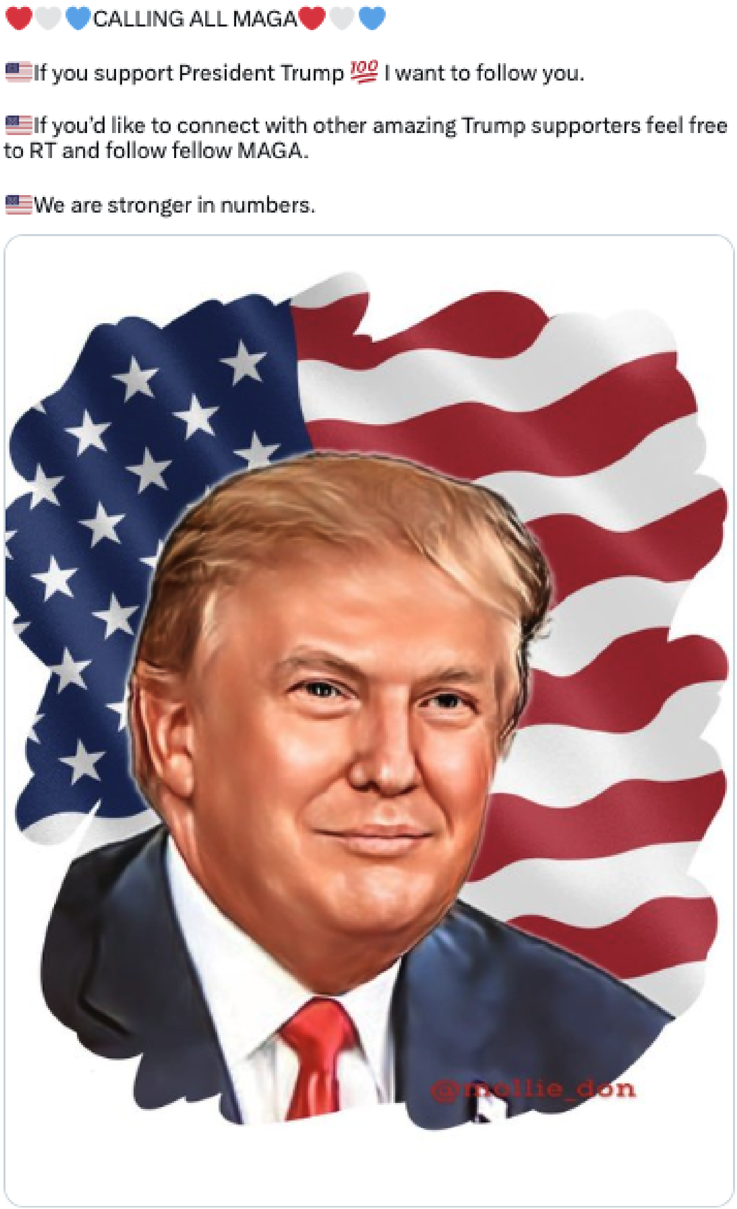
to the Twitter function of following and unfollowing users, which is a method of delineating between different types of communities.

Indications

There are several indications of a build maneuver. The topic refers to other members or is posed in a manner that convinces others to connect with each other through discussion. These references can be either through mentioning a user's name or by mentioning them using the '@' sign. The users reply, quote, and retweet similar people regarding the same topic. Alternatively, users with similar attributes may reply, quote, and retweet similar people. These messages may contain positive narrative maneuvers to create compelling reasons to connect with each other. There may be welcoming terms such as those related to fairness or loyalty, and these may be supported with positive language or emojis. Build messages may also potentially contain negative narrative maneuvers if the topic that defines the group is negative in nature. This can be seen in messages related to anti-vaccine discussions or other discussions related to sad or angry topics.

Illustrative Potential Impact

Build maneuvers create groups of people that share common beliefs, ideas, and sometimes behaviors as shown in Figure 2.5. The impact of these maneuvers is a path for information to



❤️🤍💙CALLING ALL MAGA❤️🤍💙

🇺🇸If you support President Trump 💯 I want to follow you.

🇺🇸If you'd like to connect with other amazing Trump supporters feel free to RT and follow fellow MAGA.

🇺🇸We are stronger in numbers.

Figure 2.4: Example of *build* maneuver. Building a group surrounding the topic of support for President Trump based on *follows* and *retweets*.

flow and a potential for it to become amplified as an important topic among the people who share them. As a community, they are likely to see similar information more often than if they were not within the group increasing the likelihood for the members to believe the messages or take action based on the discussions of the group. Detection of this impact can be observations of the existence of the group after the maneuvers either as a topic-oriented community or as a community joined by a common attribute.

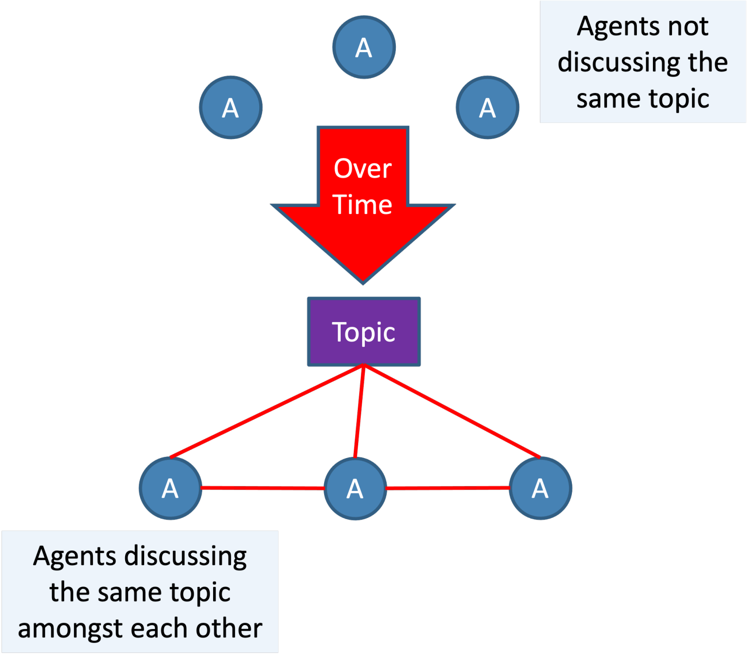

A
A
A
Agents not discussing the same topic
Over Time
Topic
A
A
A
Agents discussing the same topic amongst each other

Figure 2.5: Network Impact of *build* maneuver. Overtime, a community is created based on a given topic or attribute.

2.2.3 Bridge

> *Discussion or actions that establish a connection between two or more groups or create the appearance of such a connection*

Description

The purpose of a bridge maneuver is to establish a connection between two disparate communities that do not typically interact with one another. The maneuver attempts to join groups, combine them, or create alliances between communities that would be otherwise disconnected. This often occurs when one larger group of users with the same ideology deliberately engages with smaller groups aiming to absorb them into their community. This can also occur when two similarly structured groups combine to make a larger group that supports a general topic or share a belief or idea. An example of this is when residents across multiple countries united in support of Ukraine in their war against Russia. During elections, candidates of one party may try to connect with groups of supporters from the opposing party by extending their reach across party lines.

What can happen during a bridging maneuver is that a bridging agent from one community attempts to connect with members of another community either through mentioning, following, or other forms of interaction. In return, members of the other groups begin to interact with the "bridger" while discussing topics of the other group. Some social media platforms offer recommendations of users to connect with based on these connections and interactions. Soon multiple members of one group begin sharing and discussing topics of the other group. The groups are bridged as their discussions merge into a common discussion or ideology.

Some types of actors, such as superfriends, act as bridges between groups by engaging in multiple two-way communication with different groups. Bots can also act as bridges as they engage with other users. This occurs not necessarily because the bot knows the person or is particularly interested in their conversation but because an algorithm links to them based on specific features within their messages. If two-way communication ensues with targeted users of another community, and the bridger is confidently within multiple groups, then this indicates a successful bridging maneuver.

Indications

Indications of bridging maneuvers are mentions or interactions between users of one community with another. These communities are clearly defined in the network either by their topics or a similar attribute, and the bridger straddles both communities. The messages for these maneuvers may include inclusive terms that will facilitate a bridger becoming more accepted by a new community. They exude feelings of trust or a shared belief. Furthermore, bridging can be done through the use of hashtags. For example, anti-vaccine proponents may desire to unrelated communities by tying completely unrelated hashtags such as #kimkardashian or #simpsons just to gain the attention of communities interested in those topics. Because of how Twitter and other social media platforms use hashtags to index topics, as these communities search on these topics, anti-vaccine proponents who use these hashtags will be able to insert their messages in the newsfeeds and search of members from these unrelated communities.

Illustrative Potential Impact

Combining groups creates a larger network consisting of members from both communities as shown in Figure 2.6. This means that information at the far end of one community has a path toward the end of another community. Members of the communities are connected to more users. This makes them more vulnerable to the ideas of others but also gives them the power to influence the beliefs and behaviors of more people directly. This can be detected through identify the existence of two communities in an earlier time period and then observing the joining and strengthening of ties between the groups. Users joining the two groups in the initial period would have a high betweenness centrality, and overtime, the groups will have more connections between each other through other community members.

2.2.4 Boost

Discussion or actions that increase the size of a group and the connections among group members or the appearance of such

Description

The purpose of a boost maneuver is to take a pre-existing community and increase its size or to increase the interaction and strength between the group members and new group members. The

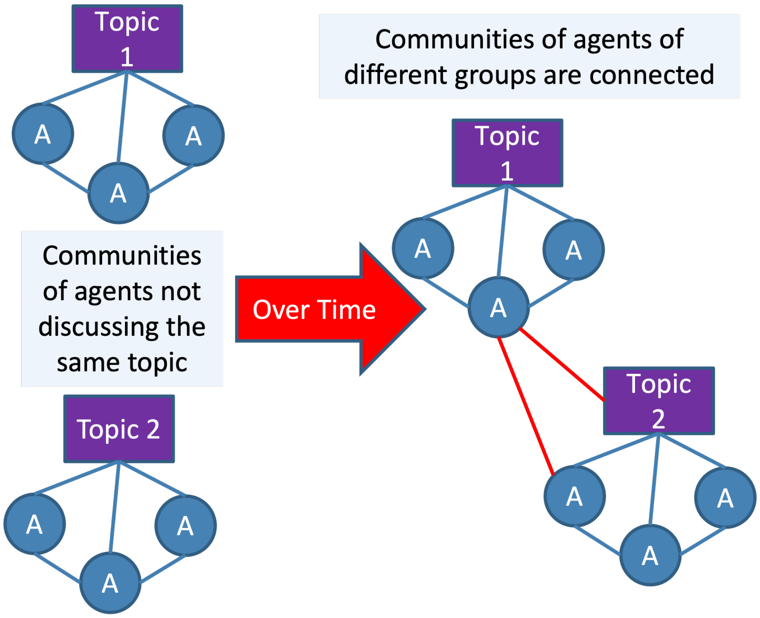

Topic 1
A
A
A
Communities of agents not discussing the same topic
Topic 2
A
A
A
Over Time
Communities of agents of different groups are connected
Topic 1
A
A
A
Topic 2
A
A
A

Figure 2.6: Network Impact of *bridge* maneuver. Overtime, two communities are joined.

actions for conducting a boost maneuver are similar to those for a build and a bridge maneuver. Whereas build maneuvers form groups and communities and bridge maneuvers join them, this maneuver boosts the size and density of a pre-existing group. Though the increased connections with more users do facilitate the spread of messages throughout the network, this is not to be confused with the concept of boosting discussions through propagating messages with retweets and quotes.

There are many examples of groups that attempt to boost the size of their communities. Religious organizations attempt to proselytize the unbelievers of their faiths. Political parties strive to gain support for their causes and candidates. Businesses and sports teams seek to increase their consumer or fan base. These groups target individuals through messaging and mentioning in an attempt to increase the strength of their communities and their ideas and beliefs. Furthermore, bots enable the quick boosting of communities as they can interact with large amounts of users at one time. As targets of these bots interact with them, they are growing the size of the group.

Indications

Boost maneuvers are conducted for already formed communities. An attempt to increase that community may include mentions to specifically target individuals or be generally directed to a target audience. Some indicators within boost messages include inclusive and positive language to create a welcoming community and positive and loyalty-related terms to encourage others to join and remain part of the group. The network structure in overtime for a group impacted by boosting would have more user nodes and/or a greater density of communications between users.

Illustrative Potential Impact

The impact of a boost maneuver is a larger community of members discussing the same topic with the same sentiment. Figure 2.7 shows how a group as related to a topic grows over time. Members of this more expansive community share a similar belief, stance, or ideology or have a more extensive reach. This means that messages reach more people and have the potential to become amplified by the users who see, agree, and share these messages within the group.

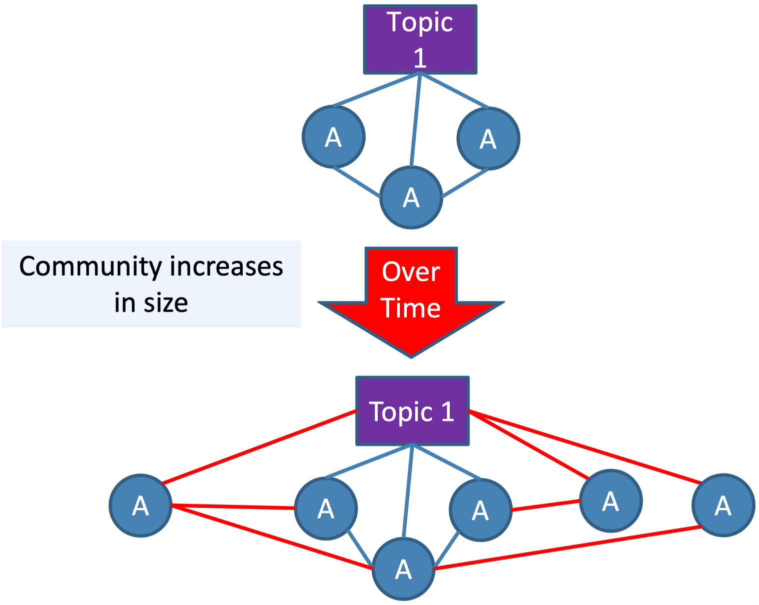


Figure 2.7: Network Impact of *boost* maneuver. Overtime, a community increases in size and/or connections.

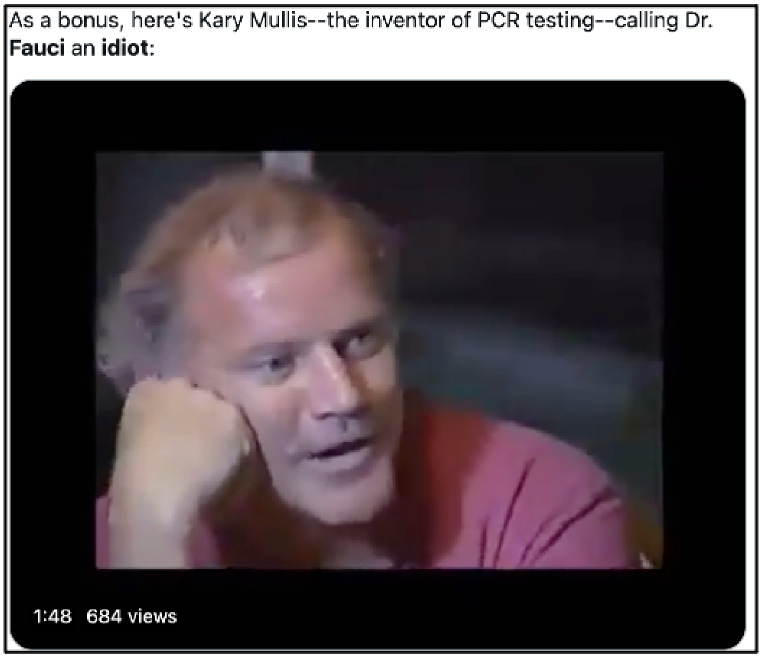
2.2.5 Neutralize

Discussion or actions that limit the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic

Description

The purpose of a neutralize maneuver is to reduce an actor's support, importance, credibility, and ability to effectively influence and communicate within the network. This can be through large amounts of negative discussion that drowns out or invalidates a leader's message, or it can be actions that remove the connections that tie a leader and their message to a community of followers. This maneuver is often used as a complementary maneuver to the back maneuver.

Neutralize maneuvers are applied against leaders or potentially influential leaders. During elections, candidates are neutralized as their opponents aim to discredit them and diminish their support base. Organizations can be the targets of neutralize maneuvers as was the case with the angry messages that stemmed from when Southwest Airlines canceled thousands of flights during the holidays in 2022. In Figure 2.8, the author of the message attempts to neutralize Dr. Anthony Fauci, the medical advisor to the President of the U.S. who recommended policies to mitigate the spread of the COVID-19 vaccine during the pandemic.



As a bonus, here's Kary Mullis--the inventor of PCR testing--calling Dr. **Fauci** an **idiot**:

Figure 2.8: Example of *neutralize* maneuver. Decreasing importance and credibility of Dr. Fauci, a COVID-19 policy advisor.

Indications

Neutralize maneuvers reference or mention the leader that is being targeted. The messages of the maneuvers may contain any of the negative narrative maneuvers to help discredit or decrease the effectiveness of the actor. These can contain negative terms, angry or fear-related terms or emojis, expletives, or abusive language. They may use positive narrative maneuvers such as explain or enhance to create supportive arguments based on a negative narrative against the leaders.

Illustrative Potential Impact

The impact of a neutralize maneuver is more negative discussion against the opinion leader as shown in Figure 2.9. Their credibility, support, and ability to influence the network are lessened. They may have fewer connections, or degree centrality, if followers decide to unfollow them because of their decreased interest, but this may not necessarily be the case. The leader may continue to maintain or grow in connectivity to others or degree centrality, but the predominant sentiment or stance surrounding the leader is negative or counter to the leader's ideology. Ultimately, the leader has limited reach and effectiveness.

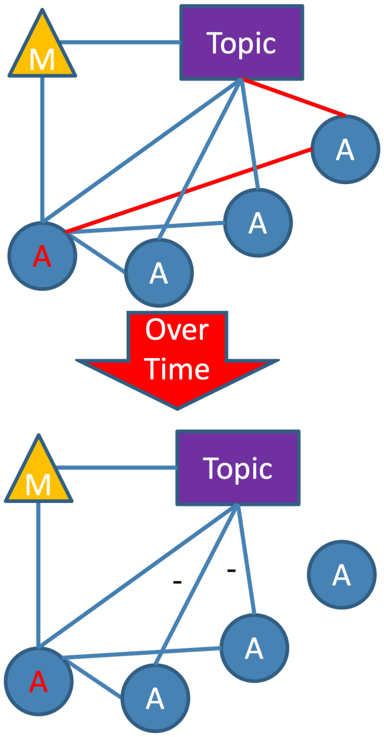


Figure 2.9: Network Impact of *neutralize* maneuver. Overtime, an actor as related to a topic loses connections to other actors or has more negative connections to actors resulting in a decrease of importance and/or reach.

2.2.6 Nuke

Discussion or actions that cause a group to be dismantled or appear to be dismantled

Description

The purpose of a nuke maneuver is to dismantle or destroy an online community. A group that is formed around a specific topic, ideology, or user attribute can no longer spread information related to that group to members of that community. This dismantling of a group can occur in multiple ways. The central topic of the group may no longer be relevant or be discredited as a topic no longer worth discussing. Members of a group may choose to leave the group because they do not feel that their ideologies are no longer in line with that of the community. Beyond the available actions of the individual actors, a social media platform may decide to completely remove users of a particular community and their messages if their discussions fall outside of the platform's policies. Figure 2.10 shows a headline for Twitter's policy against anti-COVID vaccine tweets. The ban eventually lead to the extensive removal of anti-vaccine users and tweets, nearly destroying the online presence of the anti-vaccine community.

Anti-Covid vaccine tweets face five-strikes ban policy

2 March

Figure 2.10: Example of *nuke* maneuver. Twitter removes accounts and tweets based on policy against spreading COVID vaccine misinformation, dismantling the COVID-19 anti-vaccine community

Indications

Nuke maneuvers are aimed at dismantling a group. Negative narrative maneuvers are useful tools for making a topic that binds a topic-oriented community irrelevant or not worthy of discussion. Negative or abusive language or emotions related to the topic or the group members are also methods used for making members quit a community and a community's discussion. Furthermore, though not detectable within individual messages, policies directed at removing users based on ideologies or attributes are platform-level methods for nuking communities online.

Illustrative Potential Impact

The impact of a nuke maneuvers is that a targeted community ceases to exist. Members either no longer exist or they are no longer communicating with each other about the same topics. This means that the ideas that brought the community together no longer have a viable path to efficiently reach a target audience. Not only does this limit the spread of the group's ideas and beliefs, but it also limits the ability of members to coordinate, mobilize, or affect behaviors. Overtime the number of nodes and connections that are used to characterize the group are significantly less or non-existent.

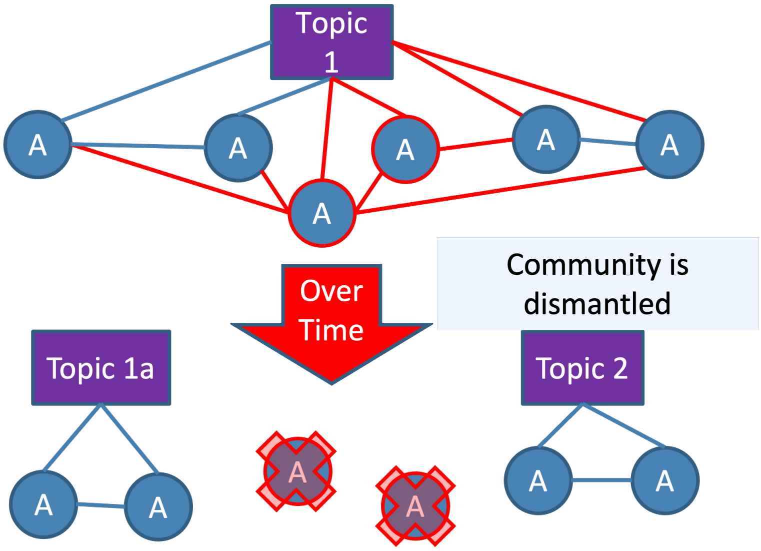


Figure 2.11: Network Impact of *nuke* maneuver. Overtime, a community is dismantled.

2.2.7 Narrow

Discussion or actions that lead a group to fission (or appearance of fission) into two or more distinct groups

Description

The purpose of a narrow maneuver is to take a larger group and break it into smaller groups that no longer identify with the larger community. This is similar to how civil wars can split a once unified country into separate factions. For online communities, this can be seen as marginalizing or singling out sub-communities or users with particular attributes or ideas.

Indications

Indications of a narrow maneuver may include terms of exclusivity that cause members to no longer feel part of the larger community. This can be directed in ways that isolate, marginalize, or exile a group from the larger community. Negative and abusive terms may be used in conjunction with exclusivity terms to make a group feel unwelcome.

Illustrative Potential Impact

The impact of narrow maneuvers is that the community is divided into sub-communities that are no longer a part of the whole community as shown in Figure 2.12. The larger group is no longer as large and their focal topics are no longer shared or discussed by all of the sub-communities. The sub-communities are separated because they do not share the same unifying attributes or ideas that joined them as the larger community. The sub-communities no longer communicate seamlessly with each other as they did when they maintained connections with all of the other groups. The spread of messages is stifled as the paths across the communities touch fewer members and the ability to mobilize more people and share beliefs and ideas is generally limited to the group that has formed as a result of the maneuver. This can be shown through the overtime development of sub-topic-oriented groups or smaller communities within the larger community emerging within the dataset defined by more distinct qualities, attributes, or conversations.

2.2.8 Neglect

Discussion or actions that decrease the size of the group, or the connections among the members, or the appearance of these

Description

Neglect maneuvers reduce the size of communities. This is done through the decrease in interactions and connections between the group members. The aim of a neglect maneuver is similar to the nuke maneuver, but the desired end state is not to completely destroy the community. The group members and their topics are targeted so as to reduce the cohesion a user feels with the

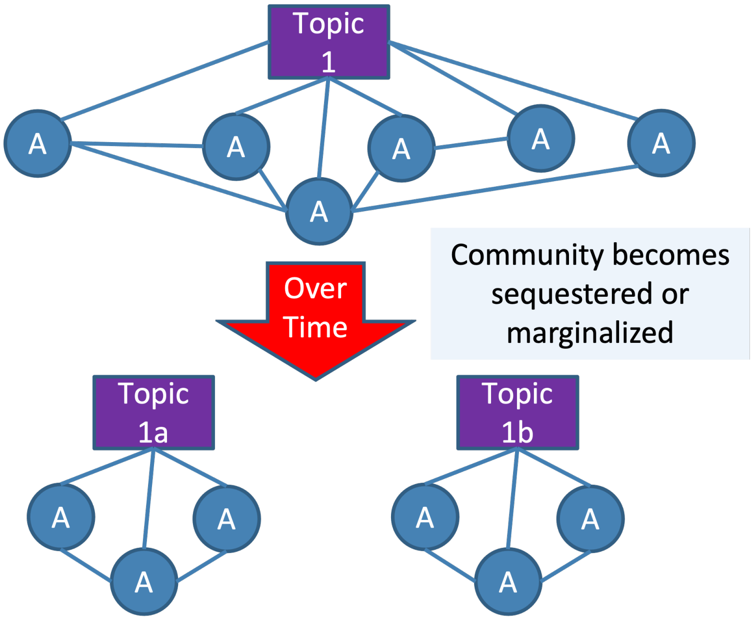


Figure 2.12: Network Impact of *narrow* maneuver. Overtime, a community is fissioned into distinct groups.

community and to decrease the relevance of the topic as an idea to maintain the existence of a group. Pro-vaccine proponents attempt to reduce the numbers of anti-vaccine communities. They either attempt to nullify their arguments or they make fun of them as idiots, trying to dissuade members from wanting to be a part of their community.

Indications

Indicators for a neglect maneuver are negative language and terms that reduce the discussion of the topic for topic-oriented groups and/or exclusivity terms that discourage members from being a part of the community. These can be in the form of negative narrative maneuvers. On the contrary, positive narrative maneuvers may also be used within a neglect maneuver if it bolsters the arguments for leaving the community. Attempts to reduce the cohesion that keeps the group together are all indications of a neglect maneuver.

Illustrative Potential Impact

The impact of a neglect maneuver is that the community is reduced in size overtime as depicted in Figure 2.13. The reach of messages within the community is limited and the extent of influence is reduced to the size of the group. The community is less cohesive with a decreased density of interactions between users. With that, the ability to form amplifying messages of influence among users is also lessened with the reduced connections and group size.

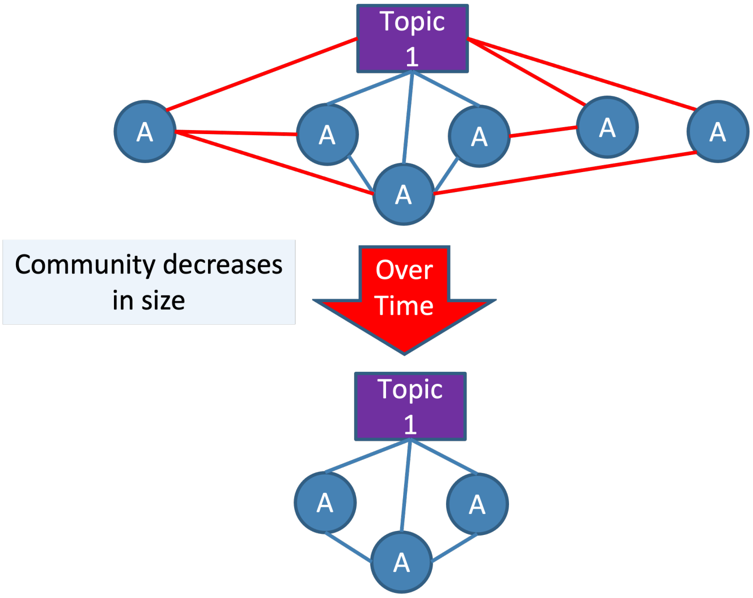


Figure 2.13: Network Impact of *neglect* maneuver. Overtime, a community decreases in size and/or density.

2.3 Narrative Maneuvers

Narrative maneuvers, sometimes called information maneuvers, focus on altering the information within the network. These maneuvers are contained within the content of messages and are "what" the actors are talking about making them more straightforward for humans to detect than community maneuvers. They may be a single original or retweet of a message, a reply or quote of an original message, or part of a larger discussion regarding a particular topic. The maneuvers center around a topic and frame messages to support an influencer's overarching narrative about their ultimate objectives.

The narrative maneuvers are divided into positive and negative based on how the maneuver increases or adds in a metric for the network. The positive narrate maneuvers are the 'E' maneuvers: engage, explain, excite, and enhance. Explain and enhance add information or details to clarify or support an idea. Engage increases the personnel connection in the topic. Using a different metric type, excite attempts to elicit positive emotion from the readers. The negative maneuvers are the 'D' maneuvers: dismiss, distort, dismay, and distract. They attempt to decrease or reduce a particular metric for the narrative. Dismiss and distract minimize importance or attention from a particular topic. Distort reduces the veracity or integrity of the original message. Similar to the excite maneuver, the dismay maneuver focuses on eliciting emotion. In this case, negative emotions.

In a well-connected network, narrative maneuvers can cause increased discussions or virality or slow or completely stop a trending topic. Narrative maneuvers can make topics appear more important or interesting, make them appear less important or interesting, or change the original message's meaning. They affect how readers perceive the ideas and whether they are ideas worth passing on.

The online impact of the narrative maneuvers involves increased discussion of the topic in the same sentiment as the author. The descriptions, interpretations, or views of the topic are more prevalent throughout the network. Some tools for measuring the extent of the topic are stance detection, polarization measures, and sentiment analysis of the number of the individual messages or the number of users who are characterized by the ideology expressed through the narrative.

2.3.1 Engage

Discussion or actions that create a personal affinity between the target community or actor and the topic

Description

The purpose of an engage maneuver is to create a personal connection between the topic and the target audience. This involves making the topic appear important to them or applicable to their lives. This can be done through the use of anecdotes or examples that may create a relationship between the topic and possible personal experiences. If successful, that affinity between the topic and the reader may ultimately lead to the readers taking ownership of the narrative.

In the example in Figure 2.14, the author attempts to use an engage maneuver to encourage their target audience to take precautionary steps to mitigate the spread of COVID-19 during the holidays. They state the need for "everyone's help" and that "hospital beds may not be available when you need them" if the spread does not stop. The message attempts to put the onus on the reader with #ThisIsOurShot, meaning that this is important for everyone, and doing "your part to break the chain of transmission" will help stop the pandemic. These are then followed up with hashtags to wear a mask, get a booster, and get vaccinated, which are the aims of the message.

Indications

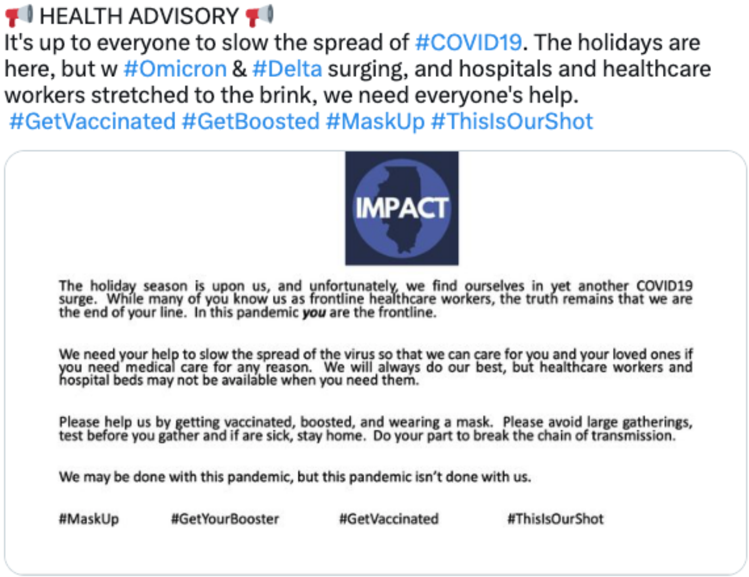
Engage maneuvers may contain inclusive terms and second-person pronouns that direct the topic to the reader. Other indicators may be images, videos, or URLs that may serve to enhance the personal connection with the target audience.

Illustrative Potential Impact

The impact of an engage maneuver is the personal affinity between the reader and the topic that ultimately convinces the reader to believe in the ideas or take part in the actions presented in the message. The feelings and sentiments surrounding the topic spread throughout the network with more users discussing the topic and buying into the message.

2.3.2 Explain

Discussion or actions that provide details on, or elaborate on, a topic to the target community or actor



📢 HEALTH ADVISORY 📢
It's up to everyone to slow the spread of #COVID19. The holidays are here, but w #Omicron & #Delta surging, and hospitals and healthcare workers stretched to the brink, we need everyone's help.
#GetVaccinated #GetBoosted #MaskUp #ThisIsOurShot

The holiday season is upon us, and unfortunately, we find ourselves in yet another COVID19 surge. While many of you know us as frontline healthcare workers, the truth remains that we are the end of your line. In this pandemic ***you*** are the frontline.

We need your help to slow the spread of the virus so that we can care for you and your loved ones if you need medical care for any reason. We will always do our best, but healthcare workers and hospital beds may not be available when you need them.

Please help us by getting vaccinated, boosted, and wearing a mask. Please avoid large gatherings, test before you gather and if are sick, stay home. Do your part to break the chain of transmission.

We may be done with this pandemic, but this pandemic isn't done with us.

#MaskUp #GetYourBooster #GetVaccinated #ThisIsOurShot

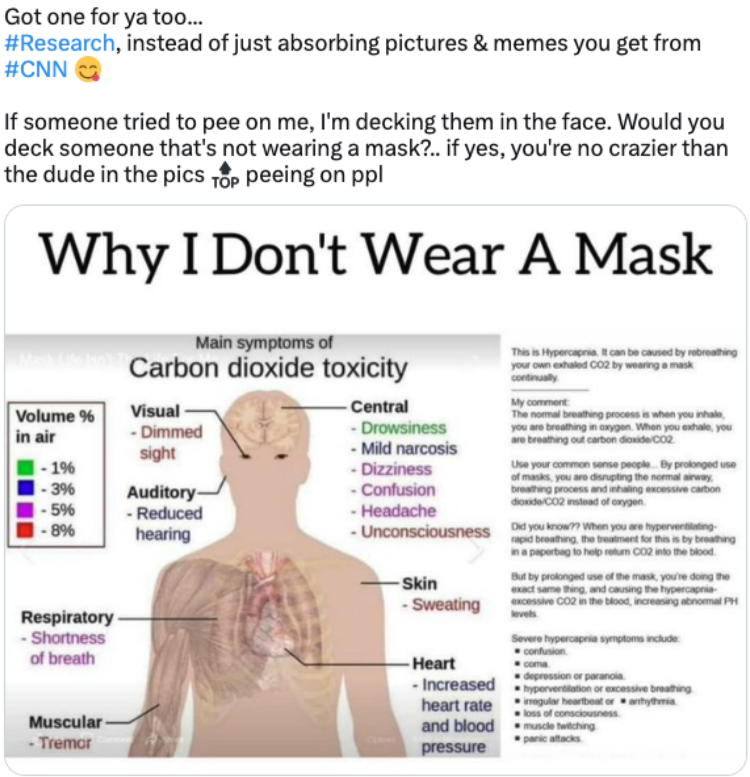
Figure 2.14: Example of *engage* maneuver. Author is making the actions for slowing the spread of COVID-19 of personal importance to their audience, attempting to make them take ownership of the narrative.

Description

An explain maneuver takes a concept and elaborates on the topic through the use of details and explanations. The messages use logic or a pedagogical approach to developing a topic. This can be used to teach or instruct the target audience on specifics about a particular topic. These messages are usually less sentimental and may contain numerical values and charts to help expand on the topic. Explain messages are intended to instruct, teach, or logically expand on a topic. They may elaborate on why something is the way it is, how something works, or what the reader may need to be informed about the topic.

Explain maneuvers are very similar to enhance maneuvers in that they expand on a particular topic with supportive explanations and evidence. They differ, though, in that sometimes an enhance maneuver may provide non-explanatory information to support the topic such as comical memes, buy-in from celebrities, or anecdotes.

Figure 2.15 is an example of an anti-mask explain maneuver during the COVID-19 pandemic. First, the author starts off with #Research to prepare the reader for an intellectual discussion. They then try to use a logical fallacy to compare being peed on by someone to being near someone without a face mask. Finally, an infographic is presented as evidence of the reasons why not to wear a mask. Numerical values, images of the human anatomy, and medical terms are all used to support the author's argument. The author uses all of this to create detailed instructions to explain their point of view.



Got one for ya too...
#Research, instead of just absorbing pictures & memes you get from
#CNN 😋

If someone tried to pee on me, I'm decking them in the face. Would you deck someone that's not wearing a mask?.. if yes, you're no crazier than the dude in the pics 🔝 peeing on ppl

Why I Don't Wear A Mask

This is Hypercapnia. It can be caused by rebreathing your own exhaled CO2 by wearing a mask continually.

My comment:
The normal breathing process is when you inhale, you are breathing in oxygen. When you exhale, you are breathing out carbon dioxide/CO2.

Use your common sense people... By prolonged use of masks, you are disrupting the normal airway, breathing process and inhaling excessive carbon dioxide/CO2 instead of oxygen.

Did you know?? When you are hyperventilating-rapid breathing, the treatment for this is by breathing in a paperbag to help return CO2 into the blood.

But by prolonged use of the mask, you're doing the exact same thing, and causing the hypercapnia-excessive CO2 in the blood, increasing abnormal PH levels.

Severe hypercapnia symptoms include:
- confusion.
- coma.
- depression or paranoia.
- hyperventilation or excessive breathing.
- irregular heartbeat or ▪ arrhythmia.
- loss of consciousness.
- muscle twitching.
- panic attacks.

Figure 2.15: Example of *explain* maneuver. Author uses statistics and instructional infograph and discussion to elaborate why he/she does not wear a mask.

Indications

Several indicators for an explain maneuver include numbers, images, URLs, and minimal negative and positive terms as an attempt to provide an unbiased argument. The maneuvers have a pedagogical tone and attempt to use logic to sway their audiences. There may be references to science, statistics, graphs and charts, or other types of supportive evidence to elaborate on and strengthen favor regarding a particular topic.

Illustrative Potential Impact

The impact of an explain maneuver is a more informed audience regarding the topic. As the premise for the topic is based on logic and facts, there is a more compelling reason for the reader to accept the narrative. This justification may lead to more discussion of the topic with the same sentiment as the original message throughout the network.

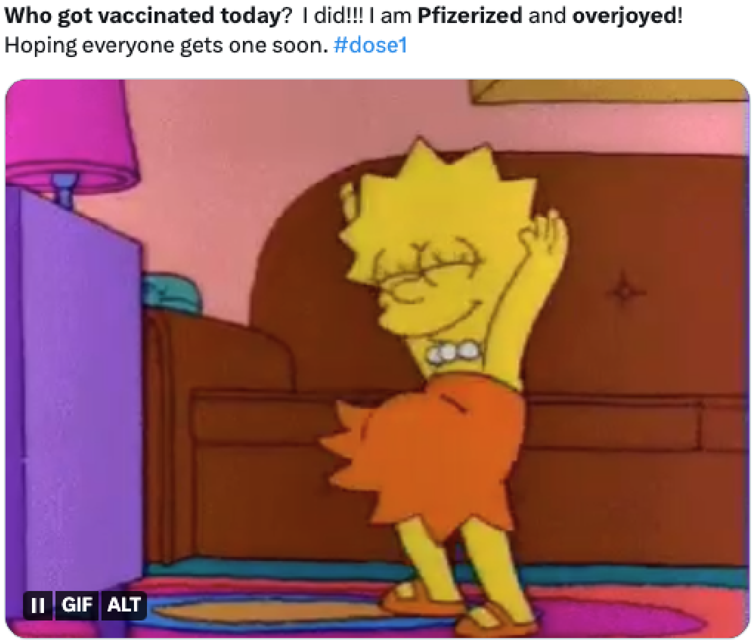
2.3.3 Excite

Discussion or actions related to the topic that bring joy, happiness, cheer, enthusiasm in the targeted community or actor

Description

Messages of this type attempt to create a state of euphoria, happiness, joy, or excitement to elicit an emotional behavioral response. This overwhelming emotion can lead to an ‘amygdala hijack’ [59]. This is the condition when critical thinking gives way to emotions resulting in people acting in ways different than they typically would have.

Users executing excite maneuvers take a topic or theme and add positive emotion or sentiment to gain support from the target audience. Some indications that a message is likely this type of maneuver are positive emojis and emoticons, positive emotional tone, positive terms, capital letters, or exclamation points. In Figure 2.16, the tweet contains exclamation points, the term “overjoyed”, and an image of Lisa Simpson doing a happy dance to show excitement for the user getting the COVID-19 vaccine. The author of this tweet uses these techniques within the body of the message to elicit positive emotions from the reader and then influence others to also get the vaccine.



Who got vaccinated today? I did!!! I am **Pfizerized** and **overjoyed!**
Hoping everyone gets one soon. #dose1

Figure 2.16: Example of *excite* maneuver. Author expresses positivity to encourage users to get the vaccine.

Indications

Many indications of an excite maneuver are the same indications for positive emotions. This can include positive terms and emojis, happy terms, terms related to caring, exclamation points, and capital letters. First person pronouns may also be used to demonstrate the author’s personal feelings about the topic in hope of passing those feelings on to the reader.

Illustrative Potential Impact

The impact of an excite maneuver is the increased spread of the topic throughout the network with users sharing the same feelings and sentiments as the original author in their discussions.

2.3.4 Enhance

> *Discussion or actions that provide supportive material that expands the topic for the targeted community or actor*

Description

An enhance maneuver adds supportive or interesting material to a discussion to provide context and maintain interest in the topic. This can be done through anecdotes or references to celebrities or authority figures acting on or supporting the topic. Entertaining or engaging videos, images or memes, URLS, and hashtags are also effective supplements to a message to make the topic interesting.

As enhance maneuvers are meant to expand on a topic, there may be an overlap between an enhance maneuver and an explain maneuver. Enhance maneuvers go beyond the instructional nature of an explain maneuver and may enhance using other compelling means.

In Figure 2.17, the author uses an enhance maneuver to convince readers to get the COVID-19 vaccine. This message is supported by references to three former US Presidents standing together to encourage people to get the vaccine and proclaim their confidence in its safety. The message is supported with a link that elaborates on the event and shows an image of the presidents standing next to each other.

Indications

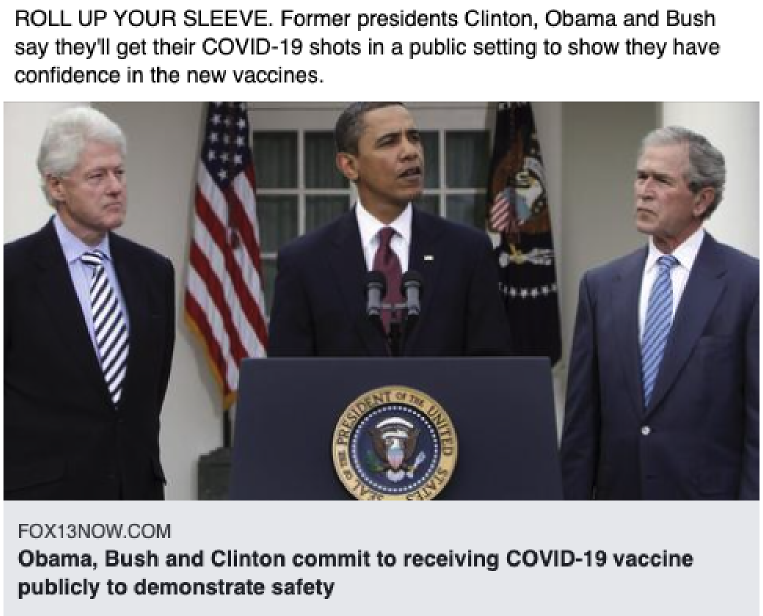
Some of the indications for an enhance maneuver include compelling or amusing videos, images, and URLs. References to other actors and their opinions are also useful to validate or support a topic.

Illustrative Potential Impact

The impact of an enhance maneuver is to generate more discussion and supportive material about the topic throughout the network with the same sentiment as the original message.

2.3.5 Dismiss

> *Discussion or actions that suggest that the topic is not important to the targeted community or actor*



ROLL UP YOUR SLEEVE. Former presidents Clinton, Obama and Bush say they'll get their COVID-19 shots in a public setting to show they have confidence in the new vaccines.

FOX13NOW.COM

Obama, Bush and Clinton commit to receiving COVID-19 vaccine publicly to demonstrate safety

Figure 2.17: Example of *enhance* maneuver. Author draws on the expertise and importance of former US presidents to encourage audience to get COVID-19 vaccine.

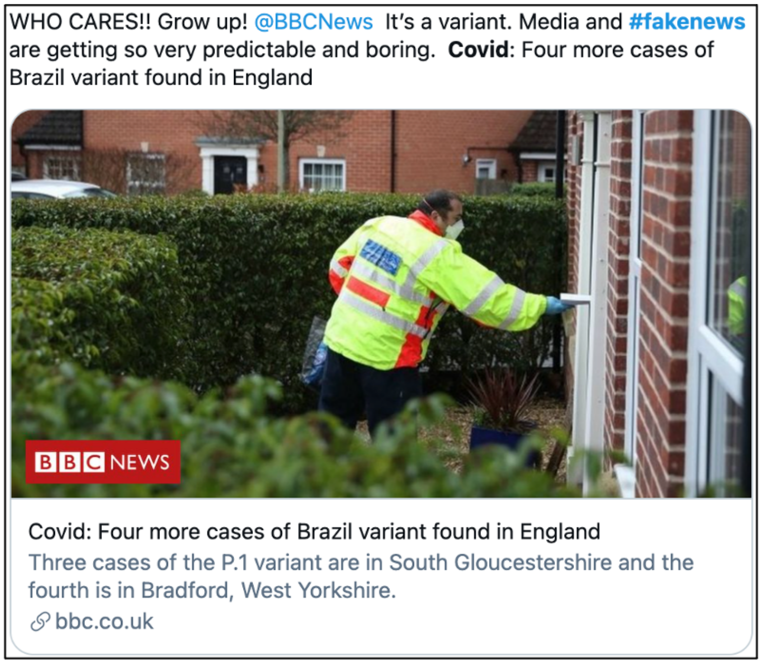
Description

The purpose of a dismiss maneuver is to take a possibly trending topic and reduce the importance and interest in that topic. This may be because it is against an actor's current narrative or potentially discredits them or something important to them. A dismiss maneuver can be executed by providing reasons about why the topic is not relevant or that it is a foolish idea. Additionally, the author of a dismiss maneuver may simply choose to deny the existence of the topic, completely ignore it, or tell others to ignore it, regarding the topic as inconsequential. Furthermore, if the topic is aimed to discredit someone, a dismiss maneuver can be more offensive by either accusing the author of the opposing narrative so as to reduce the credibility of the accuser and the significance of the topic.

In Figure 2.18, the author is attempting to dismiss the issues surrounding COVID-19. They belittle the authors by saying, "WHO CARES!! Grow up!" They provide reasons about why COVID-19 is no longer relevant, which is because it is "very predictable and boring." The author then provides a link to an article that they insinuate is another boring report about COVID-19.

Indications

Indications of a dismiss maneuver may include terms that discuss a lack of care or interest in a particular topic. It can also contain accusatory language directed at another actor in an attempt to discredit messages they may be stating against the author. There may also be indications of humor or sarcasm within the message aimed to make the original topic appear unimportant or ridiculous. This can be done through replies, quotes, and original tweets.



WHO CARES!! Grow up! @BBCNews It's a variant. Media and #fakenews are getting so very predictable and boring. **Covid**: Four more cases of Brazil variant found in England

Covid: Four more cases of Brazil variant found in England

Three cases of the P.1 variant are in South Gloucestershire and the fourth is in Bradford, West Yorkshire.

bbc.co.uk

Figure 2.18: Example of *dismiss* maneuver. Author dismisses discussions of COVID as unimportant and uninteresting.

Illustrative Potential Impact

The impact of a dismiss maneuver is less discussion about the topic within the network. This can be in the form of less messages that contain the topic being dismissed.

2.3.6 Distort

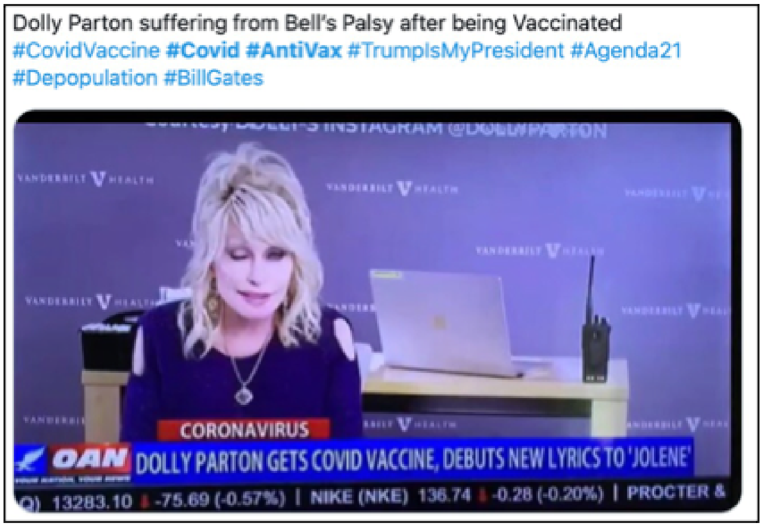
Discussion or actions that provide unsupportive material that slant the topic for the targeted community or actor

Description

The purpose of a distort maneuver is to retell a narrative to support a preferred story. This is done by changing or reinterpreting facts, omitting context, or providing facts to alter the message and change the target audience's perspective on the narrative. The message delivered through a distort maneuver is actually a form of fiction.

Distort maneuvers can be difficult to detect from a single message. In isolation, the structure of the message can appear as a positive narrative maneuver developing a desired storyline. For example, in Figure 2.19, here is a convincing message aimed against the COVID-19 vaccine. An image clearly depicts an unflattering image of Dolly Parton with a caption on the news clip headlining her getting the COVID vaccine. The author of the message's comments on the image say that her unflattering image was because she was suffering from Bell's Palsy after being vaccinated. However, in truth, this image was an unfortunate snapshot of Dolly Parton during an interview, and she did not suffer from Bell's Palsy after getting the vaccine. The author took the image out of its original and added text to fit their anti-vaccine narrative. All distort maneuvers are forms of disinformation.

Therefore, it is possible that a distort maneuver may only manifest itself when placed in its proper context alongside factual information or when juxtaposed with the topic that it is aimed at distorting.



Dolly Parton suffering from Bell's Palsy after being Vaccinated #CovidVaccine **#Covid #AntiVax** #TrumpIsMyPresident #Agenda21 #Depopulation #BillGates

CORONAVIRUS
OAN
DOLLY PARTON GETS COVID VACCINE, DEBUTS NEW LYRICS TO 'JOLENE'
Q) 13283.10 -75.69 (-0.57%) | NIKE (NKE) 136.74 -0.28 (-0.20%) | PROCTER &

Figure 2.19: Example of *distort* maneuver. Author sends disinformation in the form of an altered interpretation of Dolly Parton's image. She does not have Bell's Palsy as the author suggests.

Indications

Specific indicators for distort maneuvers vary depending on whether or not the original narrative is contained within the text. They may include sarcastic comments, false information, altered information, and information taken out of context. These indications may be difficult to detect if the truth of the message is unknown or if the reader has a bias toward the message.

Illustrative Potential Impact

The impact of a distort maneuver may be the spread of erroneous or manipulated versions of the topic that do not fit the original narrative as well as less discussion of that narrative.

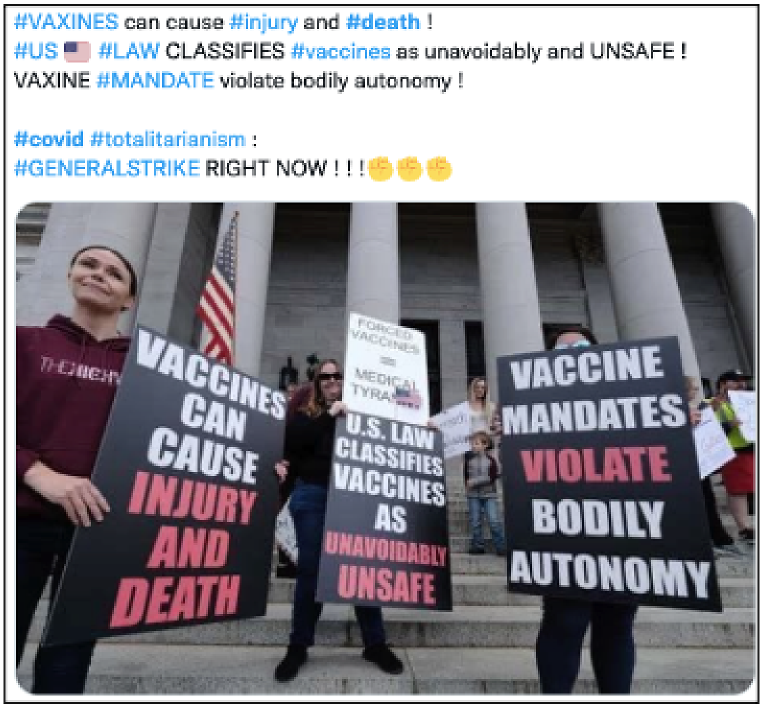
2.3.7 Dismay

> *Discussion or actions related to the topic that create worry, sadness, anger, or fear in that targeted community or actor*

Description

Dismay maneuvers attempt to evoke negative emotions from the target audience to achieve a particular end state. This may be by inducing a state of sadness, fear, anxiety, or anger. Similar to the excite maneuver, this emotional maneuver is meant to reduce the target audience's critical thinking skills using an 'amygdala hijack.[59]' This is the human fight or flight response of behavioral actions responding to an emotional stimulus. A dismay maneuver uses the amygdala hijack to force the reader to respond emotionally and less rationally about a topic.

The example in Figure 2.20 uses dismay to protest gains for mandatory vaccines. They instill fear by relating vaccines to injury and death and stating that they are legally unsafe, which is typed in all capital letters complemented with exclamation points. These comments may or may not be true, but to the reader who becomes fearful from these comments, they may not necessarily consider fact-checking the injuries and deaths and the US laws regarding vaccines.



#VAXINES can cause #injury and #death !
#US 🇺🇸 #LAW CLASSIFIES #vaccines as unavoidably and UNSAFE !
VAXINE #MANDATE violate bodily autonomy !

#covid #totalitarianism :
#GENERALSTRIKE RIGHT NOW ! ! !✊✊✊

VACCINES CAN CAUSE INJURY AND DEATH
FORCED VACCINES = MEDICAL TYRA
U.S. LAW CLASSIFIES VACCINES AS UNAVOIDABLY UNSAFE
VACCINE MANDATES VIOLATE BODILY AUTONOMY

Figure 2.20: Example of *dismay* maneuver. Author attempts to incite fear about vaccines to their audience.

Indications

The indicators for dismay maneuvers generally represent indications of negative emotions. These include negative terms, abusive and expletive terms, terms for fear, anger, sadness, and sarcasm. They may contain exclamation points, capital letters, and emojis or emoticons related to the dismay maneuver emotions.

Illustrative Potential Impact

The impact of a dismay maneuver is the spread of the topic throughout the network with users sharing the same feelings and sentiments as the original author in their discussions. This may mean a greater number of angry, fearful, sad, or worrisome messages about the original topic and a greater number of users spreading these messages.

2.3.8 Distract

Discussion or actions that redirect the targeted community or actor to a different topic

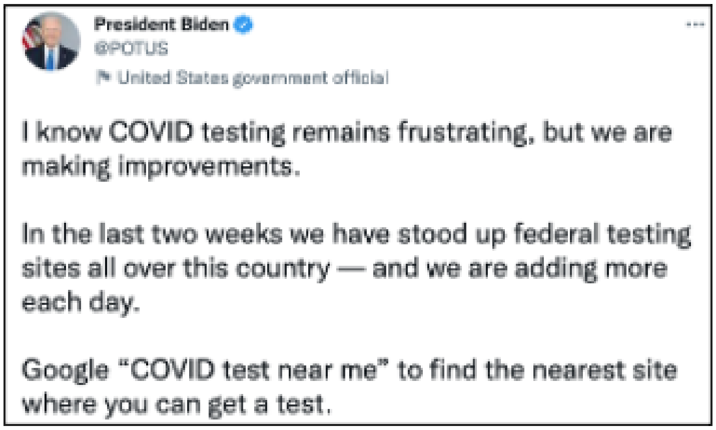
Description

The purpose of a distract maneuver, simply put, is “changing the subject’ to discuss other topics or ideas that seem more compelling. When an actor is threatened by an ongoing narrative or disagrees with undesirable conversations, this maneuver attempts to change e direction of the narrative and limit the conversations discrediting the targeted actor or the topics they support. For example, in Figure 2.21, President Biden attempts to do multiple things within his message. He is trying to encourage others to get COVID tests while attempting to gain trust in the government’s efforts to gain control of the pandemic. In reply to the message, one user responds with ”we need student relief”. This person is completely downplaying the noble efforts to mitigate the spread of the virus by not mentioning the original topic and is attempting to change the focus of the conversation to what they perceive is a shortcoming of the administration. They are distracting readers from President Biden’s original message and intent and redirecting the conversation to what may be perceived as a more pressing issue.

Indications

In a message, an actor may use a small reference to the original topic followed by a more extensive and enthusiastic discussion about something else. These messages may contain coordinating conjunctions that are tools to counter the original message such as ”but” or ”although.” It is also possible that distract messages may have a mix of positive and negative terms or emojis. A user may use these terms to condemn the original topic and then speak positively about a seemingly more compelling issue. ”Hey, this is bad, but look at all these good things!” Though, the distract part of the message may also follow in the same sentiment. ”Hey, this is bad (or good), but look at all of these things that are even worse (or better)!”

Distract maneuvers are not always visible in a single isolated message. They can occur as parts of replies or quotes of other messages. They may also be original tweets that occur in the context of an overarching campaign of misdirecting audiences from a particular topic. In these cases, distract maneuvers are observed in the context of the messages and the narratives that surround them within the network.

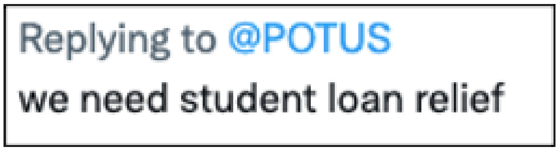


President Biden
@POTUS
United States government official

I know COVID testing remains frustrating, but we are making improvements.

In the last two weeks we have stood up federal testing sites all over this country — and we are adding more each day.

Google "COVID test near me" to find the nearest site where you can get a test.



Replying to @POTUS
we need student loan relief

Figure 2.21: Example of *distract* maneuver. Replies to the original author are unrelated to the issue of COVID. The replier discusses student loan relief as an alternate topic.

Illustrative Potential Impact

The impact for this maneuver is the original and undesirable narrative is no longer discussed or discussed less frequently across the network. If applicable, the author of the original narrative has decreased credibility. Subsequently, the alternate narrative is more prevalent, which can confuse readers as they must decide which narrative is more reliable or worthy.

2.4 BEND Tactical Categories: Combinations of BEND Maneuvers for Effects

The BEND maneuvers can be grouped into categories that provide a higher-level description of an influence campaign. Many leaders care less about describing isolated maneuvers and more about how the maneuvers are working in concert to affect the bigger picture.

2.4.1 Develop the Narrative: engage, explain, enhance

Engage, explain, and enhance are positive narrative maneuvers, which can be used to develop the narrative of an influence campaign. Engage uses techniques to pull the target audience toward a particular topic, possibly with relatable anecdotes or topics relevant to the reader. The topic can then be expanded upon using both explain and enhance. Explain supports the narrative with descriptions with supportive facts or logical instructions, and enhance supports it with additional features such as memes and validating quotes from celebrities or experts in the field. For example, in an effort to convince more people to vaccinate their children, one might engage with the audience by discussing the value of their children and showing how their children are no different than the children who fell sick after not being vaccinated. This could then be followed up with scientific research to explain the efficacy of vaccines and how they have saved thousands of lives. Then maybe enhance the narrative with images of *insert popular celebrity* standing in line to get vaccines for their children. These maneuvers, in combination, are creating a stronger narrative.

2.4.2 Counter the Narrative: dismiss, distract, distort

Distract, dismiss, and distort are negative narrative maneuvers that have been used extensively to discuss Russian propaganda [98]. They represent methods for countering narratives, but combined, they reinforce each other as counter-maneuvers. When faced with an undesirable narrative, dismissing the narrative as unimportant or foolish and then distracting the audience with either unrelated narratives or ”more important” discussions can devalue an adversary’s major argument. Distortion alters that narrative with ”alternate facts” or creates disinformation based on half-truths or lack of context. A politician intent on clearing their name after a bombshell report on some wrongdoing may attempt these maneuvers. Maybe he will tell the media that his actions were not *that* extraordinary because ”people do it all the time - look at *famously loved politician*.” Then maybe distract his constituents with talk about how ”we need to focus more on the economy and less on what the tabloids are saying.” Sometimes, however, these tactics are not easily identifiable in single messages. They may be in chains of reply messages or quoted retweets. They may also be an isolated message that occurs in response to an action that occurred offline, in real life. These are some examples and indications of how countering a message can play out online.

2.4.3 Emotionally Influence: excite, dismay

As mentioned earlier, excite and dismay are emotional maneuvers as they aim to trigger an emotional response from their intended audience. An emotionally charged message has the ability to break down cognitive barriers and cause the unsuspecting recipient to potentially do actions that they may not otherwise have done. These are the reasons people become aggressive drivers when they get road rage, give money to charitable organizations after seeing videos of sick children, or how seeing the happiness of hardworking role models inspires people to do better. In a campaign to support Ukraine, dismaying images of displaced Ukrainians and destroyed homes may be the types of messages that convert empathetic Russian supporters or convince Western allies to provide aid and munitions. Legislation on gun control in the US is often reinforced with stories of children lost in mass shootings in school. These examples elicit emotional responses and are tactical methods for influencing thoughts and behavior.

2.4.4 Grow Groups: build, boost, bridge

A large and active community has the ability to extend its reach and convey messages more effectively. They touch many users, and with many of them sending out the same messages, they become reinforced and more convincing to those who see them. Building, boosting, and bridging are methods for creating and growing those communities. On Twitter, this can be done by mentioning users in messages and supporting them with positive information about a topic. Communities of various religious organizations are commonly known for seeking to grow the size of the members of their faiths. Some may hope to increase the amount of good behavior in the world, while others are looking to recruit members for their religious wars. Regardless, community members aiming to grow groups using these maneuvers send messages supporting their groups to connect with outsiders, strengthen the current community, and, if desired, combine other communities with their own.

2.4.5 Reduce Groups: neglect, narrow, nuke

Influence campaigns may also aim to reduce the size of competing or adversarial communities. These are conducted with neglect, narrow, and nuke maneuvers, which lessen the social network connections and decrease the size of or destroy the community or the discussions that bring them together. If not destroy, reducing a group also could divide them into factions of the larger community. Examples of these actions include unfavorable rhetoric used to discredit the community or introducing exclusivity in topics or language within community discussions that cause members to no longer identify with the group. This can lead to accounts no longer being followed, reduced interactions, and the eventual dissolution of the community. Furthermore, reducing groups can occur with the implementation of platform policies. Twitter nuked communities of accounts with their policies to remove accounts and tweets spreading false or misleading COVID-19 vaccine information [21, 128, 129]. As a result, fewer of these messages permeated the social platform, and the community became less effective as an entity for spreading anti-vaccination information.

2.4.6 Affect Leaders: back, neutralize

Back and neutralize are community maneuvers that attempt to change the importance and effectiveness of actors as related to a particular topic. Backing maneuvers increase the connections and positive discussion surrounding opinion leaders, whereas neutralize maneuvers may reduce connections and positive discussions or increase connections of negative discussion surrounding an actor, ultimately reducing their reach and effectiveness as a leader. Though leaders usually refer to actual people, they can also be non-human entities such as news organizations, government entities, or automated accounts. Together, back and neutralize are useful in combination when two actors are at odds with each other. In an election, supporters back their candidates and neutralize their opponents. In regards to Ukraine-Russia, the Russian state accounts back President Putin and neutralize President Zelensky. These maneuvers become a push-pull tactic for gaining support for a particular actor while discrediting another. Furthermore, these types of maneuvers can be amplified with the use of bots. These automated accounts can quickly send large amounts of information for or against an actor. In one case, thousands of Twitter bots posted praise and support for Donald Trump over an eleven-month period, while criticizing potential opponents Nikki Haley and Ron DeSantis, potential Republican candidates for the US 2024 presidential election [78]. Even though these bots are not real people, the expansiveness of their messages provides the illusion of support for the Republican candidate.

2.5 Conclusion

The BEND maneuvers are an integral part of the larger framework to understand influence operations and have actionable information that allows leaders to make decisions. This chapter expanded on the current definitions with detailed descriptions and examples, provided some indicators for identifying them, and illustrated some of their impacts. Additionally, the maneuvers were categorized into categories that coalesced into influence tactics. These help leaders understand general themes supported by fundamental online influence techniques. In the research to characterize influence campaigns by identifying ”who is doing what to whom and with what impact,” these maneuvers answer the ”what.”

Chapter 3

BEND Maneuvers in the Wild. *The Numbers*

3.1 Introduction

The BEND maneuvers are elemental actions online actors use to achieve influence objectives. They are part of the BEND framework [17, 35], which incorporates the maneuvers as parts of a method for analyzing influence campaigns. Therefore, an essential aspect of characterizing these online influence operations is having a tool for accurately detecting them. Currently, the ORA-PRO software [39] can extract the BEND maneuvers from social media data. However, further improvements are required to improve the accuracy of detection.

In this chapter, I develop an iterative process for improving automated BEND maneuver detection. This process is scoped by the following research questions:

- How well do humans detect maneuvers? *This addresses the ability for humans to accurately and efficiently identify the maneuvers.*
- How does automated detection compare with human detection? *This highlights the limitations of both human and automated detection.*
- How can detection methods be improved? *This develops a method for improving automation using statistical analysis to improve the indicators and thresholds.*
- What is a baseline for these detected values? *This provides analysts with a standard for expected results.*
- What maneuvers commonly occur together? *This illuminates possible redundancies in detection and informs BEND tactical categories (Chapter 2.4).*

Seven data sets were used to create this quantitative analysis totaling in millions of tweets. This included data related to Marvel's Black Panther movie, Marvel's Captain Marvel movie, a 2020 Terrorist attack in Nice, France, the COVID-19 vaccine rollout, the U.S. 2020 Election 2020, the U.S. 2022 Election, and Ukraine-Russia in 2020. For each data set, I derived linguistic indicators for the maneuvers using the text-mining software, Netmapper [39], and calculated the BEND maneuvers using ORA-PRO. In this chapter, the data sets were processed using ORA-PRO version 3.0.9.154 and Netmapper version 1.0.0.94. Furthermore, all BEND calculations will use only hashtags as concepts (as explained in Section 1.2.2).

3.2 Humans Detecting BEND Maneuvers

This section discusses a process for facilitating how humans physically label the BEND maneuvers and an analysis of the ability of humans to label the maneuvers accurately. Two different types of annotators of different proficiencies were used to label the data. The first was a low to moderately-trained group of annotators, and the second was a pair of highly-trained annotators with more time and experience with working with BEND maneuvers. The results of this section will be later used to compare with the automated detection methods used in ORA-PRO as part of the detection validation process.

3.2.1 Using *Label Studio* for Annotations

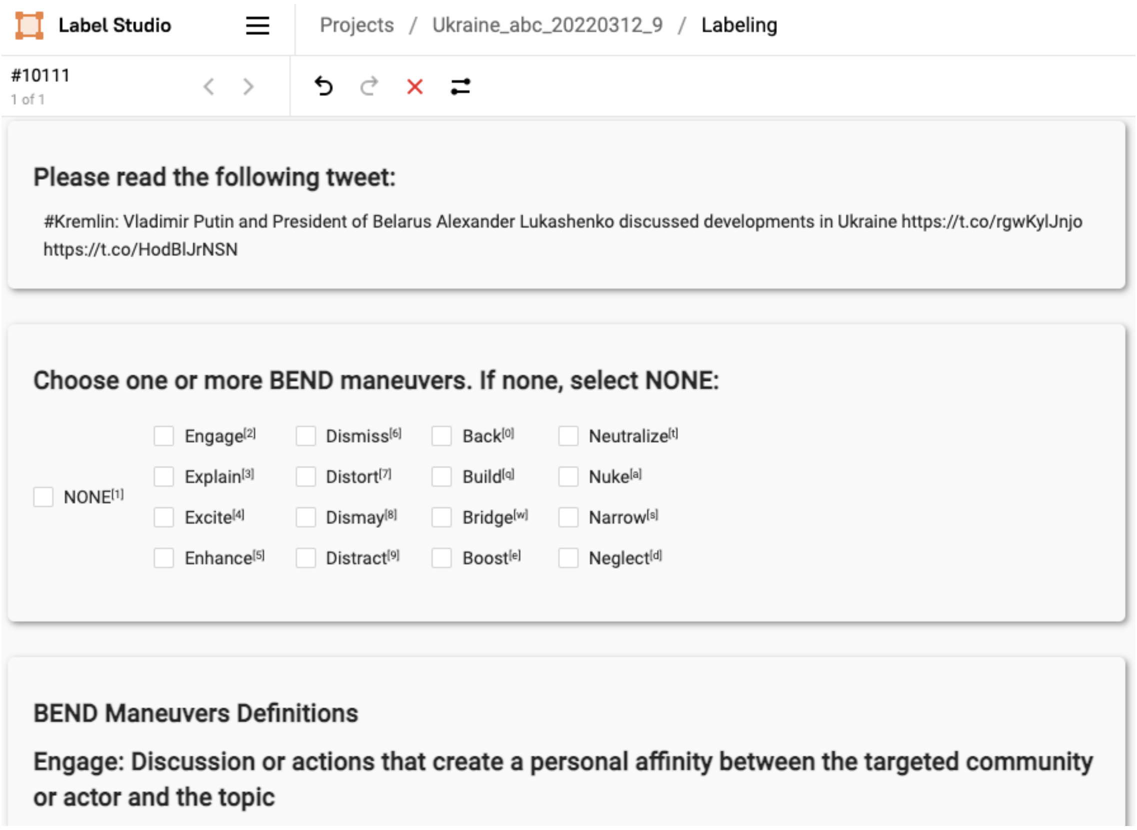
The labeling platform used for this work was the Label Studio - Community edition by Heartex, Inc. Label Studio [82] provides free and flexible software for annotating the BEND maneuvers and gathering results. Each data set was uploaded as projects of 100 tweets. The purpose of the division was to give annotators regular stopping points as they progressed through the labeling process. Users created accounts on label studio to identify them as annotators for the data sets. Then they selected the project of tweets they were assigned to annotate. An initial screen pops up to provide instructions with the definitions of the maneuvers. The definitions are also displayed throughout the process. The user is then shown a tweet and has the option to select one or more of the BEND maneuvers or NONE. An example of a tweet on the Label Studio interface is shown in Figure 3.1. If a user decides that they cannot finish a project in one sitting, they can stop and return to the project without losing their annotations. Additionally, if they feel that they may have selected an incorrect maneuver, they can return to the interface to correct the label.

3.2.2 Interrater Agreement

Eight low-moderately trained annotators participated in labeling portions of the data sets. Prior to labeling, each person received initial training on the BEND maneuvers and Label Studio. They were introduced to the BEND maneuvers and their functions within the social network, given several examples of each of the maneuvers and some of the indicators, shown how to use the labeling tool, and then given the opportunity to practice labeling before annotating. During the process, if they were unsure of selecting a response, they were allowed to ask for guidance.

1000 tweets were randomly selected from the Black Panther, Captain Marvel, COVID-19, Election 2020, Ukraine-Russia, and the French attack data sets and combined to create a diverse 6000 tweet data set for labeling. Each tweet would be labeled twice by a group of annotators.

The interrater agreement between annotators of the labeled data sets was calculated to show how consistently different people identify BEND maneuvers in text. The following analysis is from a subset of 1300 unique tweets, each labeled by two annotators. The samples of tweets come from the Black Panther, Captain Marvel, COVID-19, US Election 2020, and Ukraine-Russia data sets. A subset of tweets was used from the expected 6000 tweets because, at the time



Please read the following tweet:

#Kremlin: Vladimir Putin and President of Belarus Alexander Lukashenko discussed developments in Ukraine https://t.co/rgwKylJnjo https://t.co/HodBlJrNSN

Choose one or more BEND maneuvers. If none, select NONE:

☐ NONE[1]

☐ Engage[2]
☐ Explain[3]
☐ Excite[4]
☐ Enhance[5]
☐ Dismiss[6]
☐ Distort[7]
☐ Dismay[8]
☐ Distract[9]
☐ Back[0]
☐ Build[q]
☐ Bridge[w]
☐ Boost[e]
☐ Neutralize[t]
☐ Nuke[a]
☐ Narrow[s]
☐ Neglect[d]

BEND Maneuvers Definitions

Engage: Discussion or actions that create a personal affinity between the targeted community or actor and the topic

Figure 3.1: Label Studio interface for annotators

of the writing of this thesis, the process for manually labeling the maneuvers was ongoing for future uses of labeled BEND data. Volunteers worked at their own pace or occasionally needed to be replaced due to their inability to complete labeling.

All of the tweets were combined into data set with both annotator's labels. Cohen's kappa coefficient [44] was then calculated to determine the interrater agreement between two annotators in deciding the same maneuver for the same tweet. Values less than or equal to zero indicate no agreement and a one indicates perfect agreement. Based on the results shown in Figure 3.2, the agreement between annotators was generally weak, with values at or below 0.35. Of the 1300 tweets, annotators only completely agreed with every possible maneuver 109 times. The low agreement between annotators shows the unreliability of these annotators to detect the maneuvers.

The narrative maneuvers and the back and neutralize community maneuvers should hypothetically be easier to agree on as indications of these maneuvers are more evident in the text. However, this was not wholly the case with the data. Since most indications for narrative maneuvers occur within the text (as discussed in Chapter 2), it was not surprising that the narrative maneuvers predominately had higher kappa scores than community maneuvers. Among the top two maneuvers with the best agreement were excite and dismay. Annotators appeared more likely to agree with the annotation if they appeared to evoke emotions. Within many of these tweets with high agreement, the authors either mentioned their feelings regarding a particular

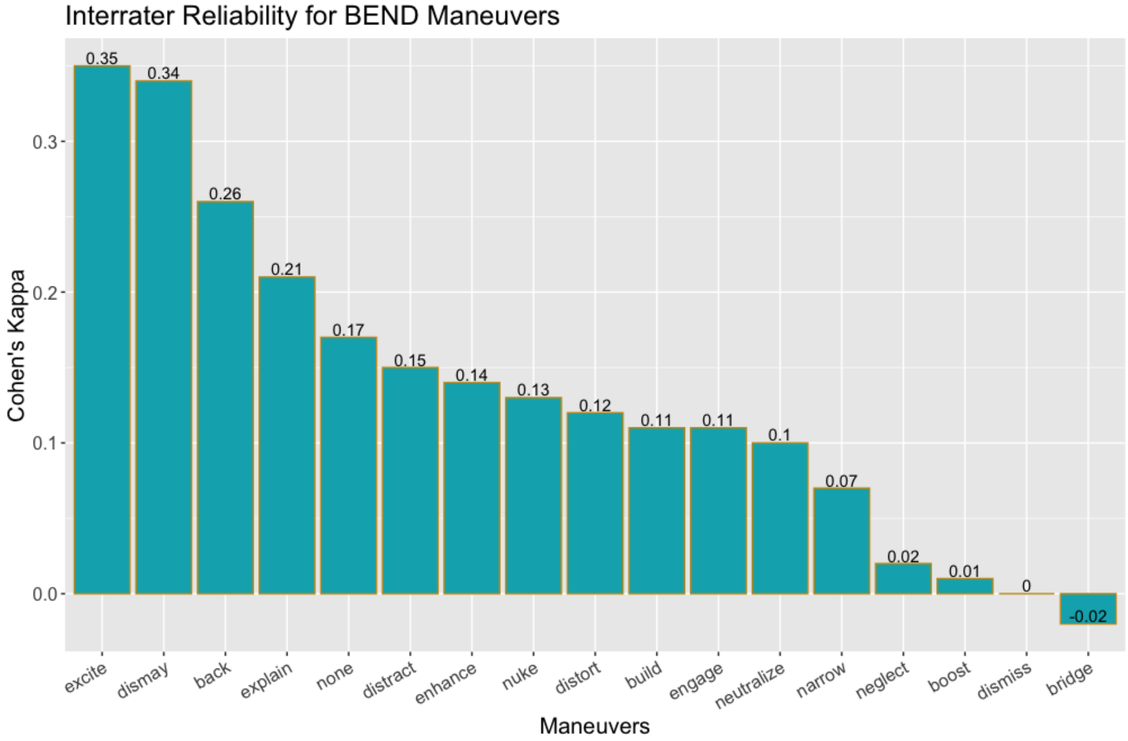


Figure 3.2: Interrater Reliability for BEND maneuvers between low-moderately trained annotators

topic or described emotional events. The annotators also had relatively high agreement for back maneuvers, which ranked third, but surprisingly, the opposite maneuver, neutralize, did not appear to have as high of a score. Particularly with the election data, annotators appeared to be able to identify when a candidate was discussed positively. Though not an emotional maneuver, explain maneuvers had the fourth highest value. Explain tweets typically contain an elaborate or detailed description of topics. For the annotators, some of the didactic explanations were easier to distinguish.

The community maneuvers appeared more challenging to label consistently. The maneuvers related to altering group sizes (all community maneuvers, except for back and neutralize) can be more subjective in differentiating which community maneuvers they may be. For example, a build maneuver creates a group, whereas a boost maneuver increases the size of a group. Depending on the message and the particular group, there can be inconsistencies between which are the correct maneuver.

3.2.3 Highly-Trained Annotator Interrater Reliability

Furthermore, the interrater reliability between two highly-trained annotators. The annotators were given 160 tweets sampled from the same data sets as the annotators discussed in the previous section. The purpose was to determine if annotators with more training were more consistent labelers. The results for the highly trained annotators are shown in Figure 3.3. They varied in their ability to agree on detecting both community and narrative maneuvers with values ranging from .53 to -.01. This is moderate or low agreement depending on the maneuver. Explain, neutralize, excite, and narrow were among the top maneuvers, whereas dismiss, distract, engage, and neglect maneuvers were among the bottom.

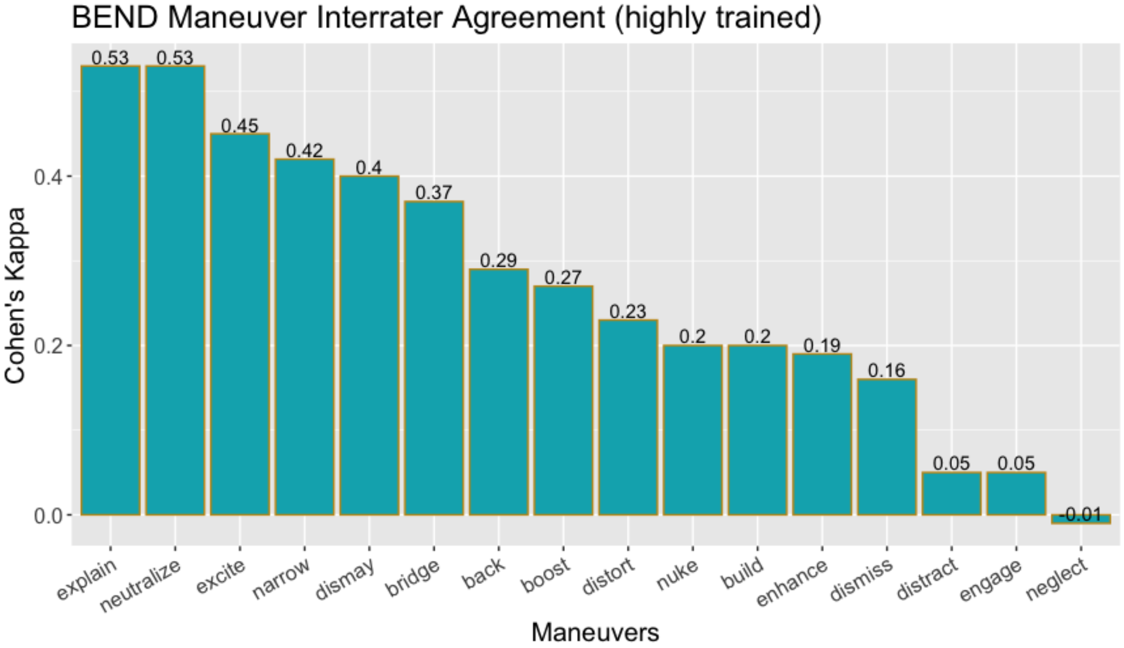


Figure 3.3: Interrater Reliability for BEND maneuvers between two highly trained annotators

The highly-trained annotators overall had higher agreement across all maneuvers than the low-moderate annotators. Figure 3.4 compares the agreement values for the two types of annotators for each of the maneuvers. Large differences in values between the annotators are observed in explain, neutralize, narrow, and bridge. However, a third of the maneuvers were still valued with low interrater agreement at less than .2. These were the build, enhance, dismiss, distract, engage, and neglect maneuvers. Annotators had difficulty in understanding what denotes each of these maneuvers within the text.

3.2.4 Conclusions and Limitations Regarding Human Annotation

There are several possible explanations for the low reliability between annotators for maneuvers. One reason for the different interpretations of the maneuvers is the result of annotators having different personal backgrounds. Some people are more susceptible to disinformation, which affects their ability to identify disinformation or distort messages [6]. Others may just be unfamiliar with the topics of the data sets, making them unable to judge accurately how the topics relate to the different maneuvers.

Some of the low-reliability values may have resulted from the implementation of the annotation process. Several adjustments may improve the interrater reliability between annotators. This

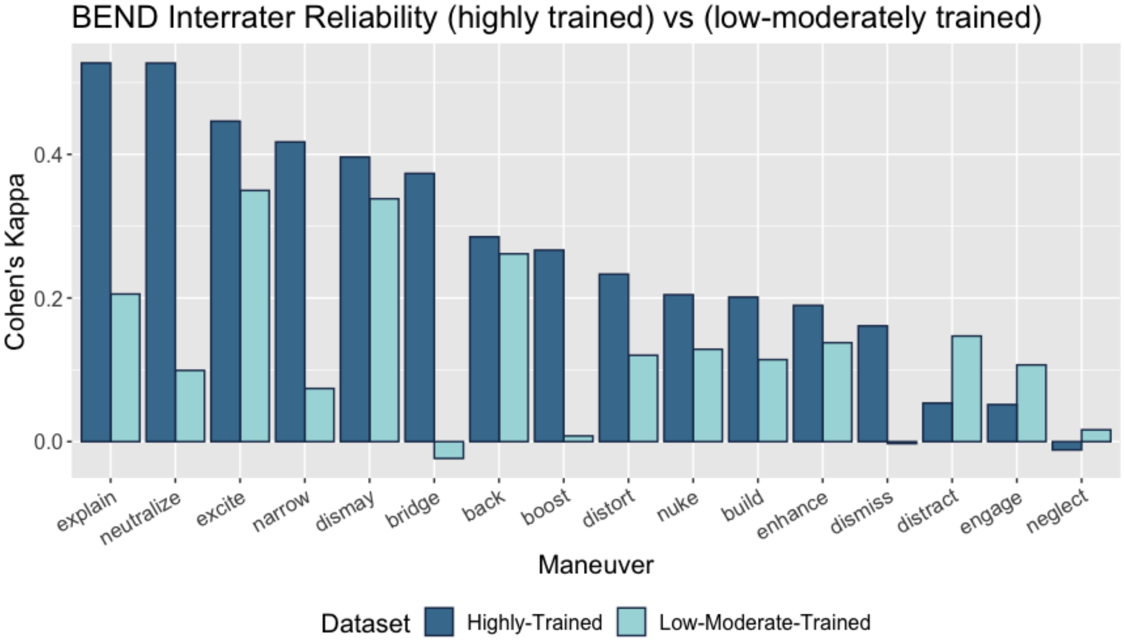


Figure 3.4: Comparison of Interrater Reliability for BEND maneuvers between two highly-trained annotators and two low-moderately trained annotators.

includes more comprehensive training or tasking the annotators to label from a shorter list of maneuvers, reducing having to evaluate all maneuvers simultaneously for every tweet. Focusing on one maneuver at a time may reduce the number of errors between annotators.

Regardless, the disagreement between labels shown in the results suggests that people unaided by computers may have trouble identifying the BEND maneuvers. Therefore, to ensure that consistently reliable results are fed into operational assessments, improving computationally detecting the maneuvers is important.

3.3 Cognitive/Emotional CUES

In this thesis, Cognitive/Emotional Cyber-mediated Usable Emotional Sensors (CUES) are the lead indicators for detecting the BEND maneuvers and extracted from messages using the Netmapper software [39]. CUES are subconscious indicators found within text-based messages, such as tweets, that can signal or influence a person's cognitive and emotional state [23, 77, 104]. While traditional machine learning algorithms could use trained data sets to identify general sentiment, the CUES are able to more accurately connect the sentiment and emotion to topics and concepts [35].

The implementation of identifying and detecting CUES uses a strategy of counting words, expressions, and other text-based linguistic variables. The method is based on techniques used in the employment of Linguistic Inquiry and Word Count (LIWC) [106, 107, 124], which similarly links word use with real-world behaviors and mental states. The process involves a library or thesauri of words, expressions, concepts, grammatical units, etc. that can be extracted from a text message. These are then categorized into hierarchical dimensions, or a ConceptTo. For example, the terms “joy”, “cheery”, and “exhilarated” would be considered an emotion, an emotion with value of positive affect, and a happy CUE. However, using lemmatization, joy can also be derived from “joyous,” “joyful,” and “joyously.”

The use of individual CUES is founded on extensive psychological research. Many of the CUES affect emotions or indicate an author’s emotional state. The addition of emoticons and emojis are typographical symbols that represent facial expressions. Users tend to use them to express feelings about particular content [49]. An example could be the addition of a smiley face at the end of a sentence to show support or approval, or it could be a sad or angry face to show their frustration [109]. They have been shown to support the emotional tone of an author’s message [135] and to even evoke stronger emotions from their recipients compared to messages without them [56]. The text of a message can influence emotions such as anger and happiness [40]. Some words may indicate stress [102], while homophones, words that sound like other words, such as *guilt vs. gilt* and *liar vs. lyre* can be related to anxiety states [85, 89]. Furthermore, non-words such as use of the like button can indicate approval, support, or that a post was “informative, useful, funny, exceptional, creative, or otherwise important and interesting ” or that a person is aligned with a particular topic or belief [53, 109, 123]. Unrelated to the actual words, studies have also shown that long informal text can have an effect on emotional responses [100].

Other types of indicators unrelated to the characters and words of the message include the effects of long informal text on emotion[100] and how pronouns can also indicate levels of honesty, thought processes, and relationships [103]. Punctuation such as question marks and exclamation points can also indicate an author’s mindset or potentially influence a reader [109, 124].

The thesaurus of CUES and their application for detecting the BEND maneuvers are constantly being updated based on updated literature and the improved understanding of the cognitive and emotional environment. The full list and a brief description of the CUES with examples of some of the BEND maneuvers they detect are found in Appendix B.

3.4 Man vs Machine: Human-Labeled vs Automated Maneuver Detection

In this section, the results from ORA-PRO and the human-labeled data are compared to evaluate the differences between both methods of identifying the BEND maneuvers.

A subset from the Black Panther, Captain Marvel, COVID-19 vaccine, Election 2020, and French data sets were used. One of seven annotators labeled 2999 tweets, and afterward, ORA-PRO was used to calculate the BEND maneuvers using the mean as the threshold.

Like the method used for determining interrater reliability in the previous section, Cohen’s kappa coefficient was the metric for measuring the agreement between humans detecting maneuvers versus an automated machine. Figure 3.5 shows the poor reliability between both sources of detection. The largest values were 0.15 for dismay, 0.13 for back, and 0.1 for excite. Though these maneuvers have a weak agreement, they were also the top three maneuvers of agreement in the inter-reliability calculations between two humans, as reported in the previous section. This suggests that the metrics for these maneuvers are detecting these maneuvers at least to a small degree. Interestingly, while the excite maneuver ranked fourth in reliability between humans, the kappa score between human and machine ranked last and was largely negative. This represents a large disagreement between the two methods of detection. Therefore, the metrics for the excite maneuver should be reevaluated to determine if they are the appropriate indicators.

Overall, given the weak agreement values, the metrics for all of the maneuvers should be reexamined to improve detection. However, this is not to say that humans are the authority on which maneuvers are occurring, as they have been shown to greatly disagree as well. Both systems have large inefficiencies that make it difficult to accurately and reliably identify the maneuvers. Therefore, more should be done to improve in both training humans to be more consistent in labeling and identifying the appropriate indicators for automated detection. Highly trained humans can inform improved detection methods for the behaviors, and accurate automated detection can uncover human error.

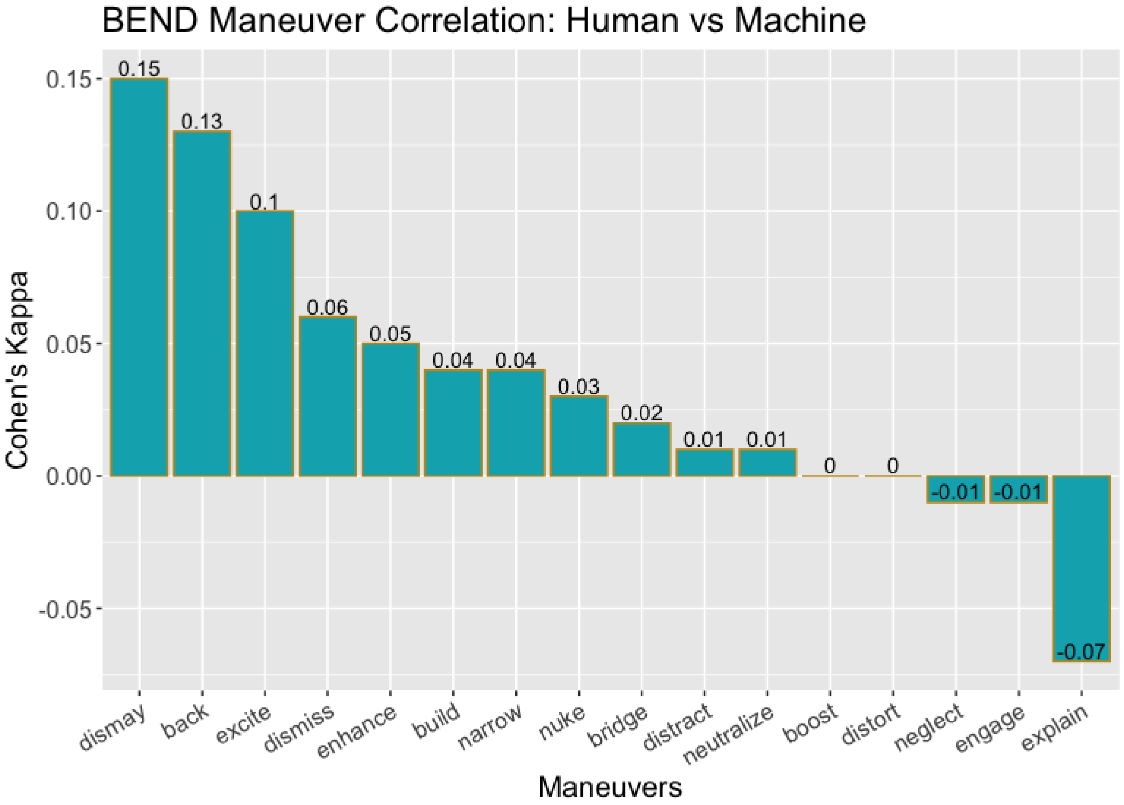


Figure 3.5: Calculations for agreements between low-moderately-trained human annotators and automated detection of BEND maneuvers (NM 94, ORA-PRO 154)

3.5 Refining Automated Detection

The massive size of social media data requires an efficient method for stepping through the network of users and tweets to gain insights on what BEND maneuvers are occurring. For many of the data sets, annotators spent approximately one hour labeling 100 tweets. For data sets that extend to tens of millions of tweets, this is a nearly impossible task for analysts with limited time to conduct their investigations. Additionally, human detection simply based on social media searches is limited as analysts may only gain a fragmented understanding of what is happening during an influence campaign. Only with automation can analysts gain a holistic view backed by quantifiable results. Furthermore, as previously discussed, humans cannot reliably identify maneuvers within messages. Therefore, it is necessary to develop a robust detection method that allows the ability to filter the data quickly using consistent metrics. This facilitates the analysis of large amounts of data, comparing data sets, or examining data over time.

The current method for the automated detection of maneuvers takes the inputs of CUES, various network measures, and other forms of content (e.g., concepts, URLs, mentions, etc.) as inputs and uses the calculations to evaluate the degree of indication (DOI) value from 0 to 1 for how likely a message is executing a maneuver. The threshold for the DOI value is determined by calculating all of the DOI values for all tweets with a value greater than zero for a particular maneuver. These DOI values are averaged, and any tweet with a value greater or equal to the mean is considered the maneuver. The tweets are then counted and combined to report the number of times each maneuver occurs in the data set. See Figure 3.6. Based on the results from several applications on multiple data sets, there is room to refine the BEND maneuver detection.

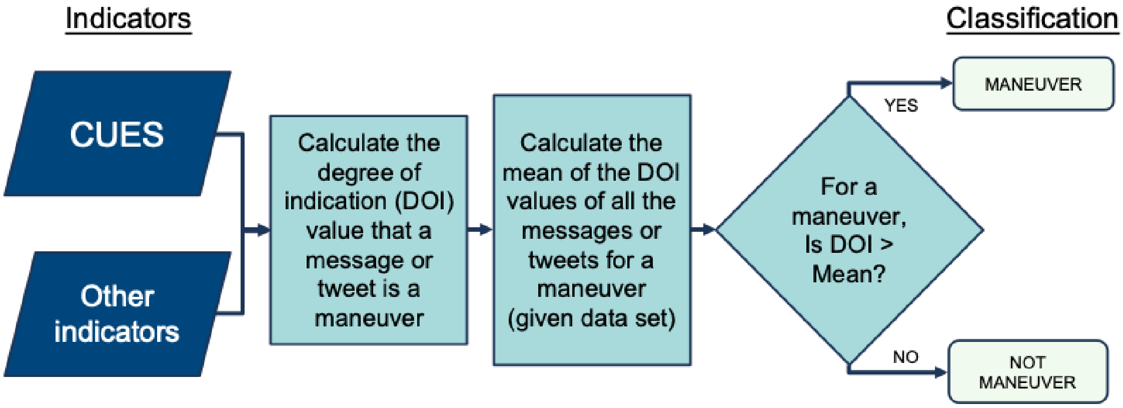

Indicators
Classification
CUES
Other indicators
Calculate the degree of indication (DOI) value that a message or tweet is a maneuver
Calculate the mean of the DOI values of all the messages or tweets for a maneuver (given data set)
For a maneuver, Is DOI > Mean?
YES
MANEUVER
NO
NOT MANEUVER

Figure 3.6: Current method for detecting BEND maneuvers. Indicators are used as inputs for calculating degrees of indication. A message is classified as a maneuver if the DOI of the message is greater than the mean of the DOIs for a maneuver for all the messages within the given data set.

3.5.1 Methodology for Refinement

The methodology for improving detection is an iterative process. The first step involves calculating summary statistics for the BEND maneuver DOI values based on the original implementation (the mean threshold). This is followed by applying statistical analysis to evaluate the CUES and current and potential thresholds. The final step is to examine the BEND maneuver results for accuracy by manually comparing results with the text. This process is repeated with the improved values until the optimized results are achieved. In this section, multiple references to the excite maneuver and its CUES will be used to illustrate the refinement process.

3.5.2 Evaluating CUES Weights

Evaluating the CUES begins with determining the correlations between CUES commonly used together for each of the maneuvers. The relationship between the CUES can potentially determine if some CUES should be weighted differently. A high correlation between two CUES may suggest that either one CUE is irrelevant or that both CUES should contribute less to the overall DOI value.

Using the excite maneuver as an example for evaluating the CUES weights, the Pearson correlation coefficient values between the CUES are calculated. The results in Figure 3.7 show the high correlations between the use of positive sarcasm and happy emoticons, positive sarcasm and positive emojis, and positive emojis and happy emoticons. Taking a closer look at the methods for identifying positive sarcasm, this calculation uses positive emoticons and emojis as inputs. Therefore, recommendations for improvement in this metric would be to remove the weight for positive emoticons or emojis if positive sarcasm is detected in the message. This will remove the redundancy of double counting positive emoticons or emojis in the DOI value calculations. Furthermore, positive emojis and happy emoticons serve similar purposes for adding the author's feelings in an excite maneuver. Therefore, if used in the same tweet, having both metrics as equally contributing values to the DOI value calculations is also a redundancy. The two CUES can be combined, if either CUES appears in a message, the existence of either would be singly weighted.

Appendix C shows the correlations for all of the remaining maneuvers.

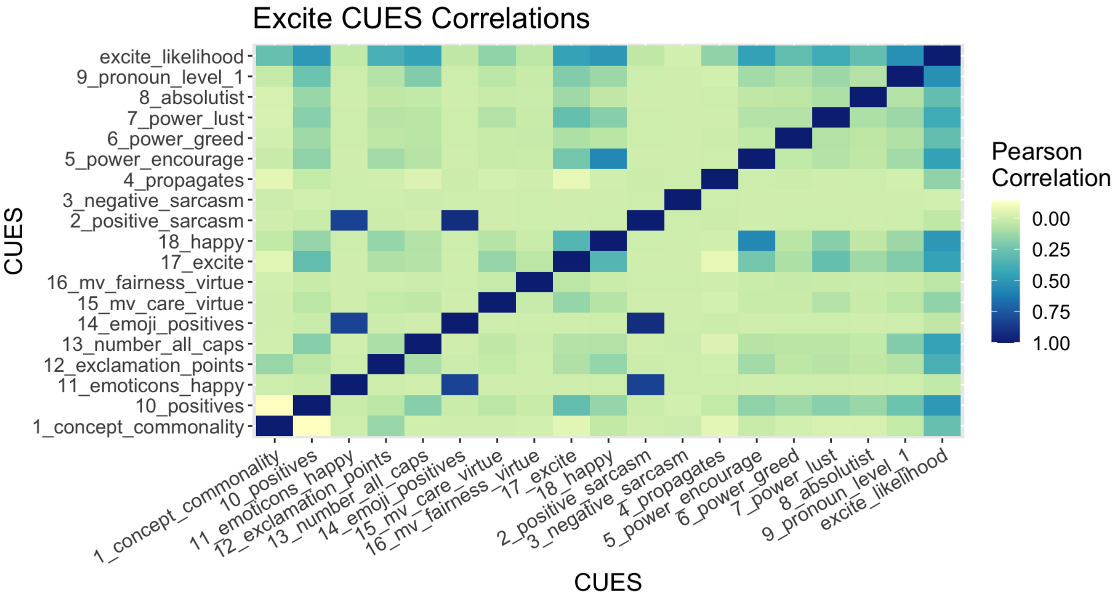


Figure 3.7: Correlations between *excite* CUES. A higher correlation indicates the possibility that pairs of CUE values should be weighted differently.

3.5.3 Evaluating Threshold Values

The threshold value for detection is the value where any document with a DOI greater than this value is considered to have a particular maneuver. To explore potential thresholds for detection, the summary of statistics and the distribution of DOI values for each of the maneuvers was calculated as shown in Table 3.1 using all of the data sets for this chapter. This differs from the original method of calculating the thresholds, which calculate thresholds based on the data set under analysis. Figure 3.8 shows the distribution of the DOI values for all of the maneuvers across all of the data sets in this study revealing non-normal behavior for many of the maneuvers. For some, such as neutralize and back, the 3rd quartile is the same as the minimum, and the median is the same as the 1st quartile and the maximum value. Values for neglect and narrow primarily occur at zero, and for engage, they primarily occur at one. Furthermore, distort and bridge have medians occurring at zero. These non-normal values are caused by low occurring maneuvers or maneuvers where the values tend to be binary, zero and one. The maneuvers with these idiosyncrasies cause issues with using certain statistics as threshold values when the values are in the extremes. For example, if using the median value as a threshold for enhance or dismiss, any message with a non-zero DOI would be considered the maneuver.

Therefore, a multi-tiered approach for deciding thresholds is recommended over the current method, which consists of labeling values greater than the mean as maneuvers.

The first method addresses maneuvers that have bimodal distributions, where values tend to be near 0 or 1. This applies to the back and neutralize maneuvers. Using the median is not appropriate as a threshold because this may result in the entire data set either being classified or not classified as a maneuver, depending on which side of the distribution the median lies. For maneuvers with these types of distributions, using a threshold of .5 DOI suffices.

The second method addresses maneuvers with normal or non-binary distributions. Though the use of either mean or median would suffice in most situations, the median would be more useful to address any skewing from outliers within the data. These include the boost, bridge, build, excite, explain, nuke, dismay, and distract maneuvers.

The third method applies to maneuvers that are extremely right-skewed in that the median value is equal to zero. This indicates that these maneuvers are detected in low quantities throughout the data sets. Consideration must be made to decide whether or not this is reflective of the actual presence of the maneuvers or if the maneuvers require better detection to assign the appropriate threshold. The maneuvers in this category include enhance, narrow, neglect, dismiss, and distort. Until then, the maneuvers should be detected using non-zero values as the threshold. This will detect slight indications of these low-occurring maneuvers so that they can be iterated through the detection refinement process.

A final exception can be made for the engage maneuver. Unlike the other maneuvers, this distribution is extremely skewed to the left where the median value equals one. This indicates a large number of maneuvers with the maximum DOI value. Therefore, if using the median as a threshold, any value that is not one would not be detected. In this case, using the mean as the threshold value is an alternate threshold for producing results that are not all or nothing. However, consideration should be made to determine if the indicators for the engage maneuver or too inclusive resulting in higher than actual detection of maneuvers. Improved indicators may form a distribution that suggests a different threshold.

The summary of these recommendations for the cutoff thresholds is shown in Table 3.2.

Table 3.1: Summary of Statistics of BEND maneuver Degree of Indication values (N=38383667)

metric	back	boost	bridge	build	engage	enhance	excite	explain
Max	1.00	0.96	0.92	0.89	1.00	1.00	0.86	0.73
Mean	0.57	0.19	0.10	0.29	0.86	0.06	0.21	0.28
Median	1.00	0.15	0.00	0.22	1.00	0.00	0.20	0.27
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q1	0.00	0.11	0.00	0.17	1.00	0.00	0.12	0.20
Q3	1.00	0.26	0.25	0.50	1.00	0.00	0.29	0.33
StdDev	0.49	0.13	0.15	0.19	0.32	0.12	0.11	0.10

metric	narrow	neglect	neutralize	nuke	dismay	dismiss	distort	distract
Max	0.93	1.00	1.00	0.64	0.77	0.67	0.78	0.78
Mean	0.04	0.00	0.70	0.07	0.11	0.06	0.05	0.14
Median	0.00	0.00	1.00	0.08	0.11	0.00	0.00	0.17
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Q1	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.08
Q3	0.00	0.00	1.00	0.08	0.16	0.00	0.06	0.17
StdDev	0.09	0.03	0.46	0.08	0.08	0.13	0.09	0.09

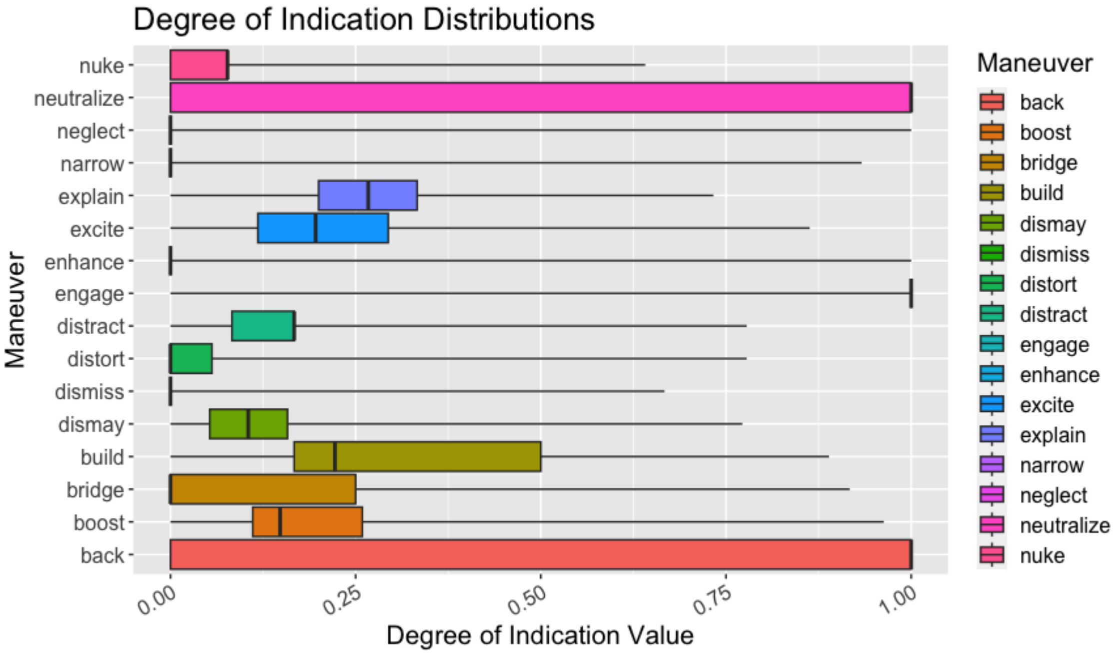


Figure 3.8: Distribution of All Maneuver Degree of Indication Values Across All Datasets

Table 3.2: Threshold recommendations (Degree of Indication (DOI) values) based on the distribution of DOI values for each maneuver from a large and diverse data set

Maneuver	Distribution	Recommended Threshold
back	bimodal	>.5
boost	near-normal	median
bridge	slight right skew	median
build	near-normal	median
engage	extreme left skew	mean
enhance	extreme right skew	>0
excite	near-normal	median
explain	near-normal	median
narrow	extreme right skew	>0
neglect	extreme right skew	>0
neutralize	bimodal	>.5
nuke	slight right skew	median
dismay	near-normal	median
dismiss	extreme right skew	>0
distort	slight right skew	>0
distract	slight right skew	median

3.5.4 Additional Considerations for Detection

Further aspects of the refinement process should be considered for determining CUES weights and thresholds.

First, this process requires a deeper inspection of the use of CUES in detection beyond a statistical analysis approach. Some CUES may still need to be weighted more not because of the correlations but because they are generally good indicators for the maneuver. Also, some CUES combinations may be better as indicators for a maneuver. For example, if examining a list of human-annotated excite maneuvers with high interrater reliability, we find that every time an exclamation point exists with positive and happy terms, the tweet is labeled as an excite maneuver. This becomes an automatic detection for the maneuver increasing the likelihood of a tweet being this maneuver to near 100 percent.

When considering thresholds, the values must be dependent on a baseline rather than the data set under analysis. Using mean or median values for thresholds based on the working data set creates an unnecessary bias for maneuvers that occur in higher or lower than normal quantities.

Furthermore, current detection methods fail to detect the maneuvers within the context of a narrative. Particularly for maneuvers used for countering narratives or expanding on ongoing narratives, some maneuvers may not be considered maneuvers unless placed in the context of other related messages. The methods poorly detect irony, sarcasm, and idioms, all of which require context in order to accurately observe the maneuvers in an overarching influence campaign.

3.5.5 Look How Far We've Come

To evaluate the progress made during the refinement process, comparisons were made between both types of annotators and an older and newer version of ORA. Figures 3.9 and 3.10 show the interrater reliability between the low-moderately trained annotator and the highly-trained annotator each against an older version of Netmapper and ORA-PRO (Netmapper v80, ORA-PRO v142). The highly-trained subject matter expert remained relatively the same except in the back maneuvers, where the expert had increased reliability for back, bridge, and distort maneuvers but had a larger decrease in reliability for explain maneuvers. The low-moderately trained annotators similarly saw higher reliability increases for back but also saw a larger change for build and neutralize maneuvers, though the kappa value was still relatively low (below .1). Furthermore, they also saw a larger decrease in reliability for explain maneuvers.

These results indicate that improvements have been made for at least the back maneuvers. However, the large decrease in reliability for the explain maneuvers may suggest that either detection is decreasing in effectiveness for this maneuver or that humans are increasingly poor at detecting this maneuver. The indicators for this maneuver need to be reevaluated.

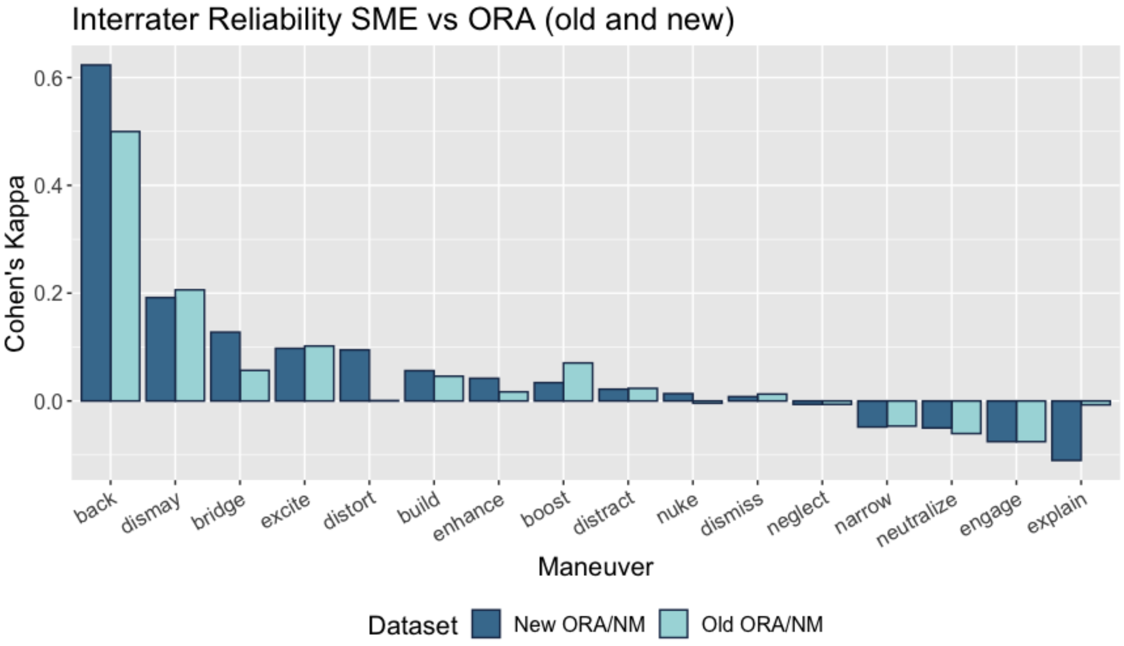


Figure 3.9: Calculations for agreements between the SME and an older and more recent version of ORA and Netmapper.

3.5.6 Limitations and Future Work for Framing Tweets for Detecting Maneuvers

The standalone message of a tweet does not contain all of the information required to make a complete assessment of the BEND maneuvers within the message. This applies to both human

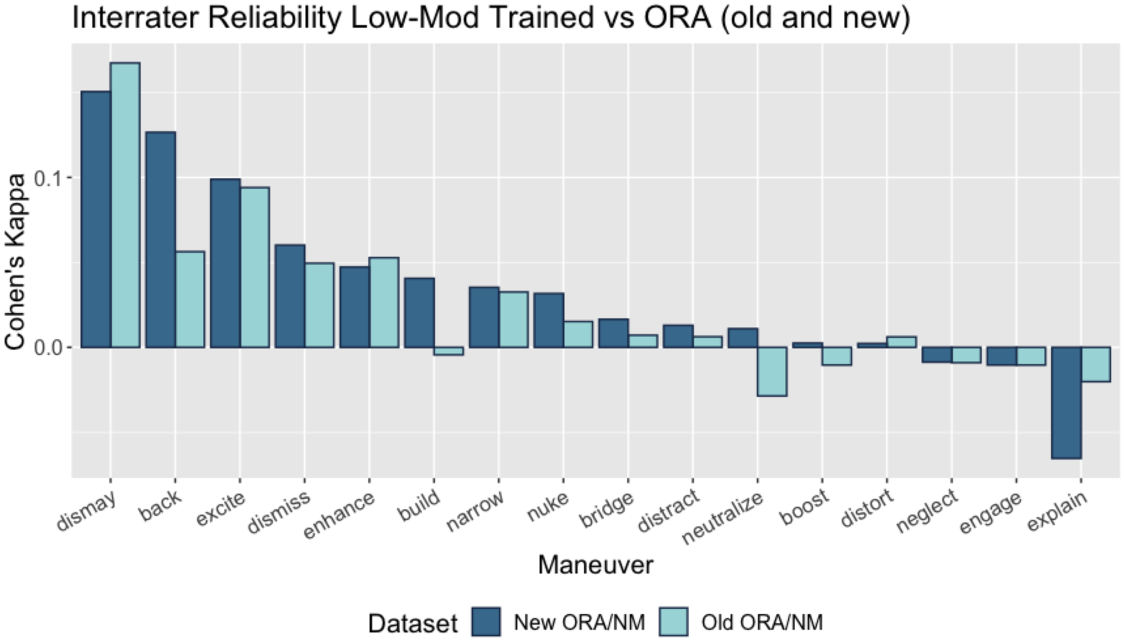


Figure 3.10: Calculations for agreements between the low-mod trained annotators and an older and more recent version of ORA and Netmapper.

and the current machine classifications. First, the tweet is not placed within the context of the overarching narratives. Therefore, maneuvers that may be a result of multiple tweets may not be properly interpreted. Additionally, the structure of the tweet message as derived from the MESSAGE field from the Twitter API does not indicate key aspects of the tweet. For replies and quotes, it does not provide the entirety of the original message. These tweets are not viewed from a conversational point of view making some tweets appear meaningless or misunderstood due to context, humor, or sarcasm. Future work should incorporate the coding of data that includes the whole conversation.

Additionally, given a tweet, sometimes human labelers may have difficulty deciding on whether or not a maneuver is present. One method for approaching the occurrence of a maneuver is in the confidence that an annotator believes the maneuver exists. This can be in the form of *none*, *low*, *medium*, and *high*. Therefore, results may show that more maneuvers match with the machine, but they only do because they fall within a certain confidence level. They may also show higher interrater reliability between the annotators. This would facilitate the training process and provide more informed results that are based on gradient versus binary view.

For automated detection, a percent likelihood value (different from the degree of indication) would be useful, but cutoff thresholds should still exist to have definitive answers or the existence of a maneuver. The difficulty is in determining a percentage likelihood value as some CUES can be binary or weighted unevenly.

3.5.7 When is it Good Enough?

Identifying when the metrics can satisfactorily detect a maneuver requires verification of refined automated results using an expert understanding of detecting the maneuvers. The process involves an interrater reliability comparison of data sets labeled by subject matter experts with a consistently high agreement and the automated detection output. Though with extensive training, humans may not be as accurate as a machine when it comes to labeling the maneuvers, a trained annotator with an understanding of the indicators can manually adjudicate discrepancies between the labeled and the automated data. Therefore, to account for human error, subject matter expert sampling and manually checking the assignment by ORA-PRO must be incorporated into the annotation process to ensure the most accurate annotations. The trained annotator should be able to decide whether or not the discrepancy is based on human error or machine error. Once the confidence level of accuracy is acceptable, the automated detection for a maneuver can be validated.

3.6 General Recommendations for Improvement

The list of CUES and indicators is expansive. As indicators for the maneuvers, each maneuver requires different CUES, and the likelihood of individual CUES varies from maneuver to maneuver. Following the latest iteration of the detection refinement process, several recommendations should be made to the current implementation of CUES regarding the necessity of certain CUES in addition to those made earlier in this chapter for the excite maneuver. The recommendations are based on observations of the CUES of other maneuvers.

Emotion related CUES need to be reevaluated. The relationship between positive emojis, happy emoticons, and positive sarcasm was discussed. However, some degree of overlap may also occur for other CUES such as between happy and power encourage, and for negative emotions, the CUES for anger and power anger have high correlations. Similar to the findings for positive sarcasm, negative sarcasm overlaps with sad emoticons, angry emoticons, and negative emojis.

For affecting leaders, a single agent reference does not appear to have any effect on the degree of indication for a back and neutralize maneuver. This may be because in many instances, these messages with these maneuvers that have at least one reference contain multiple references for either countering an opposing leader or tagging others in a community growing/reducing maneuver.

For many of the maneuvers, the maneuver terms that relate terms specifically for a maneuver (e.g., ”Excite_17_Excite”, as depicted in Figure 3.7) are not always correlated with the degree of indication (e.g., ”Excite_likelihood”, as depicted in Figure 3.7). This means that there is a disconnect between the terms used to indicate the maneuvers and the aggregation of all of the indicators use to indicate the maneuver. When it comes to Neglect CUES, the neglect related CUES are the only indicator. Therefore, these need to be explored to ensure that they are sufficient enough to detect this maneuver, especially given the low detection rates for this maneuver within the data sets.

While some CUES still need to be re-weighted, other indicators could be added to help improve detection. One example is the inclusion of verb tense. Verb tense has shown to inform the temporal focus of attention for various topics [124]. Gunsch et al. [61] discuss the different use of tense for positive and negative political ads. This could possibly be used as an indication of dismay or distort messages.

3.7 Developing a Baseline

Analysts require baseline values to understand what is normal and what is significantly different from the results of their calculations. Though the calculations will return which messages in the data set fall under each maneuver, it is hard to know if these results are meaningful without a baseline to compare them to.

In this section, the percentage that each maneuver is used was calculated from a large and diverse combined data set. The percentages were calculated using two threshold values for comparison. The thresholds are for the percentage of maneuvers occurring that are above the mean or the median. The mean is the current threshold for determining the maneuvers. The median values were calculated to evaluate a possible alternative for identifying a cutoff for what DOI value constitutes a maneuver. Figure 3.11 shows the baseline values for the maneuvers detected using the mean and median threshold values.

There are several notable features of the calculated baselines. One noticeable observation is that the number of neglect maneuvers for both thresholds observed in the overall data set is at approximately 1%. This extremely low value suggests either that the indicators for identifying the neglect maneuver are not sufficient or that neglect maneuvers are not used in high frequencies. Though it is unknown as to what value constitutes a need for better metrics versus what is an appropriate baseline value, there are too few documents indicating actors attempting to decrease the size of an opposing community. Regardless, more research needs to be conducted for this maneuver.

When observing the mean threshold, one thing to note is the high occurrence of engage maneuvers. This may appear highly abnormal, but the purpose of this maneuver is to create personal connections with the target audience and a particular topic. This particular action occurs naturally in conversations. It is possible that the indicators for this maneuver are prevalent throughout Twitter creating high or binarized values, making it difficult to use thresholds. This is apparent as using the median threshold results in zero engage maneuvers within the data set where the max, median, and 3rd quartile are all one. Therefore, maybe a more practical approach when observing an engage maneuver is to explore the actual messages being sent and how they are used in conjunction with other maneuvers to achieve a desired objective.

Though the events for each data set were characteristically different, one limitation of this methodology is that the data sets were collected surrounding notable real-world events and are not a purely random sample of Twitter. There is a chance that the types of data sets used may contain an underlying percentage of a particular maneuver simply from the nature of the tweets being sent under certain circumstances.

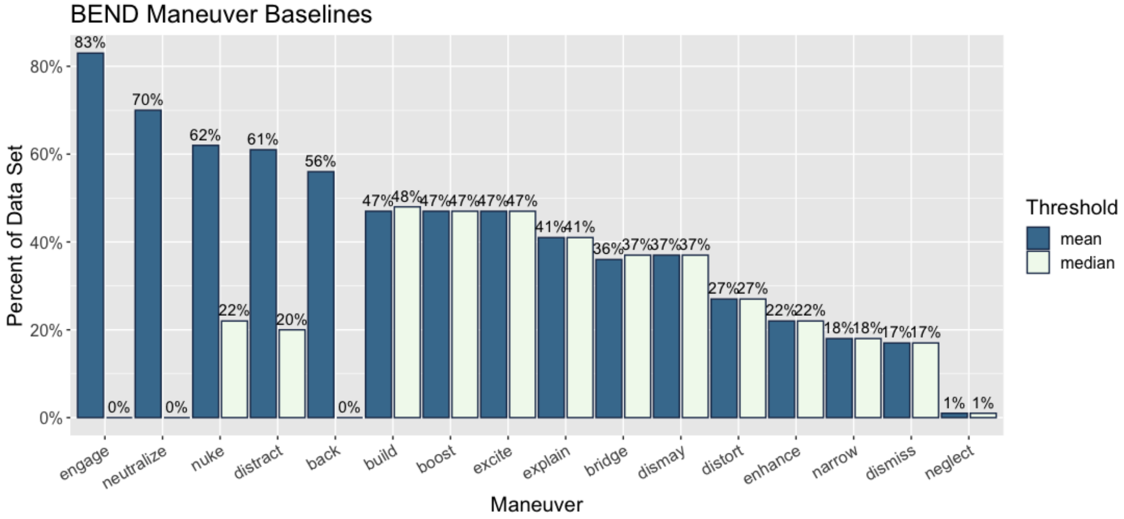


Figure 3.11: BEND Maneuver Baselines. Shown are the percentage of the data sets that contain each of the maneuvers based on the use of mean or median degrees of indication values for cutoff threshold values.

3.8 What Maneuvers are Commonly Seen Together, or Not?

Identifying which maneuvers are commonly used together can show some of the types of fundamental actions occurring concurrently during influence operations. The analysis in this section helped inform some of the techniques mentioned in the previous chapter regarding the BEND tactical categories.

Figure 3.12 shows the Pearson Correlation of maneuvers that are used in the same messages. Bridge, boost, and build maneuvers appear to be correlated. This can be because the maneuvers attempt to achieve similar objectives related to increasing the size of communities. This is contrasted with the nuke, neutralize, and narrow maneuvers, which attempt to decrease group size. Even if they are not exactly the same thing, just like the growing group maneuvers, the fundamentals of the maneuvers attempt to achieve similar goals. Therefore, it is unexpected that they did not have the same higher correlations as did the ‘B’ maneuvers.

Back and Neutralize also have a high correlation. This might result from the push-pull of backing one actor and neutralizing the other, but this can possibly be from similar non-CUES metrics. The non-CUES indicators include that a tweet is a ‘derived document’, such as a retweet, reply, or quote, and references (@mentions) exactly one agent. Also correlated with the back and neutralize maneuvers are engage maneuvers. As engage is intended to bring the target audience personally closer to a particular topic, this is a useful maneuver to create a positive or negative emotional relationship between an opinion leader and the targeted audience.

Another interesting observation of this data is the correlation between dismay, distract, and nuke. This shows how authors see it as necessary to include distract and dismay techniques within their messages aimed at dismantling a community.

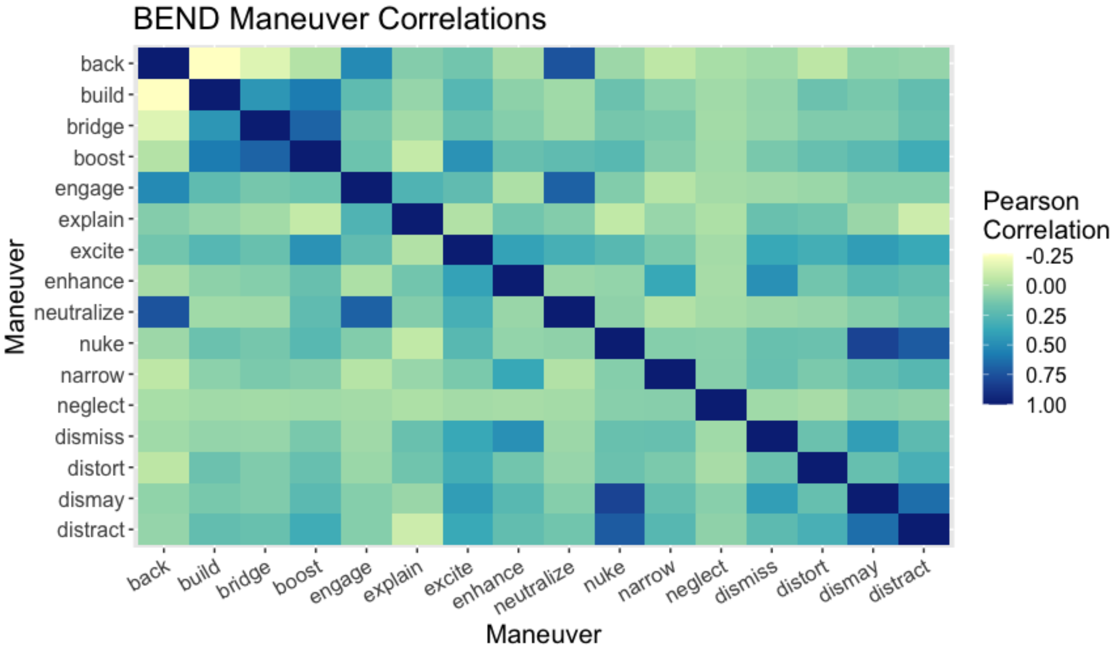


Figure 3.12: BEND Maneuvers Correlations. Darker correlation pairs indicate which maneuvers commonly co-occur within the same message.

3.9 Conclusion

This chapter described a need for accurately detecting BEND maneuvers, developed an iterative process for improving the current approach, developed a baseline for analysts to use, and examined the combination uses of the maneuvers. The methods involved developing and analyzing human-annotated data and conducting a statistical analysis of the BEND maneuvers and their metrics. While recommendations were discussed to improve the detection, further iterations along with the considerations discussed are required for refining the process.

Improving detection and creating a baseline necessitates large amounts of diverse data sets. Based on the current distribution of results, a multi-tiered approach was used for determining the thresholds of the different maneuvers. As the nature of discussions on social media evolves over time, the baseline must be reevaluated and recalculated to account for the change. This is essential for analysts to determine whether or not their results are significant.

Finally, computational detection proves vital as humans are imperfect at consistently distinguishing maneuvers, while also being extremely slow. Though both human and machine methods have limitations, both are required to improve the accuracy of automated detection. Better detection contributes to the more extensive process of better understanding the fundamental actions of influence on social media and analyzing online influence operations.

Chapter 4

How Do We Analyze Online Influence? *The BEND Framework*

4.1 Introduction

Leaders and decision-makers are keenly aware that social media is an expansive aspect of the cyber domain, where online information operations motivate and influence both behavior and large-scale operations in the real world. Users employ basic functions of social media platforms to strategically achieve their desired objectives within the online informational environment strategically. A basic understanding of the on-goings on social media can be derived from trending articles, an analyst's news feed, or what the mainstream media reports about what appears to be popular. However, these methods of achieving situational understanding are often spotty, biased, and unreliable. Various organizations have created methods of analyzing influence campaigns and disinformation, but these are typically confined to broad methods with a qualitative review of online content. Leaders often require a more systematic and metrics-based approach for discerning the extent of how influence tactics and related behaviors are transpiring on specific online platforms. Ultimately, they desire to determine the impacts of such online behaviors and any information to help them develop their response.

This chapter operationalizes the BEND framework to provide leaders with qualitative and quantitative analysis and assessment regarding online information operations. It begins with how the methodology was developed, the objectives of the analysis, and the pipeline and overall workflow for the framework. Then the chapter details the different components of the process and discusses how to develop assessments of influence campaigns that support leaders in making informed decisions regarding information operations on social media platforms.

4.1.1 What Do We Want to Know?

Not all influence campaigns are the same. Not all leaders, commanders, executives, and decision-makers are the same. Creating a framework for analyzing online information operations must remain flexible to address a constantly changing environment and audience. However, the methodology must remain prescriptive enough so that analysts understand the type of information and analysis that should be uncovered and how to frame it to tell a compelling story with actionable information.

Though this framework is similar to other frameworks in that it addresses similar essential aspects of an influence campaign, it improves upon these frameworks by giving analysts the tools and methods to develop a more comprehensive report. Francois, Alaphilippe, and Pamment's ABC/D/E (Actor, Behavior, Content, Distribution, and Effects) frameworks [3, 57, 101] lack in telling a complete overall story of the campaign, while Blazek's SCOTCH (Source, Channel, Objective, Target, Composition, and Hook) framework is only a quick and shallow assessment of the situation. Correspondingly, Nimmo's 4 D's and the DISARM/ATT%CK/AMITT frameworks only focus on the actions. Independently, none of these frameworks incorporate procedures for executing a complete assessment.

To address these shortcomings, the BEND framework integrates the fundamental analytical features of these analytical approaches, incorporates them, and then expands on them. The most important thing to address for the decision-maker is to understand the major points of the analysis, see how the analysis fits into the current situation, understand why the results matter, and be able to frame this information to make informed decisions regarding the results.

4.1.2 Developing the Social-Cyber Maneuver Analysis Framework

The development of this BEND analysis framework occurred in two primary ways. First, methods were identified in the process of analyzing the events described later in Chapter 5. This was the systematic use of social network analysis tools while still attempting to answer the questions of ”who” is doing what to ”whom” with ”what” and ”why” and with ”what impact?”

The second major contribution to the construction of the framework was through the process of designing, developing, and executing the OMEN exercise, a project sponsored by the Office of Naval Research. For the OMEN exercise, a team in the Center for Computational Analysis of Social and Organizational Systems (CASOS) developed a five-day game and the data to train analysts on how to examine social media data and provide an assessment. The scenario was based on a NATO exercise, and the analysts' assessments were shaped to support the military decision-making process. They were trained on social network tools, the BEND framework, and an approach for analyzing the data and preparing reports regarding the online informational environment for the commander. Several iterations of observing analysts explore the data and develop their tactics, techniques, and procedures (TTPs) contributed to some of the methods in this chapter.

4.2 Methodology Overview

4.2.1 The Analysis Pipeline

The social-cyber maneuver analysis pipeline is a method of processing the social network data sets from collection through the BEND maneuver analysis and assessment, as shown in Figure 4.1. This process begins with filtering the collected data to achieve a dataset that focuses on particular themes or events. This can be done by reducing the tweets to tweets that either contain specific keywords or hashtags or are from pre-identified user accounts. Another step for data preparation is using a bot detection program to identify possible bots within the collected actors. This thesis uses Bothunter for all bot detection [16]. Additionally, sometimes a dataset may revolve around conflicting groups or ideas, such as discussions regarding elections or complying with various health-related policies surrounding COVID-19. A useful technique is to apply a stance detection algorithm to separate the actors and possible hashtags and keywords by various stances.

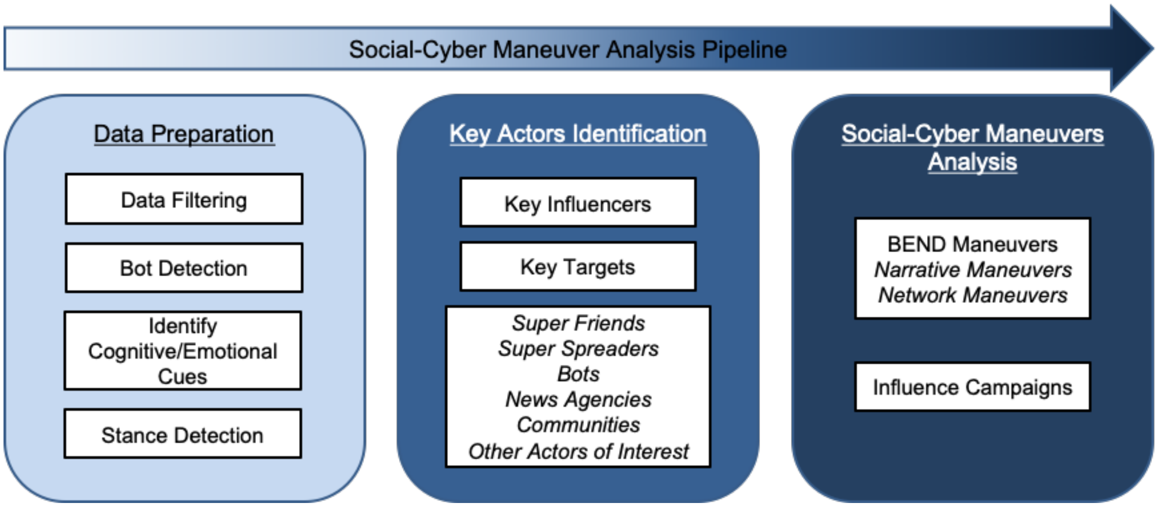


Figure 4.1: Social-Cyber Maneuver Analysis Pipeline. This pipeline includes data preparation using supplemental algorithms and software, key actors identification, and a social-cyber maneuver analysis.

After pre-processing, the next step is to identify key actors. This involves understanding who the influencers are and who they are targeting, or the maneuverers and maneuverees, respectively. Actors can be characterized by different characteristics such as if they are news agencies or other interesting actors within the data set. Furthermore, users can also be identified by their positions within the social network. They may be super spreaders, users whose content is spread often, or super friends, users who frequently engage in two-way communication.

The last part of this pipeline is where the social-cyber maneuver analysis occurs. Using the information gathered in the first two modules, influence campaigns can be characterized using the BEND maneuver framework. This final step culminates with an analysis and an assessment that is described in the next section.

4.2.2 Social-Cyber Maneuver Analysis: BEND Framework Workflow

This section outlines the social-cyber maneuver analysis methodology as shown in Figure 4.2. The purpose is to create an operational framework that can be applied to the data sets of online social networks focused on specific events or scenarios. The framework is an iterative process of analyzing social-cyber maneuvers by focusing on various aspects of the social network to evaluate influence operations and their impacts over time.

The very first step of the process is conducting the initial BEND assessment. The automated report produced by the ORA-PRO software [39] provides an overview of the data that allows analysts to get a sense of the influencers, targets, communities, and narratives. While analyzing these different aspects of the data, an analyst may decide that they may need a more in-depth of an individual or specific community or topic resulting in more focused BEND reports. The BEND report is the starting point of that investigation.

There are multiple components in the analysis process. First, there is a focus on the BEND maneuvers themselves and how they are used and received by the influencers and targets as individuals and if applicable, as communities. Narratives and topics within the narratives may need to be dissected to gain insight into the use of narrative maneuvers. Network measures are then calculated to understand the data from a macroscopic view of the network in terms of features such as density, average stance, and polarization. The process is further iterated to dynamically analyze the network over time and evaluate the impacts of the maneuvers. The result is a comprehensive analysis of the situation, and this analysis is then converted into an assessment consisting of actionable information that can be used for supporting decision-making.

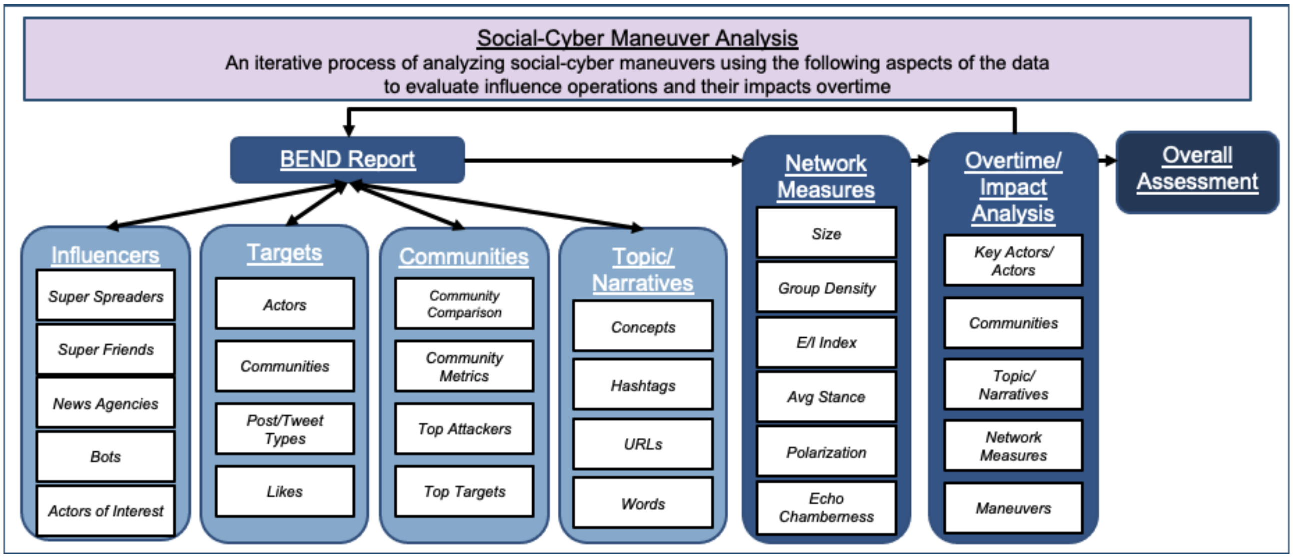


Figure 4.2: BEND Analysis Workflow

Netmapper and ORA-PRO [39] are two essential tools for detecting and calculating the BEND maneuvers for analysis. Netmapper is used to detect CUES, which are indicators for detecting the maneuvers. These CUES are then pipelined as input into the ORA-PRO software, which creates the reports containing information about the actors, communities, topics, networks, and the BEND maneuvers within the data sets. The examples in this chapter use Netmapper v94 and ORA-PRO v154.

The procedures for importing the CUES and running the BEND report in ORA-PRO are described in Appendix D.

4.3 Analysis Focus

During this process, there are multiple fundamental facets of the influence campaign that require study. Because this process is iterative, some of these facets may require more scrutiny depending on the information that comes to light during the investigation. An interesting narrative discovered while perusing the data or an unexpected highly connective actor may warrant a deeper look at a community or specific actor. However, finding these leads to follow, may only be uncovered when asking the right questions or stepping through the analysis using certain techniques. This section discusses each facet in detail, provides questions that can lead to a comprehensive report, and then describes some of the analysis procedures.

4.3.1 BEND Maneuvers

The BEND maneuvers are fundamental actions conducted by online actors to manipulate ideas, beliefs, and behavior to achieve a desired end state. They are described in detail in Chapter 2 and listed again in Table 4.1 for quick reference. Influencers use these maneuvers, and these maneuvers are used to target online users. Narratives can be more convincing or less effective because of these maneuvers. These maneuvers can alter the social network structure manipulating the way information flows and how many times a reader is exposed to certain narratives. These maneuvers, and subsequently their BEND tactical categories, as discussed in Chapter 2.4, are effective ways to describe the actions occurring as related to the primary focuses of analyzing an influence campaign.

Table 4.1: 16 BEND Maneuvers.

	Community	Narrative
Positive	Back	Engage
	Build	Explain
	Bridge	Excite
	Boost	Enhance
Negative	Neutralize	Dismiss
	Nuke	Distort
	Narrow	Dismay
	Neglect	Distract

4.3.2 Key Actors

Key actors refer to the major players of an influence campaign, and there are many ways to examine them. These actors may be pre-identified as targets of interest, but some may have to be derived through social network analysis.

These actors can fall into various types. First, they can be identified as having a particular characteristic. They may be news agencies, government agencies, or verified users. For Twitter, a verified user is an account that pays for the distinction of being “notable, active, and authentic” through a Twitter Blue. The verification process was free through December 2022. However, legacy verified users refusing to subscribe by April 2023 are expected to lose their verified status [131]. Other types of agents may be potential bots, where the probabilities of bot likelihood can be calculated through algorithms such as Bothunter [16] and Botbuster [96]. Depending on the intent of the analysis, other actor types may be of interest. These may be people of distinct identities, particular communities or organizations, or from specific locations or countries.

Another way of identifying types of key actors is by evaluating their positions within the social network. They can be highly centralized users with many followers, or they may have a high degree of interaction within a communications network. A user may also have a high betweenness centrality, indicating that they are a *bridge* between different communities or groups. When evaluating influence, calculating metrics for superspreaders can help identify the users who produce content that is shared often. This means that they efficiently diffuse information. Likewise, calculating superfriends can identify users that engage in two-way communication or reciprocity. These users can wield more influence as they have a high number of back-and-forth discussions.

From the context of an influence campaign, users are categorized as influencers and targets. Influencers are actors or a community using influence methods to achieve an end state, and they direct those actions toward a target audience. Both influencers and targets can have one or more agent types. They may be state actors, bots, or both. They may be verified users and superspreaders. As such, influencers use BEND maneuvers on their targets, and characterizing both actors is important for gaining a better understanding of who the key actors are and how they fit in the overall operation.

Questions to consider when analyzing key actors:

- How many of each key actor type exists within the data set? What percentage of the data set? What percentage of the top superspreaders or superfriends are of each of these data types? Are there any notable actors within these percentages?
- Who are the higher-scoring superspreaders and superfriends?
- What characteristics do the influencers have? What characteristics do the targets have? Are they bots, news agencies, verified actors, or any other agents of interest? What do we know about their positions within the network? Are they super spreaders or super friends?
- What are the influencers' objectives? Are there multiple objectives? What are their overarching goal and desired end state?
- Why are the influencers targeting specific actors? Are they of specific communities? Do they provide the influencers with some sort of advantage?
- What types of maneuvers are the influencers using? Do some types of influencers (e.g., bots, government agencies, etc.) use different methods? Do they have malicious intent?
- Are the targets more vulnerable or susceptible to certain maneuvers?
- What narratives are the influencers discussing? Which maneuvers are the most effective for spreading the narrative?
- Are the influencers using more complex methods such as hashtag hijacking or hashtag

latching? Hashtag latching occurs when unrelated hashtags combine unrelated topics to bring the attention of an unassociated audience to a particular message. Similarly, hashtag hijacking alters a hashtag's original purpose by tying the hashtags to an alternate message topic in large frequencies.

- How are the influencers interacting with their targets? Likes, mentions, retweets, replies, etc.
- Are there indications of influencers or targets clustering among themselves or with each other? For example, there may be a botnet of influencers, or an influencer may have infiltrated a group of targets.

Possible actions:

- Sphere of influence - Analyzing the ego network of an individual can provide valuable information about an influence campaign. This can be used for looking at opinion leaders in competing communities or other actors of interest.
- Compare BEND maneuvers between key actors.
- Calculate individual centrality measures (degree, eigenvector, betweenness, superspreader, etc.) These are calculated measures of an actor's position in the network relative to other actors and the network itself.

4.3.3 Communities

There are multiple types of communities that may emerge from an influence campaign. They may be in the form of topic-oriented communities or individuals discussing the same topics at around the same time. One way of detecting these communities is through clustering algorithms over topic and interaction networks. This same process can also help identify possible subgroups of specific ideologies or other topic-oriented groups. Communities of actors can form based on similar identities or characteristics. They may be separated into different ideologies or stances such as pro-vaccine and anti-vaccine or left-wing and right-wing political parties, or they are simply separated by actor types, such as bots or not bots.

One method of analyzing communities is to calculate various community network measures on the social network structure, as the network structure physically dictates the possible paths that information flows. The density of the group agents can show the amount of communication occurring within the group. The decrease in density over time, for example, can indicate that a topic-oriented great is either no longer conveying the same messages in coordination or that a particular topic is no longer a strong topic for the group. Echo-chamberness and reciprocity can indicate how close-knit a group is, which can amplify narratives within the community. The E/I index shows the ratio of interactions inside a community to those outside a community. These overall network measures can provide insight into the extent of interactions and relationships between agents and the messages they are trying to convey.

Several communities may emerge within the data set that may appear necessary to compare. There may be two large competing groups: an anti-vaccine versus pro-vaccine community. If focusing on sub-communities, a possible avenue of approach is evaluating why the larger community is broken into sub-community and then comparing the sub-communities. There may be groups of anti-vaccine supporters who oppose entirely all types of vaccines or those who only oppose the COVID-19 vaccine. Furthermore, some communities may only occur temporarily with the data set. Perhaps in the vaccine scenario, the anti-vaccine proponents only opposed a certain aspect of the vaccines that were eventually resolved after passing a specific policy or event.

In the same way that individuals attempt to execute BEND maneuvers to influence their targets, entire communities can be characterized by which maneuvers they predominantly use to convey their messages and affect others. Some communities may be more likely to apply the scare tactics of *dismay* maneuvers while others may rely on *back* maneuvers to support a key opinion leader in their group. Conversely, another technique could be looking at how maneuvers are targeted toward a specific group. For example, vaccine-hesitant communities are the targets of both anti-vaccine and pro-vaccine users, and it may be important to see what maneuvers are used in high frequency and which may appear to result in the most behavioral responses. Furthermore, the impact of many of the community maneuvers is the changes in the network structures. Supporters typically want to make their communities grow and for their opponents to be reduced in size. Overall, the BEND maneuvers are a very necessary part of community analysis.

Questions to consider when analyzing communities:

- How are these communities characterized? By what topics? by what attributes? By what ideologies?
- How are these communities structured? Are they centralized? Are they an echo-chamber?
- Are these communities influencers or targets?
- Who are key actors within the community? How many of each actor types are in the communities?
- Who are the leaders?
- Key actor-related questions from Section 4.3.2 but from the perspective of a community. For example, is this community a community influencers? What BEND maneuver are they using? What narratives?
- Are there sub-communities within the communities?
- Are these communities engaging in coordinated behavior?

Possible actions:

- Apply a clustering algorithm on a network relating topics and the interaction of users to detect topic-oriented communities
- Separate communities by a specific attribute or agent type, such as bots, stance, news agencies, etc.
- Calculate descriptive statistics of the communities.

- Compare communities using network measures calculations such as user and tweet count, density, centralization, echo-chamberness, E/I index, etc.
- Conduct BEND analysis on individual communities.

4.3.4 Topic or Narrative

The topics and narratives are the most prominent part of the influence campaign. Influencers focus on what they are going to say and how they are going to say it. Depending on how the narrative is framed and the maneuvers used to convey it, a target is more likely to believe the message.

Analyzing the topic or narrative can be done in several different ways. One method is to extract the hashtags and URLs and observe the most used or focus on the hashtags commonly used together. Figure 4.3 shows a visualization of the hashtag x hashtag co-occurrence network during the Pfizer/BioNTech COVID-19 vaccine rollout. In this type of network, hashtags are linked together if they exist in the same message as each other. This example illuminates the relationships between seemingly unrelated concepts resulting from the extensive use of hashtag latching. Hashtag latching is connecting two unrelated topics by combining unrelated hashtags within a message. During the stance detection process, the hashtags *#BidenCheated2020*, *#saynotomasks*, and *#endthelockdown* emerged as anti-vaccine hashtags, showing the close relationship between anti-vaccine rhetoric and discussions on the 2020 election, mandates on masks, and lockdown policies.

A more direct method for topic analysis is to use machine learning topic modeling tools such as Latent Semantic Analysis (LSA) or Latent Dirichlet Allocation (LDA). For example, Uyheng et al. [132] applied LDA in the analysis of social media data surrounding the NATO Trident Juncture exercise in 2018. The methods can help look at the important narratives or themes that are spreading throughout the data set.

Furthermore, depending on the time period of a campaign, the sub-topics that support an overarching narrative may only appear temporarily, appear only after a certain event, or exist throughout the entire time. The topics may be branches of larger topics or an evolution of the topic. While conducting the analysis, it is important to be able to differentiate between the primary narrative and the supporting sub-topics.

Questions to consider when analyzing topics or narratives:

- What are the primary narratives? sub-narratives? topics?
- Is the information factual? disinformation (intent to deceive)? Misinformation? Malinformation?
- What are the major arguments for each narrative?
- What are noteworthy hashtags?
- What are noteworthy URLs?
- What narrative maneuvers are being used?
- How pervasive are the narratives throughout the network?

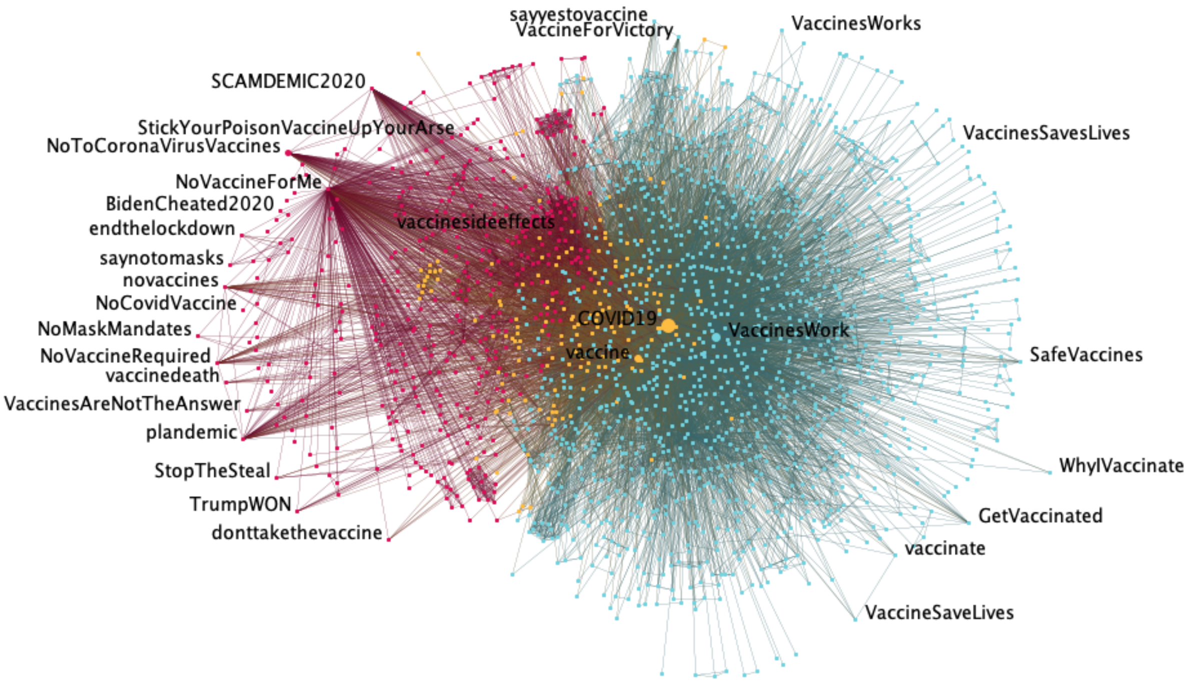


Figure 4.3: Example Hashtag x Hashtag co-occurence network based on COVID-19 vaccine data during the week of the initial Pfizer/BioNTech rollout; blue is anti-vaccine stance, red is pro-vaccine stance, yellow is unassigned; sized by degree centrality

- Are there narratives divided into stances? Which topics fall in each stance?

Possible actions:

- Apply topic-modeling tools to extract topics from the data set.
- Identify the top tweets for each maneuver and their associated topics.
- Relate topics to key actors and communities. Visualize using network agent nodes and topic labels.

4.3.5 Network Measures

While community network measure analysis may focus on the influence or target communities, there is value in looking at the network as a whole. The network measures used in community analysis are the same ones used for an overall network analysis. As a network, some of the network measures that can be calculated are size, density, Krackhardt E/I index, and echo-chamberness. This is important in looking at the different connections between users and information and the extent of the network. However, it may be useful to look at the ideological aspects of the data set. If examining election issues, the extent of polarization or the average stance of the agents within the data set may be appropriate calculations. This can be an interesting way of seeing how types of users and messages are distributed across the various data sets.

Questions to consider when analyzing network measures:

- How centralized is the network? Degree? Betweenness? Closeness?
- How close-knit or internally/externally reaching is the network? Density? Reciprocity? Echo-chamberness? E/I index?
- How polarizing is the network?
- What percentage of the network is of a particular stance?

Possible actions:

- Create network visualizations of data. Colored, sized, grouped, and labeled by attributes or network measures.
- Summarize the characteristics of the agents in the network? Number of actors, actor types, etc.
- Summarize other network characteristics. Number of tweets, hashtags, locations, URLs, etc.

4.3.6 Overtime/Impact Analysis

Overtime analysis is using dynamic network analysis to look at various characteristics of social networks and see how they compare over time. This is useful in creating an impact analysis to evaluate how the maneuvers and external events impact aspects of the data. Several overtime measures can be observing the types of key actors or actor attributes, changes in the communities, the evolution of topics and the overall narrative, general network measures and calculations for features such as average stance or information diffusion, maneuvers used, and any other relevant changes to the network that may be a result of external events. Looking at these different aspects over time can potentially be useful in extrapolating future behavioral responses or other impacts on the network.

Questions to consider when conducting overtime/impact analysis:

- How have metrics for characterizing the actors, communities, narratives, and network measures changed over time?
- Are there more or less actors of a certain agent type or community?
- How has the narrative evolved? Was one narrative more prevalent and then changed because of a particular event?
- How has network structure evolved over time? Size? Types of users? Overall network stance or polarization? Echo-chamberness? E/I density?
- How has the use of BEND maneuver changed?

Possible actions:

- Create an overtime line graph of metrics.
- Compare the first day and the last day of metrics.
- Conduct an overtime topic analysis.

4.4 Assessment

The purpose of the assessment is to help leaders understand an influence campaign by visualizing and describing their operating environment, enabling them to make decisions and take direct actions. The assessment lays out the information gathered from the analysis and places it within the context of the situation which prompted the analysis. This involves presenting the major findings for each of the sections of the framework, discussing the impact of those results, and explaining why these findings are important to the decision-maker, the stakeholders, or the commander. Then based on an overall assessment, an analyst proposes recommended courses of action to support the mission or the objectives of an organization.

4.4.1 Making BEND Recommendations

Making recommendations for using BEND maneuvers in response to actions identified during analysis depends on several variables. First, depending on the organization's role, recommendations may vary based on the rules of engagement or guidance from a commander or decision-maker. For example, a public affairs team may have different objectives than a unit primarily focusing on offensive information operations. Passive organizations may solely focus on observing, analyzing, and reporting, while those more operationally centered may decide to use direct counters to the influence campaign. The desired end state of an analysis team will vary based on the type of organization it supports.

One method for framing BEND recommendations comes from the Department of Defense's Joint Publication for information operations [75]. The diagram in Figure 4.4 describes their information-influence relational framework. First, this framework consists of understanding the information environment which consists of the human-centric cognitive dimension, the data-centric informational dimension, and the tangible physical dimension. The next steps are identifying the target audience of key influencers, mass audiences, and vulnerable populations from which information flows; the rules, norms, and beliefs that affect behavior; and the means and ways to achieve the end state. The BEND framework can fit into this framework as an application for understanding and approaching online influence. The social-cyber maneuver analysis is the method for understanding the information environment and the target audiences. Using social media as the means, the BEND recommendations are the ways to affect the rules, norms, and beliefs required for influencing behavior and achieving a desired end state.

Given guidance and the limits for operating, the recommended actions begin with knowing the main objective for the influence campaign, identifying what BEND tactical categories that support that campaign (some are suggested in Chapter 2.4), and then selecting the BEND maneuvers for those tactical categories. For example, a unit may seek to improve public relations

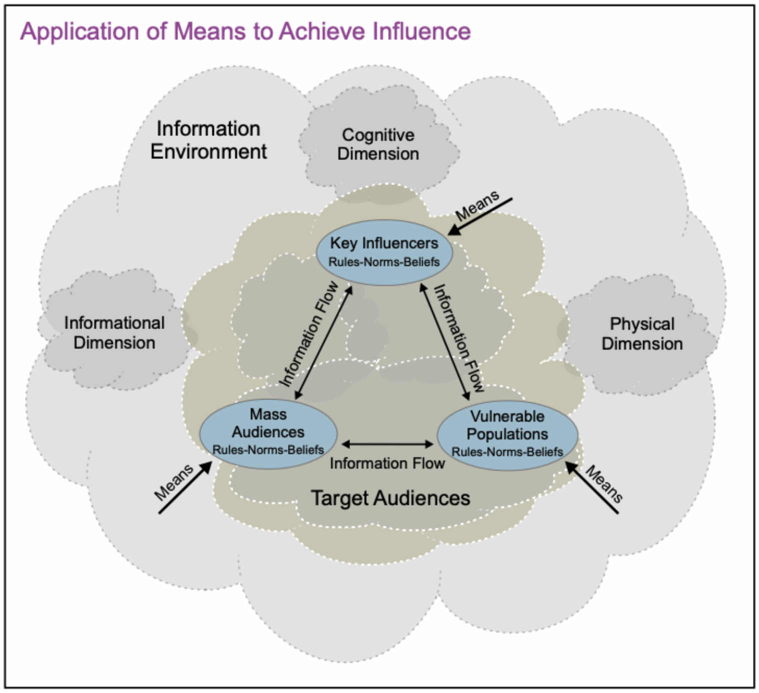


Figure 4.4: Diagram for achieving influence from Joint Publication 3-13 [75]

regarding a military exercise in a foreign country. Their goals would be to develop a narrative that supports the exercise's positive effects on readiness and conducting a collective defense operation. They may decide to *explain* how the training is making soldiers better with some of their tactical tasks, *enhance* with images or videos of helicopters conducting complex maneuvers, and *engage* with a discussion about "being proud of your sons and daughters serving the military." The campaign may consist of multiple tactical categories or any combination of maneuvers required to achieve the desired effect. Furthermore, these combinations of maneuvers can be spread over time and with varying frequency to create a complex online influence operation.

Questions to consider when developing the assessment:

- What is the context or background for this assessment?
- How do these results fit into the current tactical situation?
- What are the major points of the analysis? Who is doing what to whom and why, and with what impact?
- How do we present the material in a way that allows leaders to take appropriate action and make informed decisions based on the analysis?
- How do the BEND maneuvers help describe the entire influence campaign as a whole, and why is this information useful?
- What are the impacts on the influencers, targets, and the analyst's organization or stakeholders?

- Why do these results matter?
- What is the overall assessment?
- What are recommended courses of action? Based on our knowledge of the actors and the information, what effects are needed and what are the means for achieving those effects?

4.5 An Adaptable Framework

Though designed for a military application, this framework is designed to be adaptable to a diversity of organizations that requires an analysis of online influence operations. It provides a common language, which can lead to effective communication as well as common understanding of the different facets and tools of the methodology and the resulting analysis of the situation. For the military, this can mean that a public affairs team can develop applicable analyses from the framework just as well as an information operations or intelligence cell. Furthermore, the framework can correspondingly be applied to non-military organizations such as news agencies or private companies requiring further analysis of influence campaigns. Here are two different cases for using the framework.

Case 1: A public affairs office aims to send messages to mitigate the effects of unfavorable reports of an ongoing military exercise. These reports are having a damaging impact on the organization's image. The framework can be used to examine and assess the extent of the damaging narrative. This includes analyzing the actors involved, the methods by which the narrative is being spread, and how the network is changed by these negative messages. Furthermore, the framework offers possible online actions that can be used to reduce the impact in the form of the BEND maneuvers and tactical categories.

Case 2: An information operations cell may need to conduct a comprehensive report on an online influence operation to support real-world key leader engagements or official meetings between two organizations. Using the framework, the analysts can extract detailed information from a social network, identify potential underlying issues, examine a leader's sphere of influence, and understand the overall information environment. This information can potentially be used to develop talking points for resolving issues or improving relationships between the two parties.

4.6 Conclusion

This chapter outlined the process for analyzing and assessing online information operations. It provided a methodology, proposed questions to consider, and some guidance on how to derive the information. This process emphasizes collecting pertinent information and then framing and visualizing the analysis in a manner to support a leader's understanding and ability to make informed decisions given the results and the operational environment. In isolation, the individual BEND maneuvers are not useful for a commander. They are tools for understanding the actions occurring in the influence campaign and are part of the larger aims of the framework for knowing *who is doing what to whom and why and with what impact*?

The next chapter presents three case studies that apply the methodology and procedures discussed in this chapter. They are comprehensive illustrations of applying the BEND framework for analyzing influence operations and can be used as examples while performing an analysis.

Chapter 5

Tell Me More About This Influence Campaign. *Three Case Studies*

5.1 Introduction

In this chapter, I apply the BEND Framework to three scenarios. These case studies illustrate the use of the framework as an approach for quantitatively and qualitatively analyzing influence campaigns.

The first scenario explores the influence of pro-vaccine and anti-vaccine proponents during the rollout of Pfizer's COVID-19 vaccine in December 2020. The second scenario analyzes the influence carried about by left-wing and right-wing communities during the U.S. 2022 midterm election. The final scenario focuses on two key actors during the Russian invasion of Ukraine in early 2022 - President Volodymyr Zelensky of Ukraine and President Vladimir Putin of Russia. This case study is written in the style of a report and less of a research paper.

5.2 COVID-19 Vaccine Rollout Case Study

5.2.1 Introduction

COVID-19 claimed the lives of 2.6 million people in the first year of its discovery [72]. In a concerted effort to reduce the cases and deaths resulting from the COVID-19 pandemic, governments and major health organizations pushed for the development and rapid distribution of COVID-19 vaccines. This process, however, has been met with online expressions of resistance vaccination [25]. In view of this concerning spread of anti-vaccine sentiment online, this work has focused on identifying the specific tactics used by both the pro- and anti-vaccine communities to spread their messages over Twitter.

Though vaccinating everyone against COVID-19 may seem to be an obvious way to prevent deaths, many people and groups oppose vaccination for several different reasons. The first compulsory vaccination was established in England by the Vaccination Act of 1853. The act faced opposition to the idea that the government should impose health legislation [111]. In current times, communities speak out against the government and assert that they have the right to decide what goes inside their bodies. Some anti-vaccine proponents fear the side effects of vaccines and refuse entirely to vaccinate themselves or their children because of rumors of autism or other medical disorders [51, 110].

The Pfizer-BioNTech vaccine was the first vaccine for preventing COVID-19 to be authorized in the United States by the Federal Drug Administration (FDA). Both the FDA and European Medicines Agency (EMA) authorized the vaccine for emergency use. In 2020, the first Pfizer vaccine doses were distributed in the United Kingdom on December 8 and in the United States on December 14. Because of the rush to create the vaccine, many feel the vaccines were inadequately tested and refuse the vaccine without seeing the results of long-term studies. Additionally, some accept conspiracy theories or rumors on the vaccine. For example, one such conspiracy theory is that Bill Gates and the government created the vaccines to microchip the population for some malicious intent [119]. These are among the reasons for "vaccine hesitancy" across the world [45].

Social media has become a medium for COVID-19 vaccine discussion. Twitter is a popular platform on which government leaders, public health officials, and news organizations spread pertinent information. However, many users spread misinformation or act maliciously by conducting influence campaigns to manipulate peoples' beliefs and ideas. Bonnevie et al. [25] found that vaccine opposition on Twitter increased by 80% after COVID-19 began spreading in the United States. Misinformation is not limited to anti-vaccine users as some pro-vaccine users also share unreliable information [71]. To counter the spread of misinformation on its platform during the initial administration of the vaccine, Twitter expanded its policy by removing false and misleading tweets about COVID-19 vaccines, adding labels to potentially misleading COVID-19 vaccine information, and creating a "five-strike system" for suspending misleading accounts [128, 129].

These malicious actions online are a major aspect within the field of social cybersecurity. Social cybersecurity lies at the intersection between cyberspace and human interaction. It studies how humans can be influenced by tactful messaging and connecting the right people to the right content. Key players in an online social network can conduct influence maneuvers to change users' beliefs and affect their behavior [35]. This study aimed to identify the important actors in pro- and anti-vaccine Twitter communities as well as the social-cyber maneuvers they used to influence their audiences' stances regarding the COVID-19 vaccine.

This work focused on the time period around the approval and initial administration of the Pfizer vaccine. The objective was to determine whether there are differences between the types of social-cyber maneuvers pro-vaccine and anti-vaccine communities use toward their target audiences. Described is a methodology for determining pro-vaccine or anti-vaccine stances within tweets and identifying key players within the social network. Bot detection and linguistic cues were used to analyze the content and significance of tweets, and how the opposing vaccine communities applied social-cyber maneuvers to persuade their target audiences was evaluated. The results show how pro-vaccine messaging focused on exciting readers and explaining the vaccine issue. In contrast, anti-vaccine groups preferred to make dismaying statements and used messaging that distorted vaccine information. Also found was that Twitter's tightening of its policies on vaccine misinformation had a remarkable effect on decreasing the size of anti-vaccine communities and the prevalence of their messaging.

5.2.2 Related Work

Vaccine Stance Detection

The problem of identifying pro- and anti-vaccine communities has garnered the attention of several researchers who have sought to apply stance detection techniques from computer science to this task. Supervised machine learning methods developed for this problem have ranged from the use of transformer neural networks based on Google's Bidirectional Encoder Representations from Transformers (BERT) model [81] to the use of convolutional neural networks trained on n-grams and topics detected via Latent Dirichlet Allocation [79]. More traditional community detection algorithms have also been used to find groups with overt stances on vaccines [116]. The semisupervised stance propagation technique used for this work, which has the advantage of not requiring extensive manual labeling of pro- and anti-vaccine messages, was also used to identify linguistic differences between pro- and anti-vaccine groups [87].

Pro-vaccine and Anti-vaccine Communities

Various studies of anti-vaccine and pro-vaccine communities have sought to identify the methods used for spreading vaccine-related messages. Different communities can have contrasting messaging characteristics depending on the nature of and support for their stances on vaccines.

In 2019, a study examining influential themes and actors within the anti-vaccine community concluded that top tweeters relied on highly networked communities led by accounts that select messages expected to have high receptivity within those communities [24]. This was different from standard messages from public officials, which tended to repeat the same information to the same communities, limiting the extent of a message's reach. In an analysis of Facebook vaccine group clusters, Johnson et al. [74] observed that anti-vaccination clusters entangled more often with undecided clusters, while pro-vaccination clusters tended to be more peripheral. Furthermore, Schmidt et al. [116] examined how echo chambers reinforce the opinions of groups and how involvement within these groups could be an effective way of countering anti-vaccine beliefs.

Past research has therefore found that pro-vaccine messages tend to be supported by public health officials and governments seeking to reduce the spread of infectious diseases, whereas anti-vaccine communities are more niche and maintain a smaller following. However, although pro-vaccine messages tend to stay within pro-vaccine communities, anti-vaccine messages permeate beyond the boundaries of anti-vaccine communities.

Although these past works have analyzed the themes and targeting of vaccine messaging, they have not considered the specific types of strategies carried out in vaccine-related information operations. Thelwall et al. [125] tracked some of the anti-vaccine narratives spreading on Twitter, and Boucher et al. [26] identified the key themes in Twitter conversations about vaccine hesitancy. However, previous research has not examined the intentions behind specific choices on the language, content, and targeting of pro- and anti-vaccine messaging. This work breaks down the tactical value of specific types of vaccine messages and analyzes how those tactics have changed over time.

Social Cybersecurity: Influence Campaigns and Bots

A key development in the fight against online influence campaigns has been the growth of the field of social cybersecurity, a computational social science that aims to protect the security of democratic societies by studying the ways in which actors exercise manipulation on social media platforms [35]. Recognized by the National Academies as a new science [93], its key areas of research have been the study of information maneuvers, motive identification, and information diffusion, as well as the evaluation of the effectiveness of information campaigns and mitigation strategies. Though the field has most extensively focused on the spread of political disinformation, it has more recently expanded to tackle the problem of medical misinformation [35]. Of particular concern in social cybersecurity is the existence of automated accounts on social media platforms since they are used to spread online disinformation and influence elections [54]. They have also manipulated public health discourse by propagating misinformation on topics such as e-cigarettes, diets, and medications [4]. Because of the influence of these bots on public opinion, several studies have been conducted on the use of bots for spreading vaccine information [144]. Before the COVID-19 pandemic, Broniatowski et al. [32] examined the extent to which bots spread anti-vaccine messages, showing the high rates of vaccine content they spread and comparing it with the effects of Russian trolls, whose messages primarily sought to increase discord online. Dyer [52] determined that after Russian trolls, bots were the most prolific vaccine-related tweeters. Huang and Carley [70] found that accounts linking to coronavirus information from less reliable sites were more likely to be bots. Ng and Carley [95] also found that bots change vaccine stance more easily than non-bots. Hence, understanding the actions of automated accounts is a crucial part of vaccine-related online influence campaigns.

The BEND Framework

A crucial component of social cybersecurity's efforts to characterize online influence operations has been the struggle to establish the motives and tactics of those seeking to manipulate conversations in cyberspace. The BEND framework was developed to assist in the theoretical conceptualization of this problem by providing a taxonomy of 16 categories of maneuvers for conducting online influence [17]. These categories are divided into 2 types: narrative and network maneuvers. These types are further divided into positive and negative directions of influence. Narrative maneuvers focus on the information and content of messages. These maneuvers affect what is being discussed and how it is discussed. Network maneuvers focus on how the network and communities are shaped and the positions of key actors. The BEND framework provides analysts and researchers with a way to conceptualize the tactics used in online information operations.

5.2.3 Methods

In this work, the methodology used is similar to the pipelines in other social cyber-security studies [132]. For the social-cyber maneuver analysis, the end state is to gain a comprehensive

Table 5.1: Keywords used to collect COVID-19 vaccine-related tweets.

Filter	Keywords
Filter 1: COVID-19 tweets	coronaravirus, coronavirus, wuhan virus, wuhanvirus, 2019nCoV, NCoV, NCoV2019, covid-19, covid19, covid 19
Filter 2: vaccine tweets	vaccine, vax, mRNA, autoimmuneencephalitis, vaccination, getvaccinated, covidisjustacold, autism, covidshotcount, dose1, dose2, VAERS, GBS, believemothers, mybodymychoice, thisisourshot, killthevirus, proscience, immunization, gotmyshot, igottheshot, covidvaccinated, beatcovid19, moderna, astrazeneca, pfizer, johnson & johnson, j&j, johnson and johnson, jandj

understanding of the actors and their maneuvers used to manipulate others on social networks.

Data Collection

The data used in this work are a subset of COVID-19 tweets collected from Twitter using the Twitter application programming interface (API) and keywords related to COVID-19. The data set was then further filtered using the vaccine-related terms shown in Table 5.1. Furthermore, tweets from non-English speaking users were removed from the data.

The data was into 3 time periods surrounding the introduction of the Pfizer vaccine: December 1-7, 2020 (the week before the rollout), December 8-10, 2020 (during the week of the rollout in the United States and the United Kingdom), and January 25-31, 2021 (6 weeks after the rollout). The 3 periods consisted of 471,962, 694,200, and 662,776 users and 935,709, 1,511,344, and 1,368,035 tweets, respectively.

Identifying Bots

The probability that each user within the data set was a bot was calculated using the Tier-1 BotHunter algorithm by Beskow and Carley [16, 19]. BotHunter is a random forest regression model trained on labeled Twitter data sets. It was developed from forensic analyses of events with extensively reported bot activity, such as the attack against the Atlantic Council Digital Forensic Research Lab in 2017. This machine learning model considers network-level features (such as the number of followers and friends), user-level attributes (including screen name length and account age), and tweet-level features (such as timing and content). For this work, any score of 75% or greater was labeled as a bot to reduce the chance of false positives and ensure that the accounts classified as bots were truly bots (at the expense of missing some bots) [97].

Linguistic Cues

The NetMapper software [39] was used to extract linguistic cues from the tweet text. These are metrics helpful in identifying a tweet's sentiment and author's emotional state [35]. Examples of

Table 5.2: The number of users labeled as pro-vaccine and anti-vaccine, along with the number of tweets by users of each stance after running the stance detector.

Time period	Users labeled by stance detection		Number of tweets by users of each stance	
	Pro-vaccine	Anti-vaccine	Pro-vaccine	Anti-vaccine
Before rollout	216,156	36,609	186,726	31,200
During rollout	195,334	47,566	292,607	55,406
After rollout	430,278	19.519	338,035	30,560

these cues include the frequency of positive and negative terms, types of pronouns, emojis, and others. These tweet attributes are used to identify BEND maneuvers and actors participating in such maneuvers.

Organization Risk Analyzer - PRO Software

The Organization Risk Analyzer (ORA)-PRO software [39] is a dynamic meta-network analysis tool used extensively in this study to examine and characterize key actors, conversations, and the overall structure of the Twitter data. Key features used included a network data visualization tool, stance detection function, Twitter analysis report, and the BEND and Community Assessment report.

Stance Detection

The stance detector [80] built into ORA-PRO divided the data set into the pro-vaccine and anti-vaccine communities. This stance detector starts with a set of hashtags that the user initially labels as pro- and anti- with respect to an issue. The stance detector uses these hashtags to label the stance of the Twitter accounts that used them. The algorithm then uses the concept of influence propagation to label the stance of users who did not use any of the pre-labeled hashtags. This propagation through the user communication network proceeds by repeating 2 steps.

First, users with a known stance are used to determine the stances of some of the hashtags that have not yet been labeled. In this step, hashtags that are used overwhelmingly by users of one stance over the other are accordingly assigned that stance. In the second step, hashtags with a known stance are used to determine the stances of some of the unlabeled users. Users who have overwhelmingly used hashtags of one stance rather than the other are labeled with that stance. Additionally, both steps allow stance to spread directly from user to user. In both steps, unlabeled users who are predominantly connected to users of the same stance are assigned that stance.

The algorithm also provides a confidence level for each stance classification. After running the stance detector on the data, pro-vaccine users had a mean confidence level of approximately 99% to 100% for each time period. However, the anti-vaccine users for the before, during, and after rollout periods had mean confidence levels of 84%, 85%, and 67%, respectively. Table 5.2 shows the number of users classified by stance for each time period and the number of tweets by these communities. There were noticeably fewer anti-vaccine users and tweets than pro-vaccine users and tweets. Though the stance detector also identified neutral nodes, these were excluded from this study.

Table 5.3: BEND maneuvers organized into application categories.

BEND maneuver and application categories		Maneuvers
Narrative maneuvers		
	Developing narrative	engage, explain, enhance
	Emotional influence	excite, dismay
	Countering narrative	distract, dismiss, distort
Network maneuvers		
	Affecting leaders	back, neutralize
	Making groups	build, boost, bridge
	Reducing groups	neglect, narrow, nuke

Examining Key Actors and Social-Cyber Maneuvers

The reports within ORA-PRO provided insight on key actors, individual tweets, BEND maneuvers [38], and the entire network. The reports were used to analyze each of the 3 time periods on each of the subsequent stance communities, and each time period was examined in isolation to observe the interactions of the users between those of opposing or neutral stances. Afterward, a more fine-grained analysis was conducted by focusing on the individual communities by stance.

ORA-PRO's Twitter report can identify and analyze key agents or actors, hashtags, tweets, and other Twitter attributes on Twitter data. Key actors are useful in understanding who are the most influential entities and what are the most influential conversations. The first type of key actors observed was super friends, or users that exhibit frequent 2-way communication with others, such as reciprocal mentioning or retweeting. The second type was super spreaders. These users generate content that is shared often, facilitating the diffusion of information across the network. Once these influencers were identified, the list of tweets and hashtags for each of these key actors was extracted for further inspection.

Additionally, the Twitter report identified valuable tweets. In this study, the focus was on the most propagated tweets within a data set. These are tweets that have the highest combined values for retweets, replies, and quotes. This information aided in understanding social-cyber maneuver narratives and actions.

The ORA-PRO software uses NetMapper's linguistic cues as input for detecting BEND maneuvers in tweets using the BEND and Community Assessment report. Of the most propagated tweets, this report in conjunction with manual labeling was used to gain insight into the social-cyber maneuvers used within the data sets.

In this analysis, the BEND maneuvers were organized into 1 of 6 application categories based on the similarity of the maneuvers: developing the narrative, emotional influence, countering the narrative, affecting leaders, making or growing groups, or dissolving or reducing groups 5.3. These represent macro-level actions occurring as a result of multiple BEND maneuvers. These combinations were observed over time to identify a concerted effort to influence target audiences of their stances. The narratives and actions for the 100 most propagated tweets for each stance community within each time period were manually labeled and grouped into these application categories.

5.2.4 Results

Key Influencers: Super Friends and Super Spreaders

ORA-PRO calculated the super friends and super spreaders for the 3 data sets for each time period and identified the types of entities that fell into each of these categories.

The top 10 super friends throughout the 3 data sets were predominately pro-vaccine, though varied in the types of actors. All of the top 10 super friends identified before the rollout were unverified Twitter accounts and relatively low-profile users. ORA classified all of the tweets as pro-vaccine, and 3 of them were identified as amplifier bots [60]. During the rollout, the top 10 included several anti-vaccine users and a single neutral stance user. Of the 3 bots on the list during this period, 2 news bots emerged alongside one of the pro-vaccine bots from the before period. At 6 weeks later, several higher-profile users from health and government organizations appeared as super friends. These included the World Health Organization, the India Ministry of Health, and the India Official COVID Response account.

Except for one instance, the top 10 super spreaders were either classified as pro-vaccine or neutral within the 3 data sets. All of the users before and during the rollout were high-profile verified Twitter accounts. During these 2 periods, the super spreaders were primarily health organizations, vaccine manufacturers, news organizations, and senior government leaders. After the rollout, the types of accounts identified as super spreaders changed. Though a couple of news organizations and health-related accounts remained on the list, the users were more community leaders or professionals with a substantial reach, such as actors or journalists.

Bot Influencers

The BotHunter results revealed that anti-vaccine agents consisted of a higher percentage of bots than the pro-vaccine agents 5.1. Though anti-vaccine bots decreased over time, the number of bots remained relatively higher than the total percentage of pro-vaccine bots of the same time periods.

The number of bots within the top 100 super spreaders and super friends 5.2 were also calculated. The high percentage of super friends shows that users are interacting with the bots and engaging in 2-way communications. The super spreaders show that bots are effectively diffusing tweets through the network. These bots have managed to connect with users on Twitter, which makes them susceptible to the information or disinformation that these bots can be posting. Furthermore, the data show a noticeable decline in the number of anti-vaccine influencers after the rollout. Super spreaders, for example, reduced from 15 and 16 bots during the first 2 periods to only 4 after the rollout. This difference is likely a result of the Twitter policy against anti-vaccine disinformation enacted mid-December 2020 [128].

BEND—Narrative Maneuvers

Using ORA-PRO, the most propagated messages for each of the stances and periods were identified. These messages were used to identify and evaluate how narratives manifested themselves as maneuvers to persuade others. This section provides an accumulation of narrative BEND maneuvers observed within each of the communities.

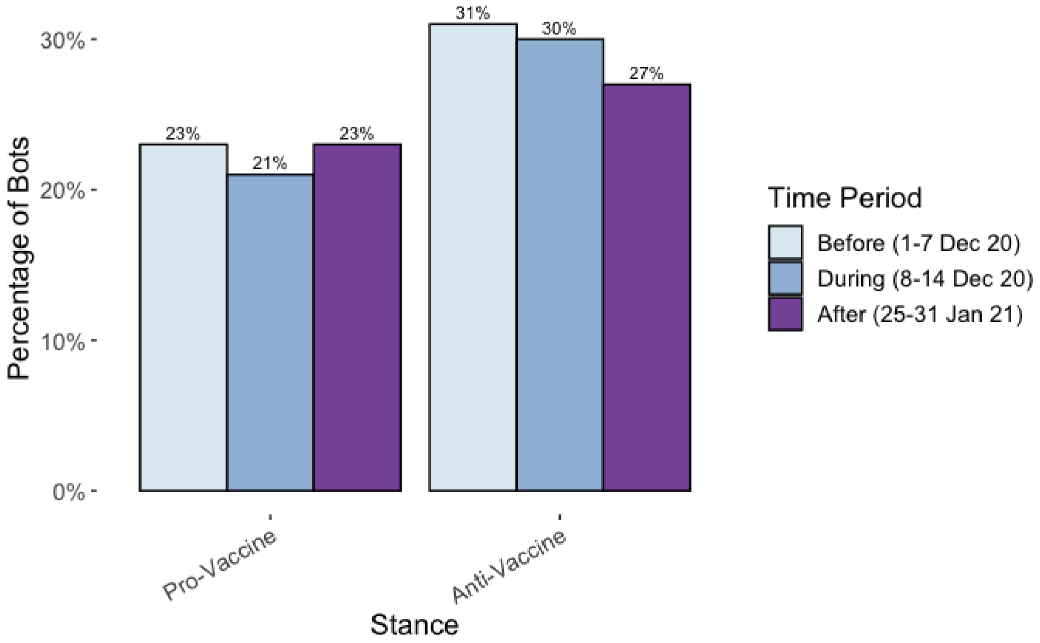


Figure 5.1: Percentage of bots by stance by time period

Pro-Vaccine Communities—Narrative Maneuvers

Pro-vaccine communities had varying messages before the vaccine rollout. Many of them were excite messages, defined as messages that elicit a positive emotion such as joy or excitement. Users posted positive tweets about the vaccine's approval and then encouraged others to get their vaccine when it became available. At the same time, many users also attempted to compel others to curb the growing number of COVID-19–related illnesses and deaths using the dismay maneuver. This is messaging to elicit negative emotion such as sadness or anger, to warn users of the consequences of not getting the vaccine. Health officials and organizations used the maneuver, explain, which is to educate on a topic using details and relevant facts, to inform the science behind the vaccine and build confidence in its use. To counter vaccine myths, users used the dismiss maneuver, the maneuver used to downplay anti-vaccine information as either irrelevant, inconsequential, or foolish. Many users also dismissed many of the fears that stemmed from the vaccine's rapid development. These maneuvers were typically followed up with explain maneuvers that attempted to debunk these myths using scientific evidence or detailed forms of justification. Finally, users would enhance their pro-vaccine ideas or encourage their views with the support of prominent actors or interesting content. Many, for example, tweeted and quoted articles about 3 former US Presidents volunteering to get the vaccine to promote trust in the vaccine.

During the rollout, pro-vaccine communities continued to post similar types of messages as in the week prior. Many users expressed excitement about the first person to receive the Pfizer vaccine, a 90-year-old woman from the United Kingdom leading the vaccine rollout. Other types of optimistic excite messages included those from users of various countries approving and purchasing vaccines as well as many excited at the sight of the logistics vehicles containing the

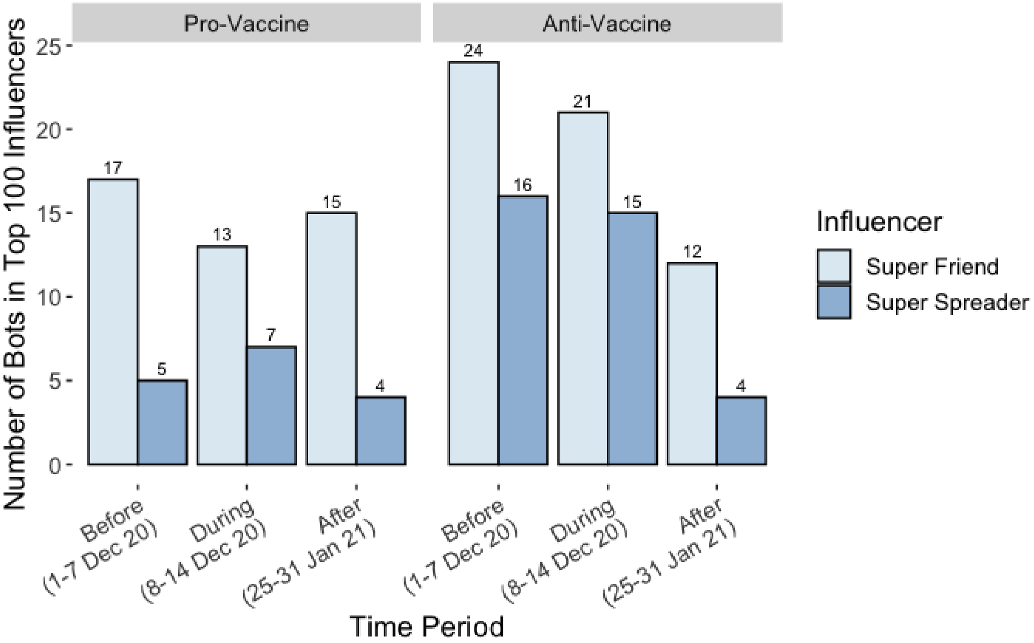


Figure 5.2: Percentage of bots among top 100 influencers by stance, time period, and type of influencer

vaccines within the distribution process. Additionally, pro-vaccine proponents also added to their explain messages about how the vaccines work by emphasizing the vaccine's effectiveness after the first dose and supporting the overall narrative with charts and results from the vaccine trials during the development process. Furthermore, medical professionals made efforts to engage with their more hesitant audiences to instill confidence in the vaccine and encourage them to get vaccinated.

After the rollout, messages continued to explain science-backed research for the development, safety, and efficacy of the vaccine to build trust in its use while countering anti-vaccine myths and narratives. Pro-vaccine users during this period showed general excitement and optimism about how the vaccine will benefit themselves, their families, and their communities. There were general excite tweets about the authorizations and distributions of the vaccine worldwide. Individuals also spread excite messages about finally getting the vaccine, getting an appointment for the vaccine, or just desiring to get the vaccine. Many users combined these messages with the engage maneuver by taking ownership of the vaccination process by setting the example as a vaccinated individual and encouraging others to also get vaccinated.

Throughout these periods, pro-vaccine communities engaged in hashtag hijacking by tying pro-vaccine narratives to hashtags intuitively associated with anti-vaccine messages. By adding #antivax, #antivaxxer, and other similar anti-vaccine-related hashtags to their tweets, pro-vaccine communities used these hashtags in large numbers to draw attention to pro-vaccine messages with anti-vaccine keywords (see Table 5.4). In one case, they used it to enhance the pro-vaccine messages, typically by attaching this hashtag to pro-vaccine explain messages intended for vaccine hesitant users. In another case, hashtag hijacking tied the hashtag to satirical messages

Table 5.4: Hashtag hijacking: usage count of anti-vaccine–related hashtags by pro-vaccine users.

Hashtag	Before rollout (n=2118), n	During rollout (n=1221), n	After rollout (n=768), n
antivaccination	0	26	5
antivaccine	55	47	68
antivax	457	281	247
antivaxer	5	3	0
antivaxers	26	11	13
antivaxx	83	54	63
antivaxxer	133	39	62
antivaxxers	1459	760	310

related to anti-vaccine individuals' actions. The pro-vaccine message distorts the anti-vaccine message with a quote or a reply or somehow ties their narrative to a specific anti-vaccine incident. Furthermore, in some uses of the hashtags, pro-vaccine users engaged anti-vaccine users to condemn or insult them for either spreading disinformation or other anti-vaccine behavior.

Anti-Vaccine Communities—Narrative Maneuvers

In the week leading up to the vaccine approval and distribution, users were already expressing their COVID-19 anti-vaccine views on social media. The most popular types of messages were the emotionally appealing dismay messages about the side effects of the vaccine. Anti-vaccine users shared messages about how the vaccine causes female infertility, destroys the immune system, or leads to death. These messages were further enhanced with references to scientists, doctors, former Pfizer representatives, and politicians. In many of these messages, the side effects were explained using plausible arguments and pseudoscientific methods and information. To counter pro-vaccine messages, anti-vaccine proponents attempted tactics such as dismissing the vaccine's effectiveness, suggesting that a person's immune system is more than sufficient against the virus. They also countered with the distract maneuver, which uses misdirection by making other topics seem more important. In one example, the 3 US Presidents volunteering for the vaccine mentioned earlier sought to build confidence. However, opponents of the vaccine focused on distracting their audiences with negative political news from the Presidents' pasts relations with China. These insinuated negative links between the Presidents and the country where the virus first began to spread. In another narrative, messages specifically targeting pro-life supporters described the use of fetal cells derived from an abortion during the vaccine development process. These began as dismay messages to anger pro-life supporters about its use and then was supported with explanations on how the different vaccines used the fetal cells in different phases of the process, making some vaccines more ethical than the others. Proponents then enhanced these messages by attaching supportive messages from major religious organizations.

During the rollout, anti-vaccine narratives continued to emphasize many of the negative aspects of the vaccine. Still, many dismaying messages about the vaccine side effects dominated anti-vaccine conversations. New explaining messages to support these dismaying claims included citing the vaccines' published lists of adverse effects and a cost-benefit analysis on the benefits versus the severe reactions resulting from getting the vaccine. Again, these were enhanced and validated with statements from medical professionals and scientists. Additional dismaying messages emerged as popular during this early period. Topics included news reports for the vaccine causing false positives for HIV, government cautions for allergic reactions to the vaccines, claims that the vaccine is not Halal certified under Islamic dietary laws, and negative experiences from those who participated in the vaccine trials. Distort messages, or discussion that alters the main message, helped anti-vaccine messages counter many positive pro-vaccine narratives and propagate anti-vaccine conspiracy theories. They countered the scientific facts about the construct of the vaccine with the lack of peer-reviewed literature to support it, spread manipulated images of Dolly Parton purporting that the vaccine caused her to have Bell's palsy, and suggested that the mRNA vaccines contain nanobots and can change a person's DNA. Furthermore, general anti-vaccine messages from both medical and nonmedical users within this community engaged online to express their distrust with the vaccine and recommend not getting the shot without knowing the long-term safety data.

After the rollout, the decrease in anti-vaccine users resulted in spreading fewer anti-vaccine messages. By this time, Twitter removed many anti-vaccine users and their messages for violating their policy on spreading false or misleading COVID-19 vaccine information. The messages that remained, however, were still primarily dismaying and distorting messages about the adverse side effects and deaths resulting from vaccinations. Despite the Twitter policy, several distorting conspiracy theories such as vaccines connecting one's body to cryptocurrency and altering DNA still appeared in the data set. Furthermore, users continued to engage their audiences more practically by expressing hesitancy for a quickly developed vaccine without data on its long-term effects.

BEND—Network Maneuvers

Instances of the communities engaging in network maneuvers were identified within the messages. Network maneuvers alter the structure of the network by encouraging connections or disconnections between users. In Twitter, one effective tool and indication of a network maneuver is the use of mentions. These types of maneuvers, however, can exist without them.

There were several ways that pro-vaccine communities engaged in network maneuvers. The most common maneuvers were building and boosting, used for creating a group or to grow the size of a group, respectively. The primary goal for pro-vaccine communities was to urge others to get the vaccine under the premise that the more people who supported and received the vaccine, the sooner the pandemic would end. Simultaneously, these groups engaged in the counter maneuver of narrowing and neglecting to reduce the size of or marginalize the opposing anti-vaccine community. One of the most effective actions for group reduction was using the nuke maneuver to dismantle or show the appearance of a dismantled anti-vaccine community. Twitter attempted this maneuver when it created its policy against COVID-19 vaccine disinformation, affecting the entire after-rollout period data set. Another common network maneuver was backing, which is an action that increases the importance of leaders or creates new leaders. Pro-vaccine users showed support for government officials, leaders in the medical field, health organizations, and vaccine manufacturers with positive messages and references to these leaders or organizations.

The anti-vaccine community conducted similar network maneuvers to the pro-vaccine community. They aimed to build and boost their group and reduce the pro-vaccine community using narrowing and neglecting. This community, however, did not have as many leaders as its opponents. The few that they backed included critics of pro-vaccine policies such as an ex-Pfizer vice president, politicians, and scientists who petitioned against the vaccine for safety concerns. Anti-vaccine users, however, had a large selection of leaders that they attempted to neutralize or decrease in importance. These opposing leaders were largely the same people and organizations the pro-vaccine community backed.

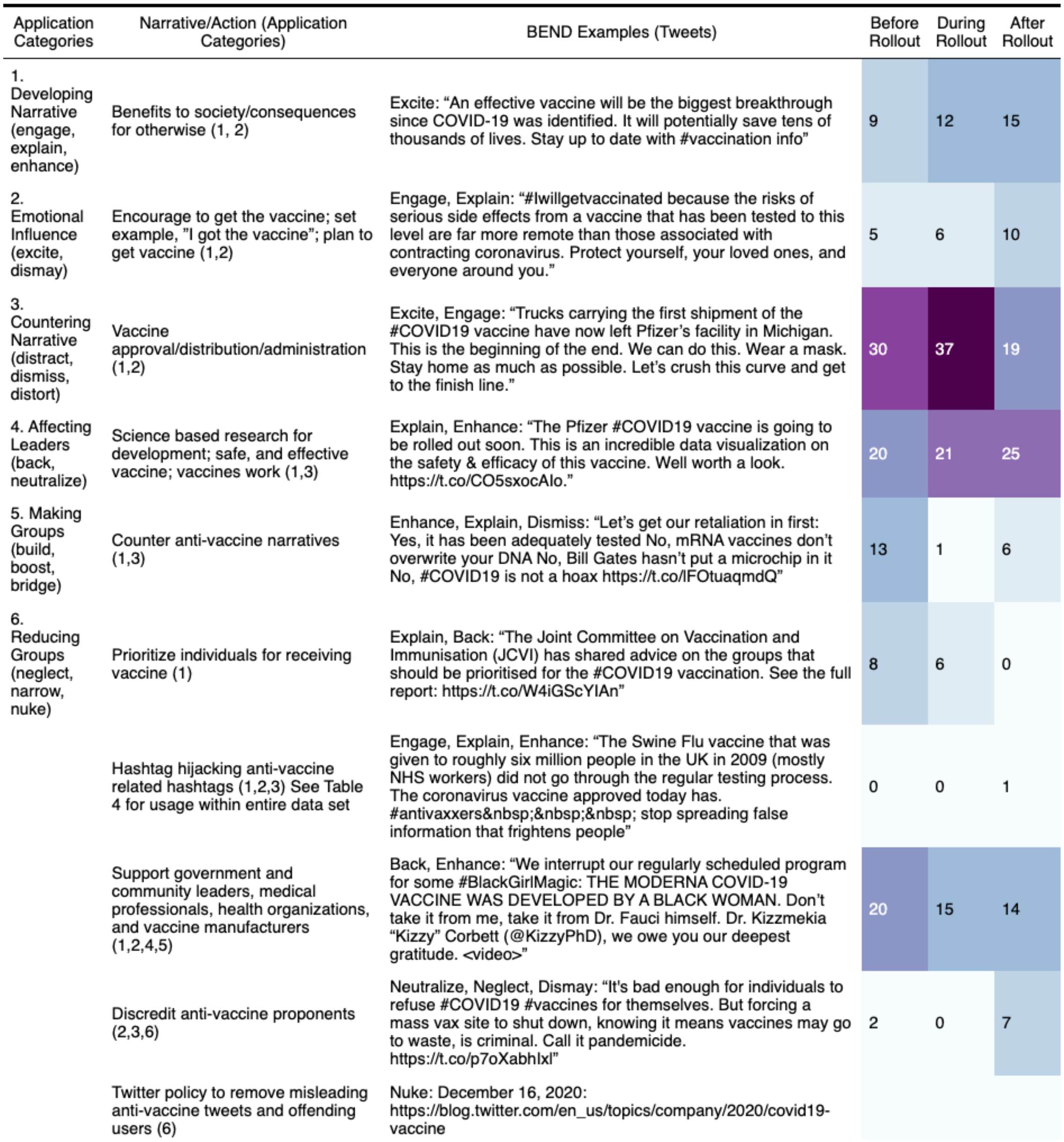
Social-Cyber Maneuvers Applications Over Time

The different narratives and BEND maneuvers from each stance community were associated with one or more application categories. Though each community used different content for their messaging, they used roughly the same techniques. Many of these techniques are used in combination over time to develop more impactful influence campaigns.

Over time, pro-vaccine communities consistently used mostly positive narrative content in their messaging while applying a pattern of developing their narrative, using emotional influence, and countering the opposing community's narrative as shown in Figure 5.3. Excite messages combined with the narratives about the approval, distribution, and administration of the vaccine were among the highest types of propagated tweets among the 3 periods. The second highest categorization of tweets developed the narrative of creating the vaccine using science-based research to build confidence in its safety and effectiveness. This narrative was also used as content to counter anti-vaccine narratives that were often fake or pseudoscience. Furthermore, it was found that directly countering anti-vaccine narratives became less common over time as the number of anti-vaccine messages decreased and the primary pro-vaccine narratives became prevalent.

Because the anti-vaccine community did not have prominent leaders to neutralize, the pro-vaccine community regularly focused on backing the leaders within their own community, such as government officials, health organizations, and vaccine manufacturers. The pro-vaccine community also attempted to expand their group by using hashtag hijacking. Though few instances occurred within the top 100 most propagated tweets, hashtag hijacking still occurs throughout the data set, as shown in 5.4. Finally, although there were many attempts to reduce the anti-vaccine community through varying narratives, Twitter's policy to counter anti-vaccine disinformation before the last time period created the most apparent change in network structure. The result was the decrease in accounts that typically propagated offensive disinformation and subsequently large amounts of anti-vaccine tweets.

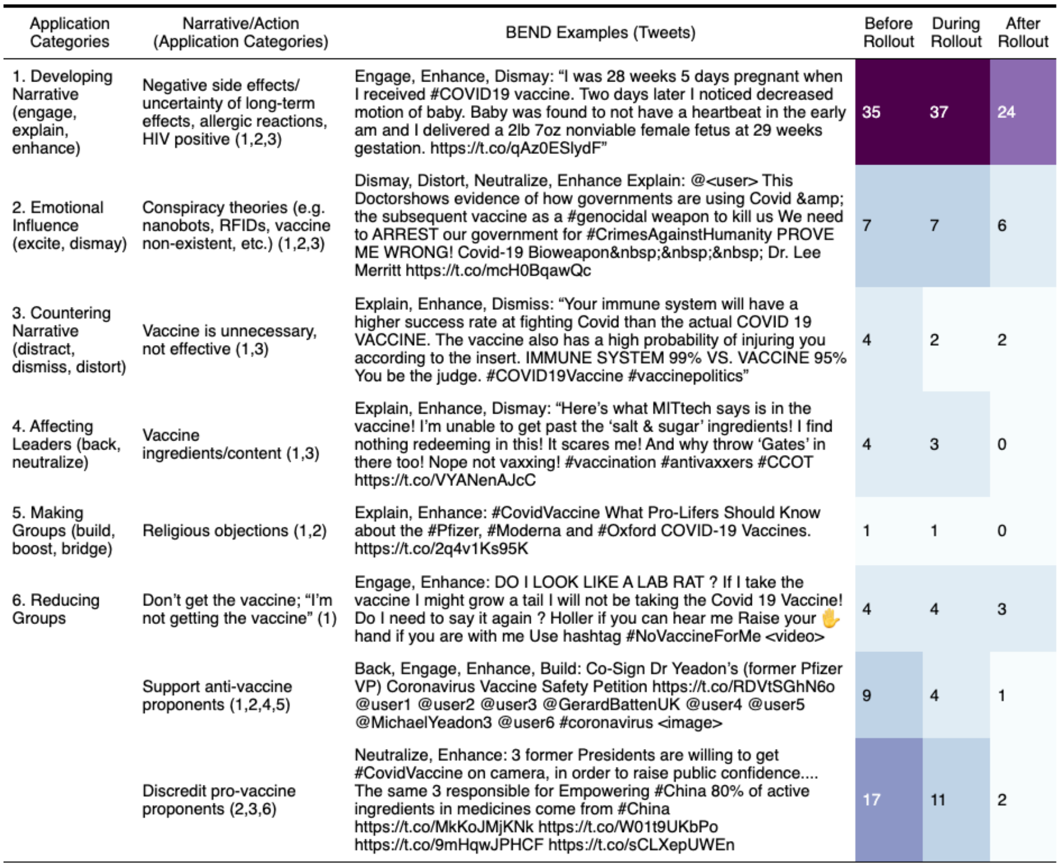
Across the 3 time periods, the anti-vaccine community also used different combinations of the maneuvers and applications to sway their audience as shown in Figure 5.4. They primarily developed their narrative and countered pro-vaccine messages using multiple maneuvers, heavily relying on the negative emotional influence using dismaying messages to highlight the side effects, long-term effects, and conspiracy theories. Anti-vaccine users also aimed to affect the relationships of leaders of both communities. They aimed to discredit the leaders of the pro-vaccine community while highlighting the negative messaging of medical professionals and scientists within their own community. During the first 2 periods, these themes appeared consistently. However, after the rollout following the removal of anti-vaccine accounts and messages,



Application Categories	Narrative/Action (Application Categories)	BEND Examples (Tweets)	Before Rollout	During Rollout	After Rollout
1. Developing Narrative (engage, explain, enhance)	Benefits to society/consequences for otherwise (1, 2)	Excite: “An effective vaccine will be the biggest breakthrough since COVID-19 was identified. It will potentially save tens of thousands of lives. Stay up to date with #vaccination info”	9	12	15
2. Emotional Influence (excite, dismay)	Encourage to get the vaccine; set example, ”I got the vaccine”; plan to get vaccine (1,2)	Engage, Explain: “#Iwillgetvaccinated because the risks of serious side effects from a vaccine that has been tested to this level are far more remote than those associated with contracting coronavirus. Protect yourself, your loved ones, and everyone around you.”	5	6	10
3. Countering Narrative (distract, dismiss, distort)	Vaccine approval/distribution/administration (1,2)	Excite, Engage: “Trucks carrying the first shipment of the #COVID19 vaccine have now left Pfizer’s facility in Michigan. This is the beginning of the end. We can do this. Wear a mask. Stay home as much as possible. Let’s crush this curve and get to the finish line.”	30	37	19
4. Affecting Leaders (back, neutralize)	Science based research for development; safe, and effective vaccine; vaccines work (1,3)	Explain, Enhance: “The Pfizer #COVID19 vaccine is going to be rolled out soon. This is an incredible data visualization on the safety & efficacy of this vaccine. Well worth a look. https://t.co/CO5sxocAIo.”	20	21	25
5. Making Groups (build, boost, bridge)	Counter anti-vaccine narratives (1,3)	Enhance, Explain, Dismiss: “Let’s get our retaliation in first: Yes, it has been adequately tested No, mRNA vaccines don’t overwrite your DNA No, Bill Gates hasn’t put a microchip in it No, #COVID19 is not a hoax https://t.co/lFOtuaqmdQ”	13	1	6
6. Reducing Groups (neglect, narrow, nuke)	Prioritize individuals for receiving vaccine (1)	Explain, Back: “The Joint Committee on Vaccination and Immunisation (JCVI) has shared advice on the groups that should be prioritised for the #COVID19 vaccination. See the full report: https://t.co/W4iGScYIAn”	8	6	0
	Hashtag hijacking anti-vaccine related hashtags (1,2,3) See Table 4 for usage within entire data set	Engage, Explain, Enhance: “The Swine Flu vaccine that was given to roughly six million people in the UK in 2009 (mostly NHS workers) did not go through the regular testing process. The coronavirus vaccine approved today has. #antivaxxers stop spreading false information that frightens people”	0	0	1
	Support government and community leaders, medical professionals, health organizations, and vaccine manufacturers (1,2,4,5)	Back, Enhance: “We interrupt our regularly scheduled program for some #BlackGirlMagic: THE MODERNA COVID-19 VACCINE WAS DEVELOPED BY A BLACK WOMAN. Don’t take it from me, take it from Dr. Fauci himself. Dr. Kizzmekia “Kizzy” Corbett (@KizzyPhD), we owe you our deepest gratitude. <video>”	20	15	14
	Discredit anti-vaccine proponents (2,3,6)	Neutralize, Neglect, Dismay: “It's bad enough for individuals to refuse #COVID19 #vaccines for themselves. But forcing a mass vax site to shut down, knowing it means vaccines may go to waste, is criminal. Call it pandemicide. https://t.co/p7oXabhlxl”	2	0	7
	Twitter policy to remove misleading anti-vaccine tweets and offending users (6)	Nuke: December 16, 2020: https://blog.twitter.com/en_us/topics/company/2020/covid19-vaccine			

Figure 5.3: Social-cyber maneuvers and narratives for top 100 most propagated pro-vaccine tweets.

negative side effect type of messages emerged as the dominant narrative. Because of the decrease in messages, it would be difficult to speculate about the types of anti-vaccine maneuvers that may have otherwise prevailed during the later period.



Application Categories	Narrative/Action (Application Categories)	BEND Examples (Tweets)	Before Rollout	During Rollout	After Rollout
1. Developing Narrative (engage, explain, enhance)	Negative side effects/ uncertainty of long-term effects, allergic reactions, HIV positive (1,2,3)	Engage, Enhance, Dismay: “I was 28 weeks 5 days pregnant when I received #COVID19 vaccine. Two days later I noticed decreased motion of baby. Baby was found to not have a heartbeat in the early am and I delivered a 2lb 7oz nonviable female fetus at 29 weeks gestation. https://t.co/qAz0ESlydF”	35	37	24
2. Emotional Influence (excite, dismay)	Conspiracy theories (e.g. nanobots, RFIDs, vaccine non-existent, etc.) (1,2,3)	Dismay, Distort, Neutralize, Enhance Explain: @<user> This Doctorshows evidence of how governments are using Covid &amp; the subsequent vaccine as a #genocidal weapon to kill us We need to ARREST our government for #CrimesAgainstHumanity PROVE ME WRONG! Covid-19 Bioweapon Dr. Lee Merritt https://t.co/mcH0BqawQc	7	7	6
3. Countering Narrative (distract, dismiss, distort)	Vaccine is unnecessary, not effective (1,3)	Explain, Enhance, Dismiss: “Your immune system will have a higher success rate at fighting Covid than the actual COVID 19 VACCINE. The vaccine also has a high probability of injuring you according to the insert. IMMUNE SYSTEM 99% VS. VACCINE 95% You be the judge. #COVID19Vaccine #vaccinepolitics”	4	2	2
4. Affecting Leaders (back, neutralize)	Vaccine ingredients/content (1,3)	Explain, Enhance, Dismay: “Here’s what MITtech says is in the vaccine! I’m unable to get past the ‘salt & sugar’ ingredients! I find nothing redeeming in this! It scares me! And why throw ‘Gates’ in there too! Nope not vaxxing! #vaccination #antivaxxers #CCOT https://t.co/VYANenAJcC	4	3	0
5. Making Groups (build, boost, bridge)	Religious objections (1,2)	Explain, Enhance: #CovidVaccine What Pro-Lifers Should Know about the #Pfizer, #Moderna and #Oxford COVID-19 Vaccines. https://t.co/2q4v1Ks95K	1	1	0
6. Reducing Groups	Don’t get the vaccine; “I’m not getting the vaccine” (1)	Engage, Enhance: DO I LOOK LIKE A LAB RAT ? If I take the vaccine I might grow a tail I will not be taking the Covid 19 Vaccine! Do I need to say it again ? Holler if you can hear me Raise your ✋ hand if you are with me Use hashtag #NoVaccineForMe <video>	4	4	3
	Support anti-vaccine proponents (1,2,4,5)	Back, Engage, Enhance, Build: Co-Sign Dr Yeadon’s (former Pfizer VP) Coronavirus Vaccine Safety Petition https://t.co/RDVtSGhN6o @user1 @user2 @user3 @GerardBattenUK @user4 @user5 @MichaelYeadon3 @user6 #coronavirus <image>	9	4	1
	Discredit pro-vaccine proponents (2,3,6)	Neutralize, Enhance: 3 former Presidents are willing to get #CovidVaccine on camera, in order to raise public confidence.... The same 3 responsible for Empowering #China 80% of active ingredients in medicines come from #China https://t.co/MkKoJMjKNk https://t.co/W01t9UKbPo https://t.co/9mHqwJPHCF https://t.co/sCLXepUWEn	17	11	2

Figure 5.4: Social-cyber maneuvers and narratives for top 100 most propagated anti-vaccine tweets.

5.2.5 Discussion

Principal Findings

The results in this study showed the differing characteristics of pro-vaccine and anti-vaccine communities on Twitter and how they manipulated narratives and the online network structure to convince users whether to vaccinate against COVID-19. Both groups of users used many of the same approaches but varied in the extent to which they applied each maneuver. They sought to build their communities while, at the same time, attempting to reduce the size of the opposing community by maneuvering as appropriate to fit their narratives. Pro-vaccine supporters tweeted excite and explain messages to encourage vaccination, whereas anti-vaccine users relied on the negative dismay and distort messages with narratives related to side effects and death. Pro-vaccine users also backed the prevalent leaders within their group, of whom anti-vaccine users targeted and attempted to neutralize. Furthermore, platform policies showed their ability to effectively nuke the anti-vaccine community by reducing the size of the online community and the quantity of anti-vaccine messages. The majority of top super spreaders and super friends for each of the analyzed time periods were pro-vaccine users. Government leaders, medical organizations and professionals, vaccine manufacturers, and, in the later period, less-mainstream community leaders emerged as pro-vaccine leaders, effectively reaching a higher number of users and engaging in more 2-way conversations. Additionally, bots had a sizeable presence within each community. A larger percentage of bots within the anti-vaccine community than the pro-vaccine community were observed, and among the top 100 key influencers for each community over time, the anti-vaccine community had more bots as super spreaders and super friends. Before Twitter's vaccine disinformation policy, they reached as high as 24% and 16% of the top 100 anti-vaccine super friends and super spreaders, respectively, before the rollout. The anti-vaccine community utilized bots to a greater extent to build their communities and spread their narratives than did the pro-vaccine community, effectively positioning them as key influencers among anti-vaccine users.

Pro-vaccine users repeated many of the same maneuvers throughout each time period, varying in different narratives that emerged as the rollout occurred. Many of the maneuvers were positive or growth-type maneuvers. They developed narratives around science-based facts explaining the safety and effectiveness of the vaccine and emotionally influencing narratives that excited their audiences about the health and societal benefits of everyone receiving the vaccine and, to a lesser extent, dismayed them with the fatal consequences of not vaccinating. Many of the topics used to develop these narratives were also used to counter anti-vaccine messages. As many anti-vaccine conversations revolved around side effects and conspiracy theories, pro-vaccine proponents rebutted with facts to explain the errors in these messages. These narratives were used consistently by highly connected leaders within the community and actors that maintained a high profile apart from Twitter, such as government leaders, news organizations, and medical professionals. Pro-vaccine users tended to back government officials and health organizations with positive messaging to build confidence in the proponents of the vaccine as well as the vaccine itself.

Throughout each of the time periods, the anti-vaccine community used similar maneuvers to those the pro-vaccine community used but with a greater frequency of the negative or reducing-type maneuvers. They developed focused narratives and countered pro-vaccine messaging to increase hesitancy and doubt. Their most consistent technique was using dismaying messages of the adverse side effects, the uncertainty of the long-term effects, and vaccine deaths. Conspiracy theories about the vaccine also added to the anti-vaccine narrative. Users attempted to neutralize pro-vaccine leaders by discrediting them and their associated messaging, and they backed leaders who criticized the vaccine and encouraged others not to get the vaccine.

Finally, as the host for the pro-vaccine and anti-vaccine engagements, Twitter is in a unique position with the ability to filter the discussion on their platform. Their policy to remove misleading and false anti-vaccine nuked the anti-vaccine community by significantly reducing the number of anti-vaccine users and tweets that had grown at the time of the rollout. The social media site made a policy to fight disinformation, which resulted in supporting the pro-vaccine effort to reduce the size of the anti-vaccine community and messaging.

Limitations

One major limitation was the ability of the stance detection to separate the nodes into pro-vaccine and anti-vaccine communities. First, many pro-vaccine maneuvers use neutral hashtags. Second, anti-vaccine mean confidence levels for each time period were lower than those for pro-vaccine, with the after-rollout data only reaching as high as 67%, even after multiple iterations of hashtag labeling. Third, hashtag latching made stance selection difficult as some hashtags commonly used for pro- or anti-vaccine messages were used to gain the attention of members of the other community. Therefore, further study is required for improving the separation between pro- and anti-vaccine agents and tweets.

Another limitation is the ability of ORA-PRO to detect BEND maneuvers. This required manual verification of select entities within the data set. Newer versions of ORA-PRO continue to refine the metrics to better identify some of the maneuvers, especially network maneuvers, that occur over time. The results of this study inform the specifications and thresholds for improving the software.

Additionally, the demographics of Twitter users may have a bias that results in larger numbers of pro-vaccine versus anti-vaccine users and data [142]. However, in this study, we focus on the content of the messages and the characteristics of the individually stanced groups over time. Furthermore, we analyze how the maneuvers and influence spread across the social media environment, and the user demographics are part of that environment. For example, the study is not about how the different communities spread their influence in the real world but how they do this on Twitter.

Finally, many tweets and users that existed during the initial data collection were either deleted or suspended due to Twitter Rules violations. This made it difficult for observing historical tweets with their associated images and videos as well as visualizing tweets with replies and mentions within the Twitter environment.

5.2.6 Conclusion

This case study analyzed how pro-vaccine and anti-vaccine communities around the initial COVID-19 vaccine administration attempted to persuade others of their stance under the BEND maneuvers framework. The BEND maneuvers enabled an examination of the different techniques used by each community. This included observations of the actions of different types of key actors within the different groups and an analysis of their varying techniques. Additionally, the main concepts and messages of tweets tweeted within each community and the extent they acted as each of these social-cyber maneuvers were explored. Furthermore, these maneuvers were combined into application categories to gain a macro-level understanding of how the maneuvers were used in combination as an overarching influence campaign over time.

Real-world events influence online discussions, and over time, the changes in these conversations reflect changes in beliefs. In this case, the efforts of these 2 communities can lead users to either vaccinate themselves against COVID-19 or not, possibly changing the direction of the pandemic. Future work should look at how these changes in beliefs mobilize into changes in behavior. Furthermore, though many influencing actions result from users interacting with other users, the policies that govern the use of social media can impact the size of a community and their ability to spread their narrative throughout the network. Therefore, research to regularly detect and evaluate the effectiveness of social-cyber maneuvers and make pointed network structure alterations based on specific narratives is needed to understand the consequences of different interventions and implement better policies to impact influence campaigns on social media.

5.3 2022 US Midterm Election Case Study

5.3.1 Introduction

While social media allows for connections between friends and families and the expansion of a high-speed and far-reaching social network, each person online becomes inescapably a target of influence. From commercial advertising and political agendas to health habits and family feuds, users are persuaded to take sides on various issues, support different groups, or take some action. With online platforms like Twitter and Facebook, people can more easily and efficiently connect with others of similar ideologies, create group discussions surrounding specific topics, and expand the reach of their topics across the globe in a matter of seconds.

Online influence operations have shaped the world, often in shocking and unexpected ways. In 2011, the Egyptian Revolution initiated on Facebook and fueled by the online community empowered a youthful generation in the overthrowing of a sitting president [133]. Democratic countries witnessed unanticipated electoral wins in 2016 with the selection of U.S. President Donald J. Trump and the United Kingdom's departure from the European Union, also known as Brexit [141] suggesting the unmonitored levels of influence undetected by mainstream media. In 2020, while the world closed businesses, schools, and major events during a global pandemic, online COVID-19 anti-vaccination campaigns gave rise to increased numbers of false vaccine information, conspiracy theories [33], and negative preventative COVID-19 actions [5] propagated online. Today, a social media war between Russia and Ukraine runs as a parallel front to the physical war itself as both countries strive to rally support from both home and abroad [115]. In this paper, we use the recently held 2022 US midterm elections, which occur between presidential elections, as a case study for a framework for assessing communities participating in online influence operations.

The battle for the 2022 elections

The 2020 US Presidential Elections culminated in a clash between two major influence campaigns over political power. President Trump, the incumbent and leader of the right-wing Republican party, fought to maintain his position as head of the country while former Vice President Joseph R. Biden, a left-wing Democrat, vied to take that position away from him. Each candidate's party sought to spread their narratives across social media to gain support for their candidates and policies. Among these online communications, manipulation occurred through the use of bots and distorted narratives [55].

When Biden emerged as the winner after a long week of counting ballots, members of the Republican party amplified narratives of a "rigged" election and electoral fraud, a theme that started before election day [58]. Trump became a sounding board for misleading post-election information. These efforts took root and spread violently across social media as disinformation campaigns [9, 15]. On January 6, 2021, Trump supporters, mobilized by social media, gathered and stormed the U.S. Capitol to prevent Biden's inauguration [92].

Following the 2020 election controversy, the 2022 midterm election witnessed a larger than normal turnout, particularly in young voters, for both gubernatorial and congressional races [41, 83]. Republicans flocked to the polls to regain their dominance in the House of Representatives and in the Senate. Democrats, likewise, sought to defend their seats in the House while aiming to gain more influence in the upper chamber. Both parties competed for governor seats in 36 state races across the country, which had become increasingly important as the Supreme Court had begun delegating power back to the states. In June earlier that year, the Supreme Court decided on the controversial ruling regarding the case of Roe vs. Wade, which gave state governments the liberty to set their own abortion policies [117].

One of the key focuses of each party's political objectives was to gain support from the battleground–or swing states. In swing states, campaigns typically use a larger number of funds and campaigning resources to garner votes from the constituents of these states where the overall political stance is uncertain. In this paper, we focus on Twitter conversations and networks aimed at these swing states and the right-leaning and left-leaning users who represent varying degrees of the Republican and Democratic parties, respectively. These two political sides represent the distinct online communities engaging in influence operations analyzed in this study.

Influencing as a community

Communities of users have the ability to efficiently create large-scale manipulation of online users. The larger the community, the greater the potential diffusion of an ideology, belief, or piece of information.

Multiple methods have been developed to identify online communities. One technique is to simply apply the Leiden clustering algorithm on the interaction network of user accounts [126]. This generally creates groups of agents, or actor nodes, who communicate with each other regardless of the topic. Iterative Vertex Clustering and Classification (IVCC) is an approach that applies community detection with the multiple data structures within a social network. This technique was used to identify online extremist communities spreading ISIS propaganda. In the work by Hristakieva et al. [69], they combined the identification of coordinated behavior within communities with propaganda during the 2019 UK general election, which lead to further insights on online behavior .

For very large data sets, using previously mentioned network algorithms to identify communities can be computationally infeasible or noisy. This makes it difficult to derive useful information from the clusters. In the case study used in this paper, a community was created based on user attributes or similar backgrounds. These communities were further defined by identities, stance, bot or not bot, and other characteristics. In our study, we created two communities of agents characterized by their political slant - democratic (left-wing) or republican (right-wing). This allowed us to focus on the communities as large ideological entities and their inherent tendencies to use various influence techniques.

The method applied in this paper, the BEND Framework, extends the analysis process and applies a more systematic method for quantitatively and qualitatively characterizing and assessing influence operations [17, 35]. In this paper, we apply the BEND framework and follow its Twitter-based pipeline for data processing and its multi-stepped iterative process to examine the dynamics and influence techniques of both left-wing and right-wing communities.

We seek two primary objectives for this paper. The first is to illustrate the BEND framework and expand on it as a methodology for analyzing influence operations, specifically in the comparison of two opposing communities. The value in the illustration of this framework is to comprehensively analyze the communities while evaluating actors, narratives, the social network, and the social-cyber maneuvers they engage in to achieve their community's goals. The second is to use the BEND framework to conduct an overtime comparative analysis of the influence campaigns from left-wing and right-wing communities in the week leading up to the 2022 US midterm elections.

5.3.2 Methodology

The methodology consists of data collection, data preparation, and community analysis and assessment using the BEND framework. An overview of the process is shown in Figure 5.5.

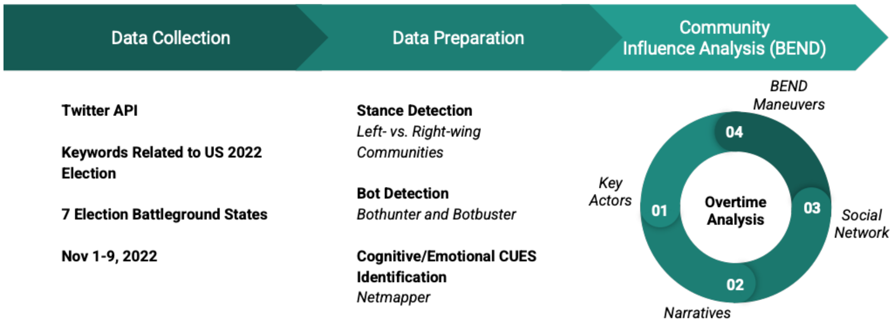


Figure 5.5: Methodology for Analyzing Online Community Influence - US Election 2022

Data collection

The 2022 midterm election data was collected using the Twitter API v2. The dates for the collection range from November 1-9, 2022, which includes the week leading up to November 8 election day. The collection extends to Nov 9 to include any tweets that may have continued to propagate immediately following the election. Data was collected using key terms related to major races and candidates for seven battleground states based on the work conducted by FiveThirtyEight [121]. The states included Arizona, Georgia, Nevada, North Carolina, Ohio, Pennsylvania, and Wisconsin. The purpose was to focus on the states that could potentially have higher use of influence maneuvers. The list of terms for collection can be found in Appendix A.1. Across the nine days, there were 257,758 unique users and 640,559 tweets.

Data preparation

Data preparation consists of deriving attributes from the nodes to structure the data to adequately conduct the influence analysis in the final section of the methodology.

Stance detection

We applied two stance detection algorithms to divide the data set into separate ideological communities and assign stances to hashtags. To identify the different stanced agents, we used the

Twitter Stance Propagation Algorithm, which propagates stance labels through user-similarity-weighted interaction networks [140]. Afterward, we applied a second algorithm to identify the stances of hashtags. This used a similar method of influence propagation through the user communication network. [80]. During the initial setup for both, several hashtags in the data set were labeled as either left-leaning or right-leaning, resulting in the complete classification of both agents and hashtags within the data set. We ran these algorithms on the combined data sets for all nine days using the classified hashtags found in Appendix A.2. Following the stance detection, we found 75,545 (29%) accounts classified as right-wing, 70,496 (27%) classified as left-wing, and 111,717 (43%) as unassigned.

Bot detection

In an effort to gain an understanding of the type of users engaging in political influence within this data set, we also identified automated user accounts, also known as bots. Bots are accounts that autonomously interact with users by tweeting, retweeting, following other accounts, or other platform actions. Though bots can serve useful purposes [54], like quickly reaching a large online population with pertinent emergency information, some social bots are known to spread influence and disinformation maliciously [120]. We used the overlapping results of two bot detection algorithms: Bothunter and Botbuster [16, 96, 97]. BotHunter uses a random forest machine learning algorithm while BotBuster uses a mixture of machine learning algorithms to differentiate bots and human accounts. For this study, if either bot detector identified an agent as a bot, the agent was labeled as a bot.

Cognitive/Emotional CUES identification

Cognitive/Emotional CUES (Cyber-mediated Usable Emotional Sensors) are subconsciously-used indicators found within text-based messages, such as tweets, that can signal a person's cognitive and emotional state. These indicators include the addition of emojis, exclamation points, and pronouns to messages, or they may also include the use of absolutist, exclusive, or angry words. Two commercial software were used for extracting and then applying the CUES [39]. We used the Netmapper software to generate the CUES for the tweets within our data set. We added the CUES as attributes to the tweets, which when pipelined into the ORA-PRO Software, calculated and detected the occurrence of various BEND maneuvers within the tweets and identified the influencers using them and their targets.

Community influence analysis

Analyzing the communities using the BEND framework is an iterative process of evaluating the key actors, narratives, social network structure, and the BEND maneuvers. Though applying the framework recommends a general order of analyzing specific focus areas of the data, sometimes analysis of one section may require further analysis of a previously reviewed section to derive a more comprehensive assessment of the situation.

Key actor analysis

During key actor analysis, we identified important actors and their actions to help understand their roles as part of the larger influence campaign. In its simplest form, identifying key actors includes determining the influencers and targeted audiences during the campaign. This can also extend to a further evaluation of users with various attributes such as bots, verified users, or news organizations. Verified users, authenticated accounts of public interest, are automatically labeled during the Twitter data collection. These typically include government accounts, celebrities, or other well-known organizations. For this study, the news organizations' labels are from an ongoing list maintained by the CASOS research group and draw from multiple sources, including `https://mediabiasfactcheck.com/`.

Users can also be key actors based on their position in the social network structure. In this paper, we explored high-scoring superspreaders, who are positioned as a highly centralized node within the network. Superspreaders are agents that share content that is spread often, indicating efficient dissemination of information and influence. To calculate superspreaders, we used the ORA-PRO software, which uses various centrality measures, including page rank centrality and K-core, to give a normalized value for the agents within the data set.

Narratives

The next thing we looked at within the data set were the narratives and topics of the political conversations for both parties. Techniques for identifying the narratives include reviewing hashtags associated with each stance, highly propagated messages, and posts that indicated the use of various influence techniques, specifically the BEND maneuvers. We focused on the narratives conveyed through the use of hashtags. Hashtags are used to index or group similar topics so that they can be easily followed. They can serve as a way to follow topics related to each other. Communities of a particular political stance can use hashtags to index the major themes of their party.

Social network analysis

The Twitter social network structure is a meta-network consisting of a network of networks joined by the interconnection of various types of nodes. A connection between agents may occur when one retweets the other creating the subgraph of a retweet network. Similar graphs could include mention networks, where users use the '@' to link someone to one of their posts, or reciprocal networks, where users are connected if they engage in two-way communication. Two-way communication is any combination of retweets, quotes (retweets with user commentary), replies, or mentions. For example, if User A retweets User B, and User B mentions User A, they have engaged in two-way communication.

In this study, we compute various network measures on an all-communication network. The all-communications network is an agent-to-agent network that joins user nodes if they engage in interactions with another user on Twitter. An interaction includes all retweets, quotes, replies, and mentions, and it is bi-directional. This means, for example, that an interaction counts if a user mentions another user and if a user is mentioned-by another user. We use this to characterize how communities interact with users within their own community, users within opposition communities, and users with an unassigned stance.

Calculated network measures of these different networks provide insight into the interaction between the different nodes and high-level analysis of the relationships between actors and other facets of the data, especially if observed over time. Degree centralization for a network uses individual total-degree centrality measures and is the extent to which network has centralized nodes similar to a hub. In a network with high degree centralization, fewer nodes have many connections and many nodes have few or none. The network density and echo-chamberness are useful in gathering information about the level of cohesion and overall internal interactions between group members. Density measures the average number of links per person over the total number of possible links. In terms of the all-communication network, the network density explains the internal cohesiveness between users by how much communication occurs between them. A high echo-chamberness in a community can result in the amplification and reinforcement of ideas without any external influence. The echo-chamberness of a community can be calculated by measuring both density and the number of reciprocal links based on their internal communications.

These agent-to-agent networks are further expanded to include hashtags, messages, tweets, and other defining parts of Twitter. Agents are linked to hashtags if they use a specific hashtag creating an agent-to-hashtag network. Subsequently, hashtags can be joined if they are used in the same post resulting in a hashtag co-occurrence network.

BEND maneuvers

At the heart of the BEND framework are 16 social-cyber maneuvers, as listed in Table 5.5, which make up the acronym for its name. Social-cyber maneuvers are actions taken by actors online to manipulate the social network or the narratives to achieve the desired end state [17].

The maneuvers are divided into community and narrative maneuvers, each with a negative and positive aspect depending on the effect on a metric related to the network or narrative. Community maneuvers are discussions or actions that manipulate the structure of the social network. This may be by mentioning other users in a post to apply a *build* maneuver, which is used to create groups. In a contrasting example, a *neutralize* maneuver can be applied to change the network structure with actions to discredit an opinion leader in an opposing community. In a network, this can be seen as either a reduction of ties from a topic to an actor or an increase in negative discussion surrounding an actor and the topic.

On the other hand, narrative maneuvers focus specifically on the messages and topics being spread and how they are manipulated through the content of the posts. Actors achieve these maneuvers through individual messages or an overarching campaign across multiple messages. For example, an *excite* maneuver can be accomplished through a single tweet that aims to elicit happiness or joy from the recipients. As an example of a maneuver influencing an overarching campaign, *distort* maneuvers can change the perspective of an ongoing narrative through multiple social media posts.

The BEND framework assesses the social-cyber maneuvers from the perspective of the actors, communities, narratives, and social network structures over time to gain a comprehensive evaluation of the influence operations. This is the framework that we use for analyzing the different partisan communities during the election.

Table 5.5: 16 BEND Maneuvers.

	Community	Narrative
Positive	Back	Engage
	Build	Explain
	Bridge	Excite
	Boost	Enhance
Negative	Neutralize	Dismiss
	Nuke	Distort
	Narrow	Dismay
	Neglect	Distract

BEND maneuver categories

In our study, we combined the BEND maneuvers to create a high-level view of how each community attempted to influence its target audience similar to a previous study conducted on influence efforts related to the COVID-19 vaccination [21] This approach resulted in five general categories where the influencers aimed to reduce groups (*neglect, narrow, nuke*), grow groups (*build, boost, bridge*), affect leaders (*back, neutralize*), emotionally influence actors (*excite, dismay*), develop the narrative (*engage, explain, enhanc*e), and counter-narratives (*distract, dismiss, distort*). We used these application categories to evaluate how the maneuvers are used as a means for the communities to achieve their ideological end states.

5.3.3 Results

In the United States, political parties compete on social media by promoting their candidates and ideas while scrutinizing policy platforms of the opposing party. In this case study, we laid out characteristics indicative of online influencers for each community using the BEND framework as a guide. We provided an overview of the social network interactions between and within the communities, compared the influence tactics between each of the communities, and explored the key actors and narratives.

Key actors as superspreaders

Within the data set, we calculated superspreader values for each of the actors. We then identified the verified actors, bots, news agencies, and news organizations in the top 100 superspreaders for each political slant. The values are shown in Figure 5.6. As expected, verified users comprised a large portion of the top superspreaders, whereas bots and news accounts were only a small number. Furthermore, the left-wing and right-wing users maintained relatively the same values for the number of key actors from each of the identified influencer categories. The left-wing community had 29 superspreaders, 22 verified actors, 2 bots, and 2 news agencies. Similarly, the right-wing community had 31 superspreaders, 27 verified actors, 1 bot, and 2 news agencies.

While the top 100 superspreaders were generally balanced across political stances, the top ten superspreaders were democratic. Of the top ten accounts, there were six Democratic candidates, one republican candidate, a former Democratic president, a republican strategist, and a left-wing writer. The top five superspreader accounts in the data set were CheriBeasleyNC, TheOtherMandela, TimRyan, katiehobbs, and staceyabrams - all democratic candidates.

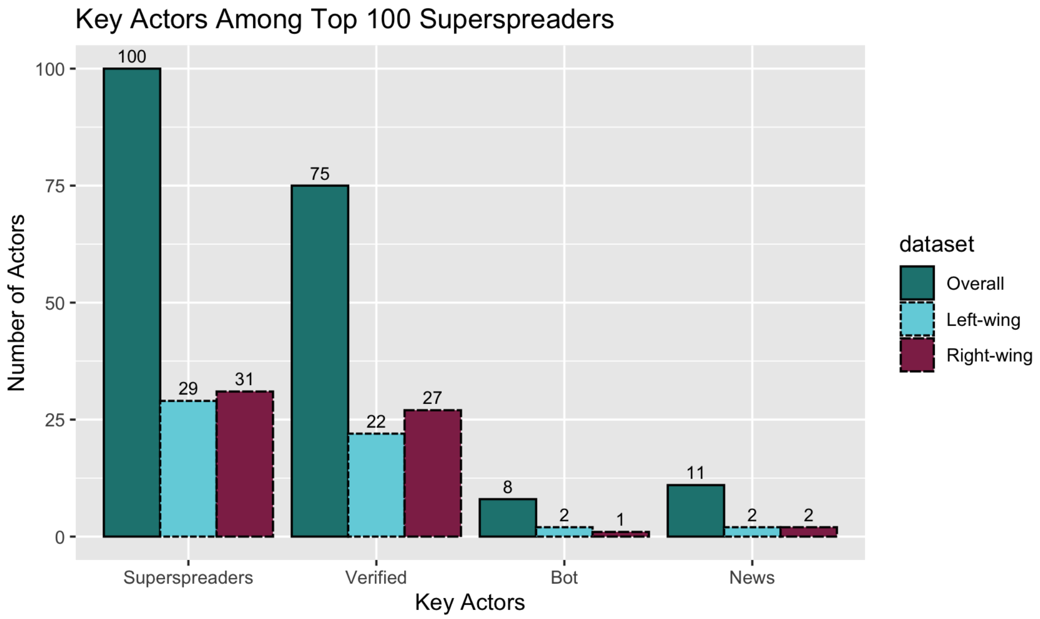


Figure 5.6: Number of key actor types among the top 100 superspreaders for Nov 1-9, 2022

After dividing the group into communities, we observed the data set as a whole interaction network from November 1-9. Figure 5.7 shows the layout of a reciprocal agent network for the time period.

First, this graph shows that larger clusters of reciprocity mostly occurred within each stance community. However, there were multiple instances of reciprocity between members of different communities indicating signs of bridging between the communities. For example, the BarackObama account engaged in two-way communication with user accounts of both stances. He made attempts to contact both communities during the election process with positive messaging in an effort to gain more votes for the democratic party. However, there were also negative cross-party communications among the superspreaders. Users used the platform to negatively call out the opposition or debate their policies and illuminate these shortfalls to their followers.

Additionally, the graph also displays the agent nodes sized by their superspreader value with the top 30 verified superspreaders labeled. We see that many of the top superspreaders are also verified users. Most of these verified users are candidates running for election or high-profile political advocates or supporters.

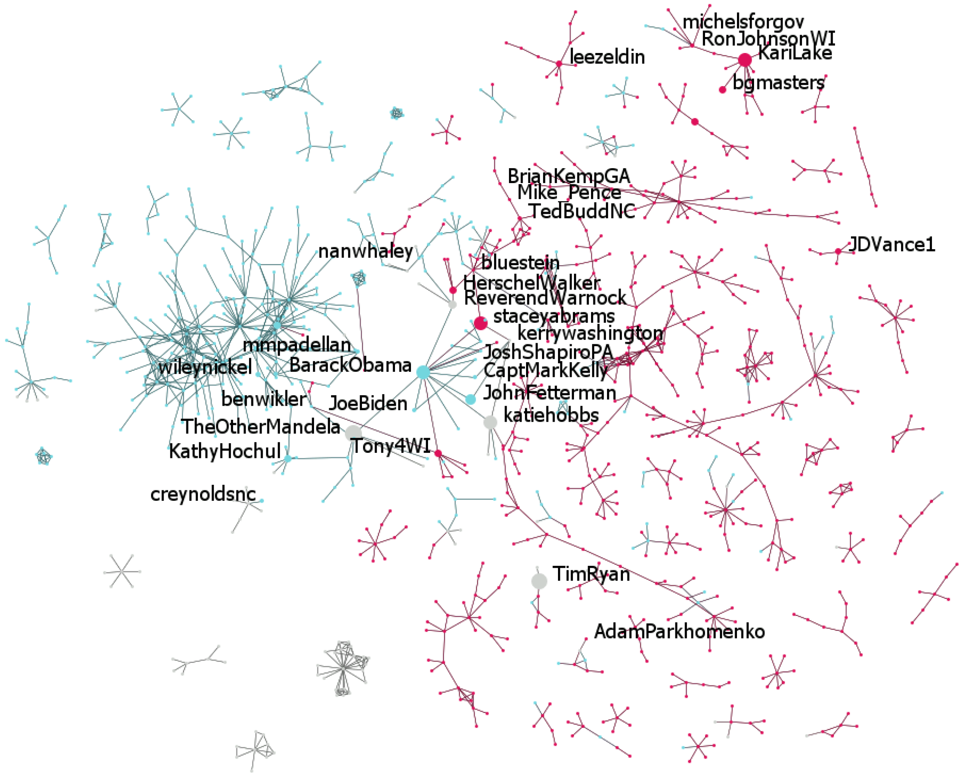


Figure 5.7: Agent x Agent Reciprocal Network for Nov 1-9, 2022. Sized by superspreader. Node color indicates stance. Blue is left-wing; red is right-wing; and gray is unassigned. Top 30 superspreader verified users labeled.

Pushing election narratives with hashtags

Hashtags used among communities of users can provide an idea of the topics discussed. Similar to the agent classification process, the stance detection algorithm categorized the hashtags within the data set as left-leaning or right-leaning. Using this, we identified the top 100 hashtags and their usage for each of the parties across the nine days. Figure 5.8 displays colored by stance and sized by the number of times they are used. We find that both parties used hashtags to convey the primary themes of their respective campaigns.

Left-wing hashtags were consistently used more than right-wing ones. They used general pro-democratic voting themes such as #ProudBlue22, #wtpblue, and #DemVoice1, and anti-republican concepts such as #MAGALiesCostLives and #FreshResists. They also attempted to *back* candidates as well as discredit (or a *neutralize* maneuver) on opposing candidates. Example hashtags include support for #warnock and #Fetterman but also for negatively portraying #AntiJobsJohnson and #NotYourBudd. Furthermore, hashtags provided views on political issues. This includes #ProtectMySocialSecurity and #VoteBlueToProtectWomen. Interestingly, some hashtags that may typically be considered as neutral (eg. #ncpol and #ncvotes22) were closely associated with pro-left-wing rhetoric as they appeared largely in democratic messages.

Right-wing hashtags were used similarly to left-wing hashtags. They aimed for voters to create a #redwave, #voteRed, and #LeadRight, and to save America from Democratic policies with #DemocraticPoliciesKill and #VoteRedToSaveAmerica. They supported their own candidates with #Doug4Gov and #TeamHerschel, and countered democrats with #katiehobbesisaracist and #FireKathyHochul, as examples. Additionally, they used hashtags to emphasize their issues such as their support for the overturn of Roe vs. Wade with #birthcontrol and #abortionhurtswomen and complaints with #gasprices.

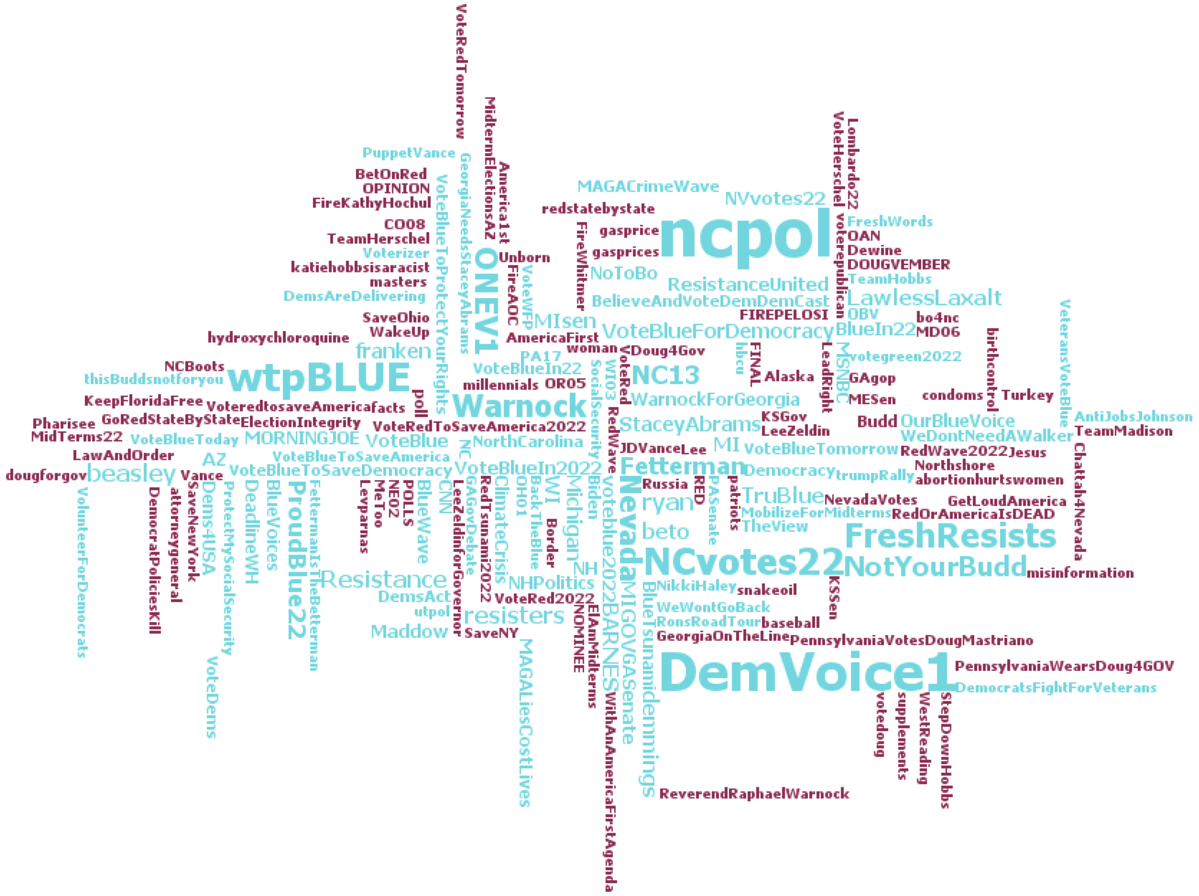


Figure 5.8: Word Cloud of Top 100 Hashtags by Each Stance. Sized by usage count. Blue is left-wing; red is right-wing.

Left-wing and Right-wing networks over time

We then calculated various social network metrics over the course of the time period for each community. This allows us to quantitatively observe various aspects of the social network that exist beyond the actors and their narratives.

In Figure 5.9, we show the number of daily users, hashtags, and tweets of each political stance for the time period. For each of the metrics, the data fluctuated leading up to election day, but the left-wing users unexpectedly had higher values in each of these categories every day. This was surprising given that there were a larger number of unique right-wing than unique left-wing accounts in the combined numbers for all nine days. This means that even though there was a larger number of right-wing users, more left-wing users were actively spreading more messages with more tweets. Right-wing users tended to be more sporadic across the days of the data set, but they existed in larger quantities collectively over time.

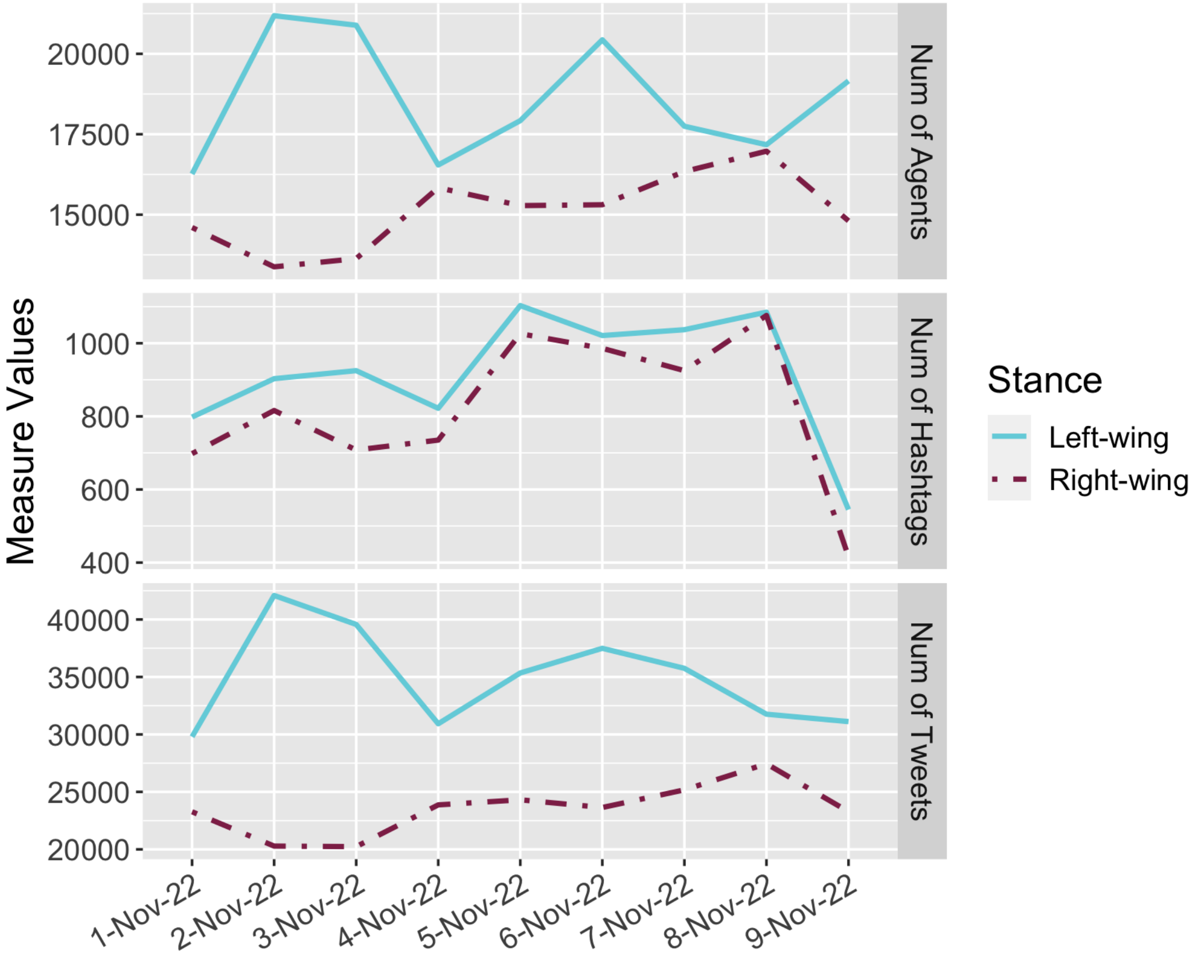


Figure 5.9: Over time description of number of users, hashtags, and tweets by usage.

We then calculated various social network measures on the all-communications networks for each stance every day, shown in Figure 5.10. First, we calculated the degree centralization for the different communities using the interactions from an all-communications network. For our case study, the left-wing network typically has a higher degree centralization indicating they have the tendency to have more individuals acting as focal points for receiving or sending information. We then looked at the users with high degree centrality based on the interactions in the all-communication network. Unsurprisingly, the top accounts of high degree centrality in our data set are political candidates as well as several bots. Democratic senatorial candidate from North Carolina, Cheri Beasley, had the largest value for degree centrality. The republican with the highest degree centrality was Kari Lake, an Arizona gubernatorial candidate and former television news anchor, who appeared sixth in the rankings. Of the top 100 users with the highest total-degree, there is approximately an even distribution between democratic and republican users.

We also calculated the density and the echo-chamberness of the all-communications networks in Figure 5.10. In this graph, the right-wing community consistently had greater density and echo-chamberness. This suggests that right-wing community members are more densely communicative between members, close-knit, and internally facing than the left-wing community. The implication of a network that has high echo-chamberness is that the community members are more likely to influence each other with their shared ideology. Therefore, the right-wing community is more likely to reinforce the biases and ideas amongst each other creating more strongly held beliefs.

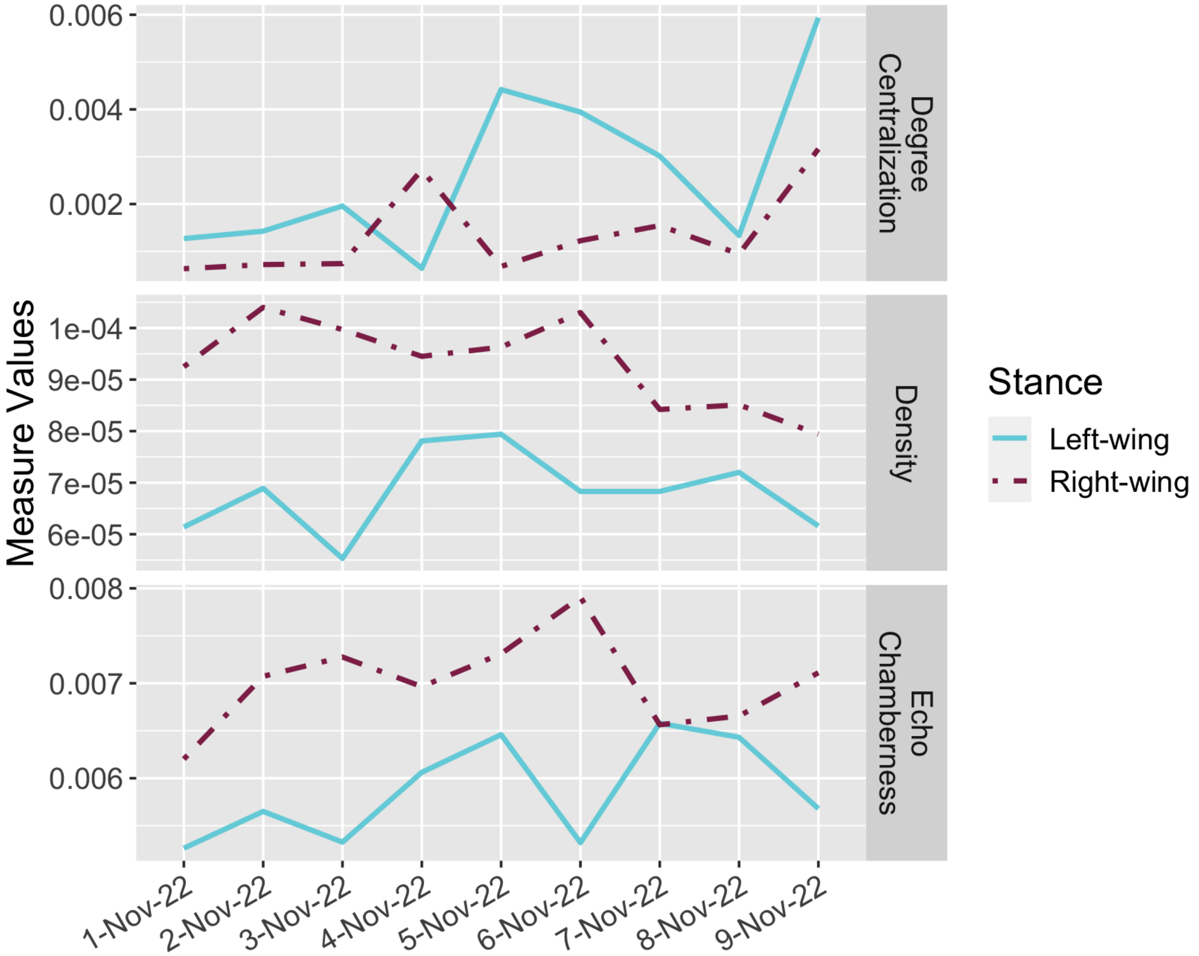


Figure 5.10: Over time network measures by stance.

Executing social-cyber maneuvers during the election

To gain an understanding of the actions of the influence campaigns conducted by each of the ideological groups, we conducted a quantitative analysis of their maneuvers as categorized into different applications. The categories, which are created by combining several maneuvers, consist of reducing groups, growing groups, affecting leaders, emotionally influencing, developing a narrative, and countering a narrative.

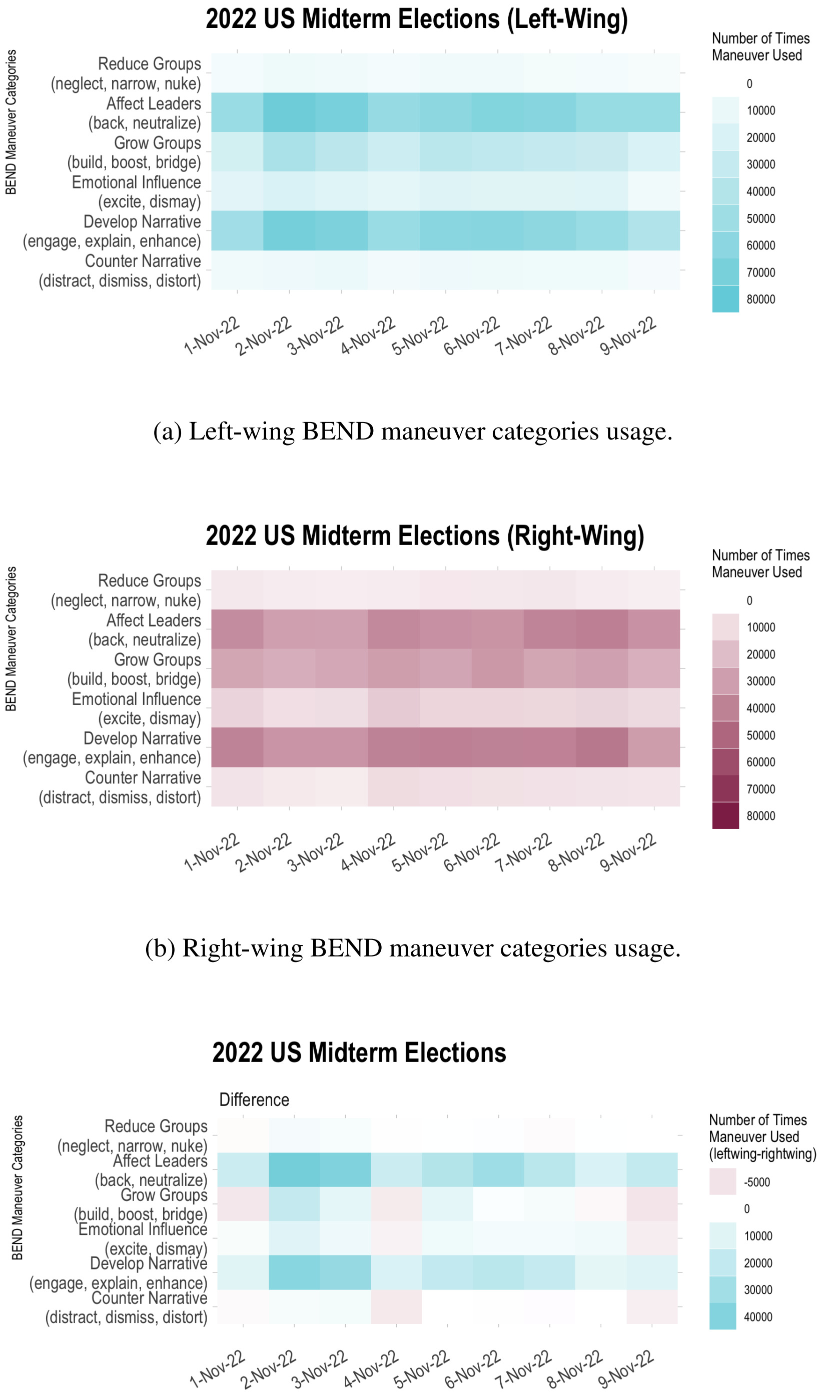
In Figure 5.11, we show the number of times each of the different stances used the various application categories. The values are a summation of the number of times a maneuver from each category was used by a community member. We found that for both stances, there were large amounts of maneuvers that affected leaders, grew groups, and developed narratives, as shown in Figures 5.11a and 5.11b. This falls in line with expectations for how an election campaign aims to support its candidates, discredit the opposition, increase party size, and develop the narratives around endorsed policies. In Figure 5.11c, left-wing users tended to conduct more maneuvers that affected leaders and developed the narrative than right-wing users. However, both sides made similar attempts to increase the size of their communities over time.

For further inspection, we conducted a chi-square test of independence to examine if there were a significant relationship between the different political sides and the use of the components of these three BEND categories over the entire data set. Table 5.6 shows the results of a statistical test for each maneuver. Our results showed most of the maneuvers had a high chi-square value and p-value less than the .05 threshold indicating that left-wing users used these maneuvers significantly more than right-wing users. The exception was the *enhance* maneuver. This maneuver refers to discussion or actions that provide supportive material that expands the topic for the targeted community or actor. Both parties used this maneuver with similar consistency.

Table 5.6: Chi-squared test for BEND maneuvers in top BEND categories (df=1, N=496048).

BEND Category	Maneuver	χ^2	p-value
Grow Groups	Build	7919.2	<.0001
	Boost	7946.9	<.0001
	Bridge	2030.6	<.0001
Affect Leaders	Back	12489	<.0001
	Neutralize	10922	<.0001
Develop Narratives	Engage	4073.8	<.0001
	Explain	1098.8	<.0001
	Enhance	1.1868	0.276

Given the nature of election campaigns toward supporting people and policies, we further explored two specific maneuvers: *back* and *neutralize*. The *back* maneuver is defined as discussion or actions that support another user or leader. The *neutralize* maneuver does the opposite in that it is used to discredit or decrease the importance of a leader. Figure 5.12 shows over time how many members of the left- and right-wing communities were maneuvering against other users and how many members were being maneuvered-upon or are targets of maneuvers. We found that left-wing users were the predominant executors of these maneuvers, whereas right-wing



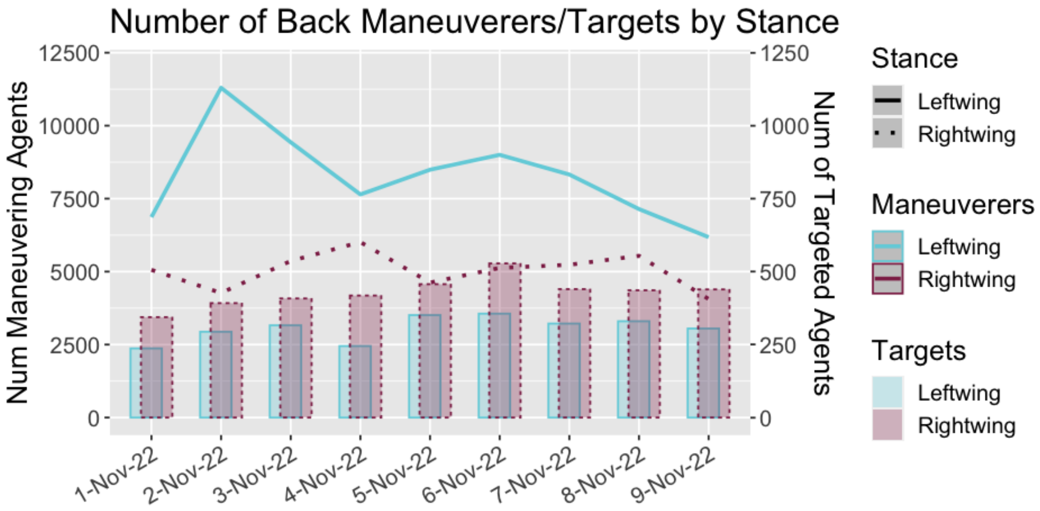
(a) Left-wing BEND maneuver categories usage.

(b) Right-wing BEND maneuver categories usage.

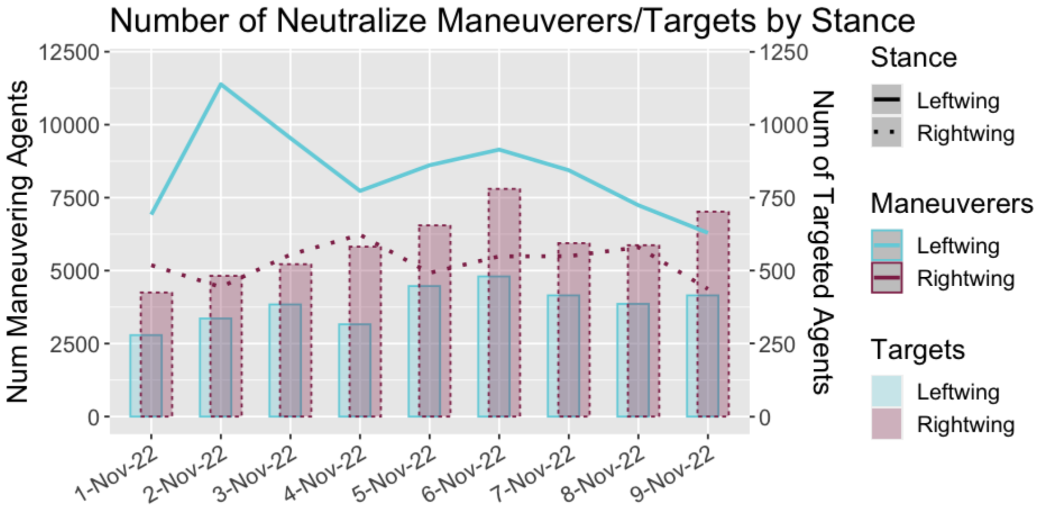
(c) Difference in BEND category message usage. Blue indicates more left-wing usage; red indicates more right-wing usage.

Figure 5.11: BEND maneuver categories usage by stance. Number is based on the summation of the number of individual maneuvers used in each category.

users were more often the targets of these maneuvers throughout the time period. This means that while more left-wing accounts created tweets that intended to *back* and *neutralize* overall, there were more *back* and *neutralize* tweets within the data set that were directed to or mentioned specific right-wing accounts.



(a) Number of Back Maneuverers and Targets by Stance.



(b) Number of Neutralize Maneuverers and Targets by Stance.

Figure 5.12: Number of left-wing and right-wing maneuverers and targets for BEND maneuvers affecting leaders.

Summary of 2022 US midterm election influence campaigns

In summary, left-wing and right-wing communities collectively engaged in online actions to influence the 2022 US midterm elections. Their ultimate offline goals were to win the elections for their candidates and the policies they supported. This translated to online behavior that focused on positioning key actors, developing narratives, and manipulating the social network.

Each side engaged in superspreader behavior through the communications of their political candidates and other verified users. The ties between these users not only focused on their political base but in many cases extended beyond party lines in attempts to persuade undecided voters or discredit the opposition leaders. Bots and news agencies were only a small part of the superspreader demographic.

Hashtags were an effective tool in tying the narratives related to both their candidates and policies throughout the social network. These hashtags included themes of general support for the party, negative messaging against the opposite party and their candidates, and indications of policy concerns such as with women's rights and gas prices.

The work on this case study also showed that though there were a larger number of unique right-wing users throughout the week, left-wing users interacted more consistently and widely every day leading up to the election. Additionally, left-leaning accounts showed more indication of extending the reach of their messages outside of their communities while the right-leaning accounts appeared to be more interconnected with their own communities showing generally greater signs of echo-chamberness.

Both parties used combinations of maneuvers that developed their party's narratives, grew their community, supported their candidates, and discredited opposing candidates with left-wing communities conducting maneuvers significantly more than right-wing. Further inspection on backing and neutralizing maneuvers showed that left-wing users acted more as maneuverers or targeters while right-wing users tended to be more maneuvered-upon or targeted.

Limitations

There were several notable limitations in this study. First, though the stance detection algorithm for separating the left-wing and right-wing communities had mean confidence levels of 84% and 96%, respectively, there were several instances of obvious political actors identifying with the incorrect community. This is possibly a result of the use of neutral hashtags used more by one political party than the other or forms of hashtag latching where unrelated hashtags are tied to the political messages. Secondly, we attempted to mitigate the errors from bot detection by using the union of two bot algorithms. Though this reduces the likelihood of false negatives, there is likely an increase in false positives. Third, BEND maneuver detection continues to be a progression of refining metrics to better detect maneuvers. In this case, some of the maneuvers are more detectable than others. However, the results from the case study are used to improve the methods of identifying the BEND maneuvers. Finally, studies have shown that Twitter users are more likely to be Democratic [142] than the general public. However, in this study, we focus on the content of the messages and the volume of users, messages, and maneuvers that are occurring within the Twitter environment. Though this may not be an accurate reflection of the amount of influence occurring in the real world, it does show how influence is spreading within social media.

5.3.4 Conclusion

In this chapter, we made two major contributions. We illustrated the BEND framework as an approach for gaining insight into the inter-workings of influence campaigns, and we provided a case study example using the 2022 US midterm elections. The framework looked at multiple facets of the left-leaning and right-leaning communities and provided a qualitative and quantitative understanding of the actors, the narratives, and the social network over time just prior to election day.

We found the two political-leaning communities engaged their candidates as highly-connected superspreaders to communicate their party's policies and to convince their followers to vote them into office. Leading up to election day, both communities shared messages consistent with the themes of their party with left-wing users tweeting daily at a higher frequency than right-wing users despite the higher number of right-wing users across all nine days. Right-wing communities also tended to have a more close-knit group with denser internal interactions and more echo-chamberness overtime. Both parties, though more executed by left-wing users, used maneuvers that aimed at growing the size of their community, supporting their party leaders, discrediting opposition leaders, and developing party-related narratives. In terms of backing and neutralizing leaders, left-wing users conducted more of the maneuvering whereas right-wing users were more the targets of these maneuvers.

While we focused on select aspects of the BEND framework to develop a more nuanced understanding of the community influence during the election, the application of the BEND Framework can be further expanded to conduct a more comprehensive analysis of the election. This would consist of a deeper investigation of specific key actors, the development of the narratives over time using topic analysis techniques, and coordinated activities within each of the communities. Additionally, this study does not evaluate the impact of the maneuvers on behavioral response, such as voting results, or change in stance or political lean. We would like to extend this analysis to use the impact aspect of the BEND framework to consider the effects of the BEND maneuvers and create a more complete assessment. Furthermore, beyond the current data set, research should be conducted on whether or not the actions executed by the different communities during the 2022 midterm elections occurred in other elections in the US or abroad.

This research continues to build upon research on how to better identify, characterize, and assess influence campaigns. Future work in this field should continue to improve the ability to qualitatively and quantitatively detect the maneuvers and the algorithms that feed into the BEND framework pipeline. As these types of influence operations have shown that they can have large impacts on society, understanding them is crucial so that leaders can make well-informed decisions when confronted with them.

5.4 Ukraine-Russia Case Study

5.4.1 Purpose

Evaluate the extent that the presidential administrations of Russia and Ukraine attempt to extend their influence using social media. This can be viewed through the methods that they use as well as the impacts they appear to have on the online social network.

Guiding questions

- What are the objectives for each administration?
- How do the presidents differ in how they use their social media account?
- What narratives/themes surround each of these leaders?
- What makes their narratives more effective in meeting their objectives?
- What are the dynamics for them as influencers and as targets of conversation?
- What key accounts discussed the Presidents, and how were their interactions?
- What are the impacts of their social media presence?

5.4.2 Background

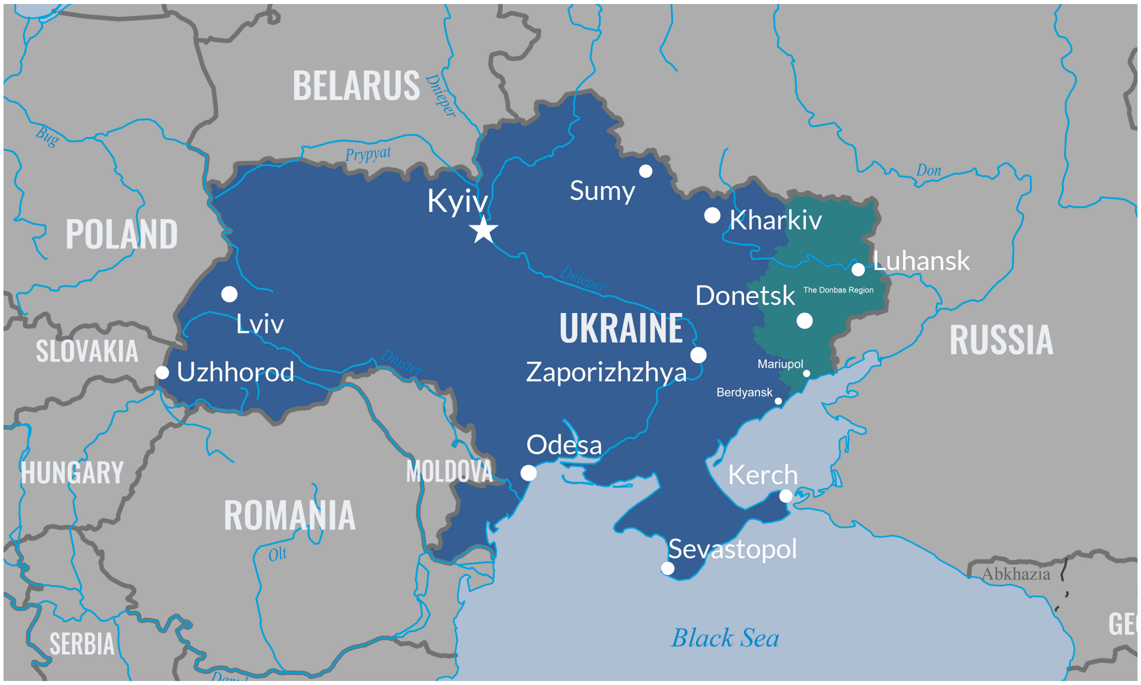


Figure 5.13: Ukraine Map

Ukraine is the second largest country located in Eastern Europe, bordering Russia to the east and the northeast 5.13. The Ukrainian-speaking western part of the country is more pro-Western, whereas eastern Ukraine, particularly the Donbas region of eastern Ukraine, is more pro-Russia.

In 2014, Russia quickly invaded and annexed the Crimean peninsula in southern Ukraine. Putin cited the large population of Russian citizens and Russian speakers in the area and called the annexation a "reunification with Russia." [112] This was the first time a European state annexed the territory of another state since World War II. Leading up to and during the event, Russia engaged in influence operations through the spread of false news and rumors on social media to set the stage for their invasion and support their cause [8, 134]. Following the invasion, the Crimeans held a referendum where they overwhelmingly voted in favor of reuniting with Russia. However, due to no international validation, the vote is considered illegitimate by the United Nations and the rest of the world.

Though in 1991, Ukraine declared its independence from the Soviet Union making it an independent country, on many occasions since then, Russia has refused to acknowledge them as a sovereign state. In 2021, Russian President Vladimir Putin wrote an article about how Russia and Ukraine were a single people [113], foreshadowing the current conflict. Additionally, Putin stated his opposition to Ukraine becoming more allied with the European Union and other Western countries, particularly through NATO. He warned that these would be indications of hostilities against Russia. Then just prior to the invasion in February 2022, Putin continued to make his stance on Ukraine by publicly refusing to acknowledge their sovereign statehood and claiming that Ukraine was being manipulated by foreign powers.

Situation

On February 24, 2022, Russia invaded the Eastern European country of Ukraine, which Putin called a 'Bspecial military operation' in what he deemed an attempt to protect the persecuted ethnic Russian minority of Donbas. Unlike during the annexation of Crimea, the United States declassified information regarding Russia making allied nations more aligned in preparation for the invasion [12]. Since then, Ukraine's President Volodymyr Zelensky has sought and achieved widespread Western support for Ukraine in its defense against Russia.

5.4.3 Sources

- Twitter Data
  - Collected through Twitter API
  - Dates: 22 Feb - 24 Mar 2022 (the first month of invasion)
  - Initial Zelensky data set
    - Collection terms: zelensky AND ukraine
    - Unique Agents: 1,118,114
    - Tweets: 4,114,395
    - Hashtags: 29,793
  - Initial Putin data set

    - Collection terms: putin AND ukraine AND russia
    - Unique Agents: 1,419,234
    - Tweets: 5,123,407
    - Hashtags: 49,673
  - Spheres of influence data sets (key actor ego networks): Both original data sets were filtered to consist of the spheres of influence of key actors of interest. These are the actors, tweets, hashtags, URLs, and locations up to one level adjacent to the key actor through user interactions (retweets, quotes, replies, mentions).
    - Zelenksy data set: ZelenskyyUa
    - Putin data set: KremlinRussia_E
- Tools
  - Approach for analysis: BEND framework (Chapter 2)
  - ORA-PRO Software v154
  - Netmapper Software v94
  - BotHunter [16]

5.4.4 Analysis

Actors of interest

- ZelenskyyUa - Official account of Ukrainian President Zelensky
- MFA_Ukraine - Government organization account of the Ministry of Foreign Affairs (MFA) of Ukraine
- KremlinRussia_E - Government organization account of the President of Russia
- mfa_russia - Government organization account of the Ministry of Foreign Affairs (MFA) of Russia

The four primary actors of interest are the Twitter accounts for the presidents of Russia and Ukraine and their country's ministries of foreign affairs. These accounts are official government accounts verified by Twitter as authentic. Based on the positions of the account holders and the posting frequency and content, these accounts were selected to represent the public-facing accounts of the chosen administrations. During this analysis, all four actors are analyzed when addressing issues that require a general analysis of the country or the president's administration. However, in some parts of the analysis, there is a more focused examination of the individual presidents' accounts.

Spheres of Influence Network Measures (*KremlinRussia_E, ZelenskyyUa*)

Of the four primary actors of interest, two spheres of influence (ego networks) were created to focus on the accounts of Putin (KremlinRussia_E) and Zelensky (ZelenskyyUa).

Table 5.7: Key actor network characteristics

Account	Actors	Hashtags	Tweets	URLs	Density (agent interaction)
KremlinRussia_E	8,978	7	14,488	79	2.40E-04
ZelenskyyUa	78,237	18	100,182	21	8.28E-05

Table 5.7 shows some of the network-level characteristics of the two spheres. Zelensky's network is greater than Putin's in the number of actors interacting with the president, hashtags, and tweets in the network. Putin, however, had a greater number of references to URLs. Furthermore, calculations were done on the density of the all-communication, or interaction, the network for each sphere. The all-communication network is a network where agents are linked if they quote, retweet, reply, or mention another user. Of all the possible connections that can occur in the network, the density is the fraction of the network connected. A higher density indicates a higher percentage of links across the network and measures how potentially cohesive a network is. In this case, Putin's network has a greater density than Zelensky's.

Actors of Interest BEND Analysis

A BEND analysis was conducted to explore the online actions of the key actors (Figure 5.14). The key actors are generally attempting to *build* their base of support while minimizing the opposing country's using the neglect maneuver. The MFA of Russia appears to be the more active maneuvering account of the Russian key actors, and Zelensky appears to be the more active for the Ukrainians. Many of the maneuvering messages of these actors use the enhance maneuvers with the MFA of Russia also using explain maneuvers. This means that the maneuvering actors are attempting to develop the narratives that support their sides. They attempt to extend that reach by building their communities by targeting specific online actors as described in the next section.

Targeted Accounts

Each account attempted to target various actors using the mention feature on Twitter. This feature connects a user account to a message when the username is placed in a tweet preceded by an @ symbol. This draws the recipient to the tweet. The tweet appears as a notification to the recipient and will appear in the recipient's timeline view [130]. Table 5.8 lists the four key actors and up to the top ten accounts they mentioned in the one-month period of the invasion. Some of these actors did not mention ten accounts.

The Russian accounts targeted world leaders and organizations. President Putin's account only mentioned one account, the Prime Minister of Israel Naftali Bennett. This was regarding a phone call to Putin initiated by the PM to mediate between Russia and Ukraine. Zelensky initially requested the PM to act as a mediator days earlier.

Russia's MFA, on the other hand, mentioned many more accounts. These included internal accounts such as those for Russia's Ministry of Defense (MoD) and a self-amplification of the MFA account as well as international organizations such as the North Atlantic Treaty Orga-

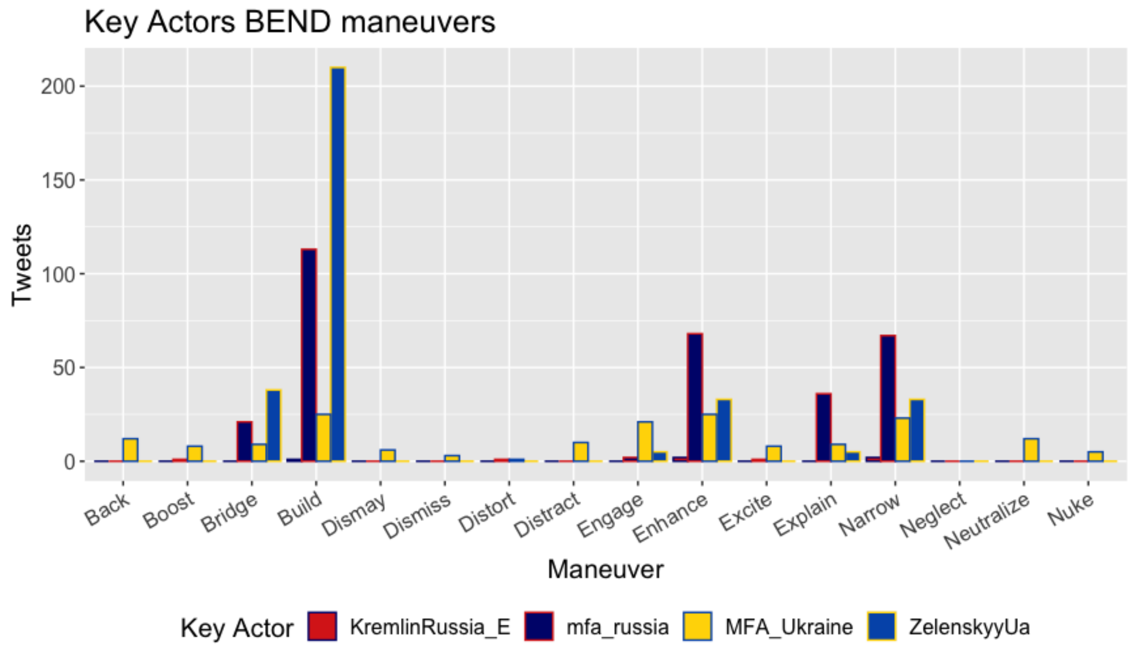


Figure 5.14: Key Actor BEND maneuvers

nization (NATO) and the United Nations (UN). They highlighted the MoD work on providing humanitarian aid to cities in the frontlines and spoke against Neo-Nazi regimes and genocide in Ukraine. The MFA discussed the multiple expansions of NATO after promises NATO had promised not to expand, marking threats to Russian security. One message right before the invasion was the MFA speaking on an unsupported letter from the US to the UN suggesting targeted killings in Russia. They call on the UN to stop taking sides and spreading anti-Russia rhetoric stating ”an information war on Russia.”

The MFA also targeted specific official accounts for several countries. Among their top ten were the Presidents of France and Ukraine, the Prime Minister of Israel, The Ministries of Foreign Affairs for Turkey and the Kyrgyz Republic, and the US Department of Defense. The references discussed conversations Putin and the Russian Foreign Minister had with world leaders about the situation in Ukraine. Messages mentioning the US Department of Defense (DoD) tied a Ukraine military biological program to US funding.

The President of Ukraine primarily focused on world leaders of other nations and the presidents of the EU Commission and the European Council. He brings the issues of Russian aggression and Russian crimes against Ukraine to the focus of conversation asking for and gaining support from foreign partners to help counter the Russian Federation. He aims to build an anti-war coalition of foreign partners to defend against Russia.

The MFA of Ukraine account did not share the same focused targeting as their president. They amplified messages by Zelensky and their ministers of defense and foreign affairs. They discuss EU membership for Ukraine with the Financial Times, a news agency, and highlight a discussion by the Global Citizen, an anti-poverty organization regarding the growing hunger issues in Ukraine since the war. The further highlight images from photographers, such as the

Table 5.8: Top Targeted Accounts by Key Actors

KremlinRussia_E	
naftalibennett	Prime Minister of Israel
mfa_russia	
mod_russia	Ministry of Defense Russia
NATO	North Atlantic Treaty Organization
EmmanuelMacron	President of France
UN	United Nations
MevlutCavusoglu	Minister of Foreign Affairs of Turkey
ZelenskyyUa	President of Ukraine
mfa_russia	Minister of Foreign Affairs of Russia
naftalibennett	Prime Minister of Israel
DeptofDefense	US Department of Defense
Kazakbaev_R	Minister of Foreign Affairs of the Kyrgyz Republic
ZelenskyyUa	
BorisJohnson	Prime Minister of the United Kingdom
vonderleyen	President of the EU Commission
EmmanuelMacron	President of France
eucopresident	President of the European Council
POTUS	President of the United States
AndrzejDuda	President of the Republic of Poland
JustinTrudeau	Prime Minister of Canada
OlafScholz	Chancellor of Germany
naftalibennett	Prime Minister of Israel
MinPres	Prime Minister of the Netherlands
MFA_Ukraine	
ZelenskyyUa	President of Ukraine
DmytroKuleba	Minister of Foreign Affairs of Ukraine
FT	Financial Times (news agency)
GlblCtzn	Global Citizen (anti-poverty organization)
Hetzner_Online	Data Center Company, Hosts websites
MagnumPhotos	Photographers
oleksiireznikov	Minister of Defense of Ukraine
serhiy_zhadan	Ukrainian writer

account MagnumPhotos, to draw attention to the ongoings in Ukraine.

In summary, targeting specific accounts through mentioning was used by both countries to draw national issues regarding the conflict to the world stage. For Russia, they spoke out against the expansion of NATO and the UN bias in favor of Ukraine. They also highlighted meetings between their senior leaders with world leaders to support their ‘special military operations,’ citing Ukrainian genocide and other crimes to validate the invasion. Zelensky’s account primarily targeted world leaders, calling for unity and assistance from Western countries against Russia. Their MFA account targeted specific accounts to draw attention to specific problems with Ukraine.

Targeting the Presidential Accounts

There was a noticeable difference in how the different presidents were targeted on social media. The messages that mentioned Putin’s account were generally negative. Many accused him of war crimes and exhibited general hate towards Russian military actions in Ukraine. These messages often contained positive support for Ukraine to contrast with the anti-Russia sentiment.

On the other hand, messages that mentioned or targeted Zelensky’s account were generally positive. Users discussed support for Ukraine and praised the Ukrainian president for his courage to stand up against Putin. Many also offered thoughts and prayers while condemning Russian actions. There were, however, a handful of anti-Ukraine messages directed at Zelensky. These discussed Russia’s topics related Neo-Nazis against Russians in Ukraine and how the president is a puppet of the West.

Bots and Superspreaders

BotHunter [16] was used to detect automated accounts, or bots, on the ego networks, or spheres of influence, of the Zelensky (ZelenskyyUa) and Putin (KremlinRussia_E) official accounts. For both accounts, there were a large number of bots surrounding both actors. These numbers were still high even after adjustments of increasing the threshold value for selection [97]. Of the actors interacting with the Kremlin account, 4560 of 8978 (50.8%) accounts were found to be bots, and 65,826 out of 78,237 (84.1%) were bots for the Zelensky account.

Analyzing the top superspreaders of each of the networks, both presidents had different types of these influential actors. Zelensky’s top 10 superspreaders consisted of verified users. Half were journalists or news organizations that supported Ukraine, and the other half were Western world leaders. On the other hand, the top 10 superspreaders within Putin’s ego network consisted only of two verified accounts, the PM of Israel and an online US Democratic-leaning influencer, and six of them are bots. Within that list, only one account supported the Russian cause, and this account was a bot.

A BEND maneuver report was conducted for the bots interacting with the presidents’ accounts. Figure 5.15 shows the BEND maneuvers comparing the bots surrounding the Russian president’s account to non-bots. While most maneuvers were generally the same, the top 3 major differences were in the back, neutralize, and engage maneuvers. The bots in the Zelensky sphere of influence had a greater difference in the number of tweets containing BEND maneuvers as shown in Figure 5.16. Similar to Putin, the top maneuvers were also back, neutralize, and engage, but the number of tweets containing maneuvers was greater across all maneuvers.

Further inspection of the bots surrounding both presidents reveals that many of the bot messages were retweets and pro-Ukraine. 95% of the bot tweets in the Zelensky sphere and 80% of the bot tweets in the Putin sphere were retweets. Many are retweets of news organizations reporting on the war in favor of Ukraine, though there were very small numbers of pro-Russian tweets.

Overall, bots made up a large percentage of both the Putin and Zelensky accounts explaining the large numbers of tweets authored by bots compared to non-bots. The bots showed the power of how a large community of bots can make a potentially large impact as far as spreading information throughout a network with large quantities of tweets. This impact is predicated on many of these tweets containing the BEND maneuvers.

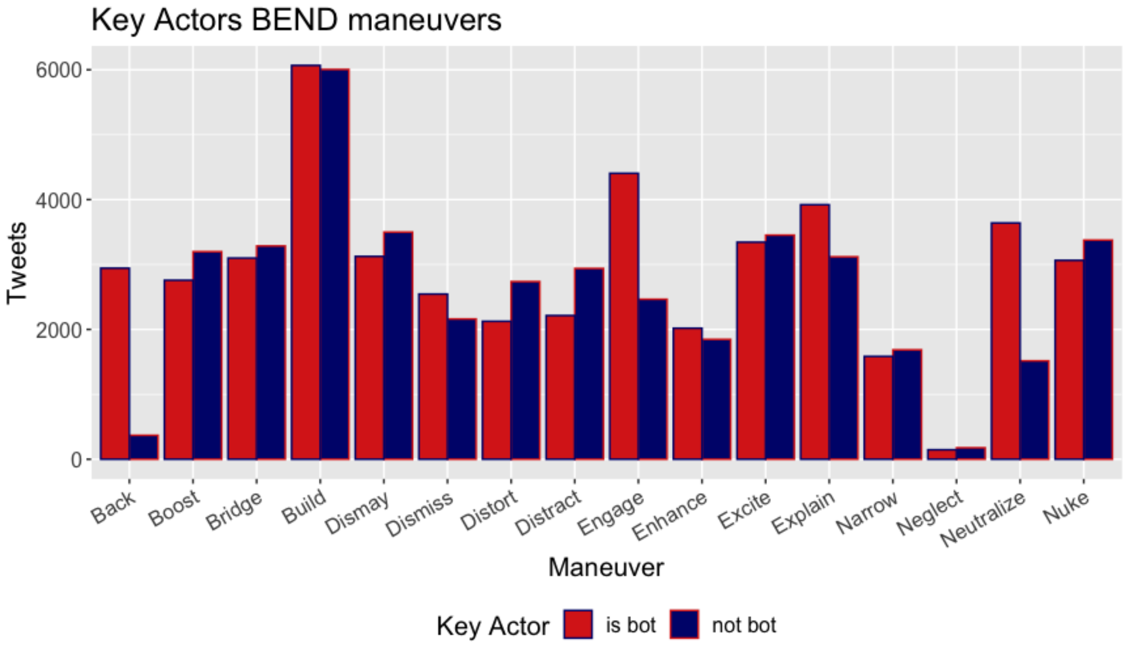


Figure 5.15: Bot BEND maneuvers in *KremlinRussia_E* ego network

Hashtag analysis

Russian and Ukrainian accounts used hashtags differently throughout the data sets. Russian accounts used them except to highlight locations with the country or actors they are engaging with. No obvious pro-Russian or anti-Ukraine hashtags emerged from any of the data sets. Ukrainian accounts, however, used hashtags to spread their anti-war and anti-Russian messages. Among the top hashtags used were #StandWithUkraine, #UkraineUnderAttack, #StopRussia, #PutinWarCriminal, and #StopRussianAggression.

5.4.5 Overall assessment

The Putin and Zelensky administrations differed in their objectives for using social media regarding the invasion of Ukraine. They utilized social media accounts in different ways to convey their overall narratives. It is evident after this analysis, that the narratives and actors on Twitter

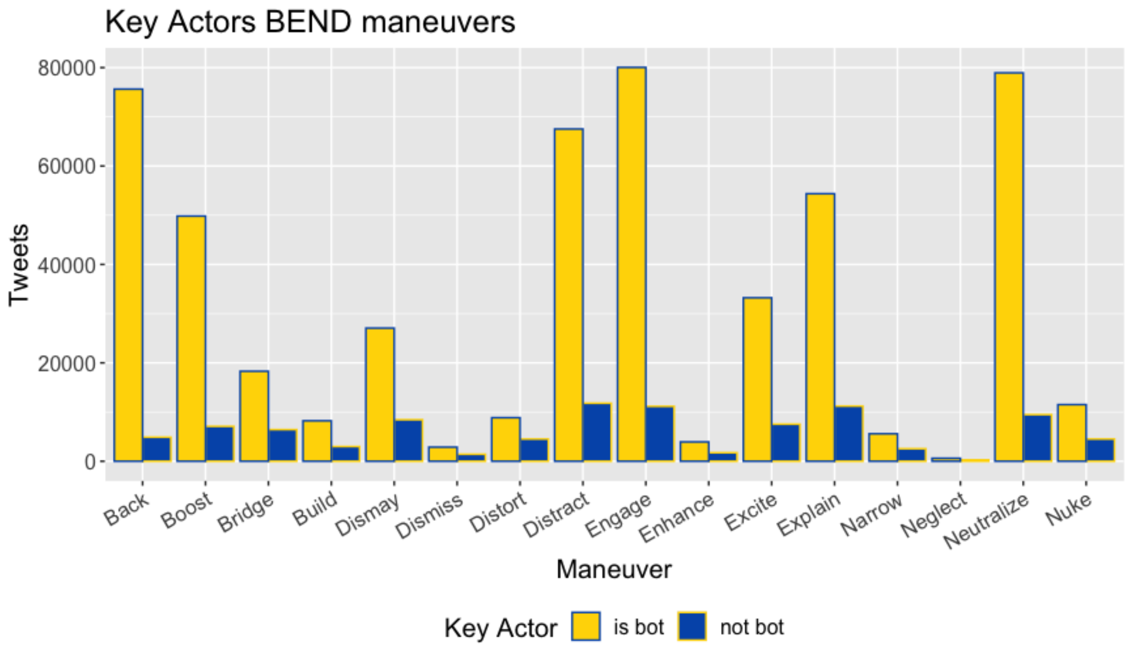


Figure 5.16: Bot BEND maneuvers in *ZelenskyyUa* ego network

surrounding the presidents of Russia and Ukraine as related to the Russian invasion were largely pro-Ukraine and pro-Zelensky.

Russia focused on defending and legitimizing its actions to conduct ‘special military operations’ in Russia based on the Neo-Nazi or genocidal crimes conducted by Ukraine. They cited many occasions of reaching out to the international community to explain the need for the operations and condemn the on-goings against Russians in Ukraine. They targeted world leaders to side with them and called on world organizations to stop anti-Russian misinformation. Many of the accounts that interacted with the Russian accounts, particularly Putin’s, sent messages against the Russian invasion of Ukraine. Though there were multiple instances of Putin discussing in person how Russia and Ukraine were one people prior to the invasion, this narrative did not emerge as any of the themes from the Russian accounts.

Ukraine sought to gain aid and support to defend its country from the Russians. They reached out to world leaders to ask for aid and support and condemned Russian actions. Ukraine accounts messaged using both pro-Ukraine and anti-Russian hashtags that were highly used showing the successful impact of narrative maneuvers. The community of agents that interacted with the Ukraine key actor accounts primarily shared the pro-Ukraine sentiment.

As Twitter is a Western-based social media platform, the users predominately support the pro-western rhetoric conveyed by the Ukrainian government. This ultimately made it an ideal platform for Ukraine to foster pro-western support. While the Russian regime attempted to use maneuvers and sway the Western users, Ukraine was able to rally more online support from an already biased platform.

Furthermore, both Russia-Ukraine data sets contained a high percentage of bots. Within both Putin and Zelensky’s spheres of influence, many of these bots were retweeting many pro-Ukraine messages containing hashtags in-line with the Pro-Ukraine narrative.

Chapter 6

Concluding Remarks

6.1 Summary

As social media continues to expand and evolve, influencing online actors continue to take advantage of the convenience of the medium to allow them to engage with their target audiences. While today, the trending campaign may be about the next presidential election or the newest regulation regarding COVID, tomorrow will bring a new controversial topic or opportunity for people to influence, persuade, or manipulate, and with social media, they will do it fast and reach across the globe. The BEND framework presented in this thesis provides a method to look at those information operations and comprehensively characterize them to support leaders of organizations who need to make decisions on social media.

In summary, I have expanded an influence operations analysis framework from its general application to a comprehensive analysis and assessment tool suited for both military and general social media analysts. I began with expanding the definitions of the BEND framework to create clarity when differentiating the fundamental influence techniques, or social-cyber maneuvers, used by online actors. I described them as individual actions and in combination in the form of BEND tactical categories. I then proceed to conduct a statistics-based analysis for developing refined detection techniques by first examining data with human-annotated BEND classifications, comparing that data to the current method for automated maneuver detection, and then developing and initiating the iterative process of analyzing the results of the automated maneuver detection to systematically improve detection. Afterward, I describe the methodology for using the BEND framework to characterize influence operations and create assessments, and then I illustrate that application using three case studies. These case studies describe the influence operations during the COVID-19 vaccine rollout, the 2022 US midterm elections, and the Ukraine-Russia War following the invasion in 2022.

6.2 Contributions

In this thesis, I have made several contributions to the field of social cybersecurity and related online social network research. The contributions include the following:

Data Sets

First, I created a corpus of at least 10000 tweets labeled as one or more of the BEND maneuvers described in the framework. This required developing a user-friendly interface and training annotators on correctly labeling the data. The data are samples of all the data sets used in this thesis.

Second, for all of the data sets, I synthesized and expanded the Twitter data sets surrounding various events to include meta-data suitable for social network analysis. Depending on the data set, this may include information on linguistic CUES, bot attributes, stance classification, topic-oriented group information, and BEND maneuver information.

Characterized Social-Cyber Maneuvers

I expanded on the current framework by developing comprehensive definitions and identifying indicators for each maneuver. This includes clear and specific definitions and examples, methods for detection, and statistical data on the usage of each of the maneuvers. I also described several examples of the maneuvers used in combination to create BEND tactical categories. These categories explain general actions or effects that can be achieved when using the maneuvers together. The information in this section is catered to support training on the BEND maneuvers.

Process for Improving Detection

In developing this thesis, I developed two levels of improving detection that improved BEND detection overall. The first was the fine-grained iterative process of using statistical analysis on CUES indicators to improve weights and thresholds, as discussed in detail in Chapter 3.

The second was a more complex methodology of improving detection encompassing all of the chapters within the thesis. The overall process is depicted in Figure 6.1. This began with a literature review of the CUES and the BEND maneuvers from Chapter 2 and 3 to improve CUES. The three case studies discussed in Chapter 5 not only informed the creation of the framework and illustrated the methodology but the lessons learned from these case studies were used to identify gaps and helped inform necessary changes to CUES and other indicators to improve the social-cyber maneuvers detection. The integration of human-annotated data and comparison with automated results were used to both assess the need for the algorithm and inform areas of improvement within the algorithm through instances of discrepancies between users and the machine. Finally, the fine-grained iterative detection refinement methodology discussed in Chapter 3 used statistical analysis to conduct fine-point analyses of CUES and thresholds. The next step would be to repeat the process with indicators for improving the use of media forms such as images, videos, and others. This method applied diverse techniques that contributed to detection through a ratcheting process of incrementally making improvements. Incorporating these techniques can be used in future situations for algorithms where training sets are not feasible to construct.

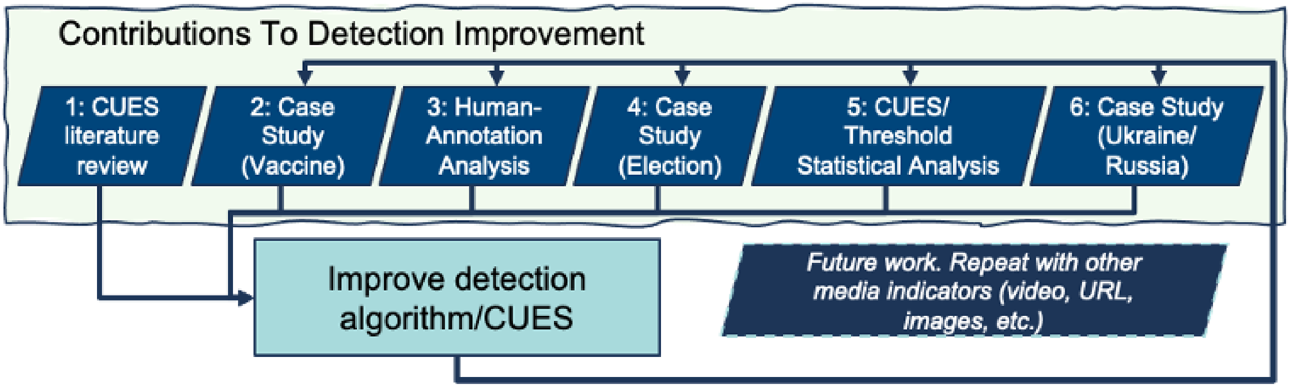

Contributions To Detection Improvement
1: CUES literature review
2: Case Study (Vaccine)
3: Human-Annotation Analysis
4: Case Study (Election)
5: CUES/ Threshold Statistical Analysis
6: Case Study (Ukraine/ Russia)
Improve detection algorithm/CUES
Future work. Repeat with other media indicators (video, URL, images, etc.)

Figure 6.1: Diagram of overarching path to improve BEND detection. Each contributing step improves the detection algorithm or CUES in a ratcheting process.

Applying and Assessing Social-Cyber Maneuvers

I developed a method for operationalizing the BEND framework to better analyze influence operations from the context of the social-cyber maneuvers. Then I create a practical method for making assessments that would be beneficial for informing decision-makers.

Case Studies

I created an analysis and assessment on three real-world scenarios, specifically discussions on the COVID-19 Vaccine, the US 2022 Midterm Election, and key actors in the Ukraine-Russia War in 2022. Not only does this add valuable insight for these specific events, but I also illustrate an effective application of the social-cyber maneuver framework.

Counter-Maneuver Simulations

I created simulations to analyze the effects of using the BEND maneuvers in a simulated Twitter environment [20]. These inform the effectiveness of the maneuvers and provide a better understanding of methods for countering the maneuvers online. This work can be found in Appendix E.

6.3 Limitations

There are several limitations within the scope of this thesis. First, though many aspects of this framework theoretically can be applied to multiple types of social media platforms, this research's foundational work and case studies were developed based on the Twitter structure and using Twitter data. Furthermore, some very specific characterizations and detection methods may only apply to Twitter data or platforms with similar features (such as those using messages, hashtags, and mentions).

Another limitation involves the inherent use of Twitter API and data collection. The Twitter API only allows us to collect a sampling of the network based on specific accounts and collection terms. Users are only allowed a specific number of tweets a month, resulting in only a small fraction of the entire discussion. This means that often we have incomplete communications and follower networks. To collect the data sufficiently enough to create a more accurate BEND assessment, there need to be fewer restrictions on the data collected across all platforms, not just Twitter. The meta-data or message data should also be detailed and accessible enough to derive the required indicators for the analysis. This means access to suspended accounts and tweets to observe conversations online; the ability to download/view media files such as images, audio, and URLs; and policies to decrease the cost barrier for conducting research in this area.

Finally, the analysis of the BEND maneuvers is based on analyzing individual data sets where any aggregations would be of the particular period under analysis. Overtime analysis results from compiling the analysis of multiple data sets rather than a single overtime calculation of the entire data set. This means tweets or conversations are not systematically tracked.

6.4 Future Research

There are several avenues of research in which I suggest future work in this area can benefit the field of social cybersecurity.

BEND Tactical Categories

This research touched upon some of the effects and actions that result from combinations of maneuvers. These maneuvers have been observed in other combinations than those discussed, and the combinations should be analyzed dynamically over time.

Disinformation and Influence

Though disinformation can be an instrument for influence, this is not always the case. As shown in the case studies in Chapter 5, influence techniques need not always accompany malicious intent. One possible future direction can be to look at the relationship between disinformation and influence to determine whether different maneuvers are used when the information spread is disinformation.

Detection

While I have previously stated in Chapter 3 that detection of the individual maneuvers needs continued refinement, other aspects of detection should be addressed.

Detection should incorporate non-text information such as images, videos, audio, URL analysis, etc. The current process primarily uses text analysis as the primary source for detection. The BEND analysis, therefore, lacks the added value provided by other aspects or modalities of a message. Social media platforms, such as TikTok and Instagram, primarily use images and videos for posts, and Facebook posts typically accompany images or URLs. For these types of platforms, the text becomes supplemental to these more prominent platform features and therefore less of the primary indicator of the maneuvers. A recommended next step for detection could be to derive emotional content from images, audio, and video to identify the emotion maneuvers of dismay and excite [137, 146, 147]. Work has also been done on extracting semantic cues and indexing images, which may lead to better detection of other maneuvers [90, 143]. Each modality would have to be separately analyzed and then combined to create a holistic view of the overarching maneuver.

Methods should also be improved to be able to detect the maneuvers within the context of their position within the narrative or a storyline. Currently, maneuvers are identified in isolation, which sometimes results in erroneous classifications of posts. Finally, detection should be extended to automating impact analysis of the community maneuvers. Actors may have been attempting boost maneuvers in time period 1. How successful were they in time period 10? How has the network evolved to reflect the effects of this maneuver?

BEND Cross-Platform

While this work primarily uses Twitter as the base for research, the BEND framework can be applied on other social media platforms. More research should be extended to the applications on other platforms. The user interaction structure and predominate form for messages (e.g., images, texts, audio, etc.) can affect some methods of the framework. For example, Twitter, networks based on interactions through retweets, quotes, replies, and mentions. Information is not limited to individual communities but can be extended easily throughout the entire network. Deriving Twitter communities is a manual process. Similarly, Facebook, Instagram, and TikTok interact by spreading information to followers who have the options to like, share, and reply. On the other hand, Reddit, Telegram, and Facebook's groups feature have communities or groups inherent to the social media platform. This type of structure facilitates the observations of community maneuvers related to the change in community growth or reduction. Tables 6.1 and 6.2 show a breakdown of popular social media platforms and their general network structures and modalities for spreading information. Depending on the network structure and modalities, BEND analysis may have a greater emphasis on community network maneuvers, and narrative maneuvers may be more effective if they are better capable of analyzing modalities other than text.

Table 6.1: Popular social media platform network comparison. Interaction networks refer to networks created by replies, quotes, mentions, and retweets. Community interaction networks refer to networks of users based on specific topics interacting through posts and replies.

Platform	Predominant Network Features	Other Network Features
Twitter	Interaction networks	Derived community networks
Facebook	Interaction networks	Community interaction networks
Instagram	Interaction networks	
TikTok	Interaction networks	
Reddit	Community interaction networks	
Telegram	Community interaction networks	Interaction networks (forwards)

Table 6.2: Popular social media platform modality comparison

Platform	Primary Modalities	Other Modalities
Twitter	Text, images, URLs	Videos, likes
Facebook	Text, images, URLs	Videos, reactions
Instagram	Images, Videos	Text, likes
TikTok	Videos	Text, likes
Reddit	Text	images, videos, URLs, votes
Telegram	Text, images, URLs	Videos, reactions

Counter-maneuvers

This work focused on identifying the maneuvers and using the maneuvers to characterize influence operations. However, organizational leaders may desire a way to combat the maneuvers that are being targeted against them. Future work should begin by exploring the tactics, techniques, and procedures for the social-cyber maneuvers and the responses used in varying scenarios. These can be observations of counter-maneuvers followed by evaluations of their effectiveness. This type of impact analysis relies on overtime analysis of the scenario. Though there may be challenges in determining whether or not beliefs or behaviors are affected, evaluation can be made on changes in network structure, overtime stance analysis, and the evolution of the narratives. This can then be followed up with simulations.

BEND Simulations

While several simulations have been conducted to explore the effects and actions of the BEND maneuvers on various networks, more can be done in this area. Beneficial simulations include refining the use of emotions, evaluating counter-maneuvers, and exploring the natural propagation of information in the presence of specific maneuvers over time. Though some simulations have partially addressed some of these recommendations, they can be improved with more accurate representations of the social network and maneuvers.

Social Media Analyst Training

The OMEN exercise was used in part to help develop the BEND framework methodology as discussed in Chapter 4.1.2. Currently, four days of training is used to provide in-class lectures as well as hands-on training on the tools and the processes for analyzing and assessing influence campaigns and suggesting potential courses of action. This course is suited for entry-level analysts. These would be the personnel who would take guidance from a more experienced analyst. A more experienced analyst would need a minimum of six weeks of training. This should involve instruction on using all of the tools within the pipeline including data collection, Netmapper, bot detection, and ORA-PRO. Furthermore, the training should include more in-depth analysis techniques and the application on multiple and varied data sets. Emphasis should be made on visualizations and forming the analysis as actionable information. As a capstone project, the analysts could apply these techniques on a real-world scenario and present them to appropriate decision-makers and stakeholders. Additionally, though this is intended for industry and government analysts, this can be expanded into a semester course in an academic environment.

Expanding Operational Utility

As mentioned in Chapter 4, the BEND framework has been used to train analysts in characterizing online influence campaigns to inform decision-making. Based on the experience from the training, the OMEN team has shown the training to be a potentially useful and accessible tool.

Several actions could facilitate the widespread use of this framework and its accompanying tools. First, this should continue to be presented to higher-level decision-makers as a critical tool for solving real-world challenges. Then the tools need to function at the high-level expected for its intended use. Not only do they need to believe the theory, but they need to believe that the tools will help them in the application of the theory. This means detection must continue to be improved and more features are required to conduct a thorough analysis. The tools, however, need to be easy to use. For example, ORA-PRO is a powerful network analysis tool, but for the analysts using it for analyzing influence operations, they may only require the functions related to their analysis process. This may also mean streamlining the pipeline into a single package where the users are able to move from one aspect of the analysis process to the other without a large amount of effort. Furthermore, the framework and tools need to be improved for social media platforms other than Twitter. As Twitter is becoming less accessible and other platforms are becoming more popular, the BEND framework should be able to get a holistic view of the information environment. Finally, as this framework has military applications, the names for the maneuvers should not represent already existing tactical terms. Therefore, the names *nuke* and *neutralize* should be reconsidered. Though it may not fit the acronym, the names should not be confined to starting with the letter ‘N’.

Bibliography

[1] Jais Adam-Troian and Thomas Arciszewski. Absolutist words from search volume data predict state-level suicide rates in the united states. *Clinical Psychological Science*, 8(4): 788–793, 2020. B.1

[2] 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. B.1

[3] Alexandre Alaphilippe. Adding a 'D' to the ABC disinformation framework. `https://www.brookings.edu/techstream/adding-a-d-to-the-abc-disinformation-framework/`. Accessed: 2023-03-01. 1.1, 1.1.2, 4.1.1

[4] Jon-Patrick Allem and Emilio Ferrara. Could social bots pose a threat to public health? *American journal of public health*, 108(8):1005, 2018. 1.3.1, 5.2.2

[5] Daniel Allington, Bobby Duffy, Simon Wessely, Nayana Dhavan, and James Rubin. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine*, 51(10):1763–1769, 2021. doi: 10.1017/S003329172000224X. 5.3.1

[6] Michelle A Amazeen and Erik P Bucy. Conferring resistance to digital disinformation: The inoculating influence of procedural news knowledge. *Journal of Broadcasting & Electronic Media*, 63(3):415–432, 2019. 3.2.4

[7] Vanessa de Oliveira Andreotti and Cash AhENAKEW. Equivocal knowing and elusive realities: Imagining global citizenship otherwise. In *Postcolonial perspectives on global citizenship education*, pages 233–250. Routledge, 2012. B.1

[8] Sinan Aral. *The hype machine: how social media disrupts our elections, our economy, and our health–and how we must adapt*. Currency, 2021. 5.4.2

[9] Kevin Arceneaux and Rory Truex. Donald trump and the lie. *Perspectives on Politics*, page 1–17, 2022. doi: 10.1017/S1537592722000901. 5.3.1

[10] Matthew Babcock, Ramon Alfonso Villa Cox, and Sumeet Kumar. Diffusion of pro- and anti-false information tweets: the black panther movie case. 25(1):72–84, . ISSN 1381-298X, 1572-9346. doi: 10.1007/s10588-018-09286-x. URL `http://link.springer.com/10.1007/s10588-018-09286-x`. 1.4

[11] Matthew Babcock, Ramon Alfonso Villa Cox, and Sumeet Kumar. Diffusion of pro-

and anti-false information tweets: the black panther movie case. 25(1):72–84, . ISSN 1381-298X, 1572-9346. doi: 10.1007/s10588-018-09286-x. URL http://link.springer.com/10.1007/s10588-018-09286-x. 1.4

[12] Julian E Barnes and Adam Entous. How the U.S. adopted a new intelligence playbook to expose Russia's war plans. https://www.nytimes.com/2023/02/23/us/politics/intelligence-russia-us-ukraine-china.html. Accessed: 2023-04-09. 5.4.2

[13] Frank M Bass. A new product growth for model consumer durables. *Management science*, 15(5):215–227, 1969. E.2.1

[14] Krishna C Bathina, Marijn Ten Thij, Lorenzo Lorenzo-Luaces, Lauren A Rutter, and Johan Bollen. Individuals with depression express more distorted thinking on social media. *Nature Human Behaviour*, 5(4):458–466, 2021. B.1

[15] Yochai Benkler, Casey Tilton, Bruce Etling, Hal Roberts, Justin Clark, Robert Faris, Jonas Kaiser, and Carolyn Schmitt. Mail-in voter fraud: Anatomy of a disinformation campaign. *Berkman Center Research Publication*, (2020-6), 2020. 5.3.1

[16] 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. 1.3.1, 4.2.1, 4.3.2, 5.2.3, 5.3.2, 5.4.3, 5.4.4, E.1

[17] David M Beskow and Kathleen M Carley. Social cybersecurity: An emerging national security requirement. *Military Review*, 99(2):117–127, 2019. ISSN 0026-4148. 1.1, 1.1.2, 2.1, 3.1, 5.2.2, 5.3.1, 5.3.2

[18] David M Beskow and Kathleen M Carley. Agent based simulation of bot disinformation maneuvers in twitter. In *2019 Winter simulation conference (WSC)*, pages 750–761. IEEE, 2019. E.1, E.2.1, E.3.1, E.3.2, E.6

[19] 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. 5.2.3

[20] Janice T Blane, JD Moffitt, and Kathleen M Carley. Simulating social-cyber maneuvers to deter disinformation campaigns. In *Social, Cultural, and Behavioral Modeling: 14th International Conference, SBP-BRiMS 2021, Virtual Event, July 6–9, 2021, Proceedings 14*, pages 153–163. Springer, 2021. 6.2

[21] Janice T Blane, Daniele Bellutta, and Kathleen M Carley. Social-cyber maneuvers during the covid-19 vaccine initial rollout: Content analysis of tweets. *J Med Internet Res*, 24(3): e34040, Mar 2022. ISSN 1438-8871. URL https://doi.org/10.2196/34040. 1.4, 2.4.5, 5.3.2

[22] Sam Blazek. SCOTCH: A framework for rapidly assessing influence operations. https://www.atlanticcouncil.org/blogs/geotech-cues/scotch-a-framework-for-rapidly-assessing-influence-operations/. Accessed: 2021-05-21. 1.1, 1.1.2

[23] Herbert Bless, Gerd Bohner, Norbert Schwarz, and Fritz Strack. Mood and persuasion: A cognitive response analysis. 2001. 3.3

[24] Erika Bonnevie, Jaclyn Goldbarg, Allison K Gallegos-Jeffrey, Sarah D Rosenberg, Ellen Wartella, and Joe Smyser. Content themes and influential voices within vaccine opposition on twitter, 2019. *American journal of public health*, 110(S3):S326–S330, 2020. 5.2.2

[25] Erika Bonnevie, Allison Gallegos-Jeffrey, Jaclyn Goldbarg, Brian Byrd, and Joseph Smyser. Quantifying the rise of vaccine opposition on twitter during the covid-19 pandemic. *Journal of communication in healthcare*, 14(1):12–19, 2021. 5.2.1

[26] Jean-Christophe Boucher, Kirsten Cornelson, Jamie L Benham, Madison M Fullerton, Theresa Tang, Cora Constantinescu, Mehdi Mourali, Robert J Oxoby, Deborah A Marshall, Hadi Hemmati, et al. Analyzing social media to explore the attitudes and behaviors following the announcement of successful covid-19 vaccine trials: infodemiology study. *JMIR infodemiology*, 1(1):e28800, 2021. 5.2.2

[27] William J Brady, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318, 2017. B.1

[28] William J Brady, Julian A Wills, Dominic Burkart, John T Jost, and Jay J Van Bavel. An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of Experimental Psychology: General*, 148(10):1802, 2019. B.1

[29] William J Brady, Molly J Crockett, and Jay J Van Bavel. The mad model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, 15(4):978–1010, 2020. B.1

[30] William J Brady, Ana P Gantman, and Jay J Van Bavel. Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General*, 149(4):746, 2020. B.1

[31] Daniel J Brass, Kenneth D Butterfield, and Bruce C Skaggs. Relationships and unethical behavior: A social network perspective. *Academy of management review*, 23(1):14–31, 1998. B.1

[32] 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. 1.3.1, 5.2.2

[33] Talha Burki. The online anti-vaccine movement in the age of covid-19. *The Lancet. Digital health*, 2(10):e504–e505, 2020. ISSN 2589-7500. 5.3.1

[34] Kathleen M Carley. Group stability: A socio-cognitive approach. *Advances in group processes*, 7(1):44, 1990. E.2.1

[35] Kathleen M Carley. Social cybersecurity: an emerging science. *Computational and Mathematical Organization Theory*, pages 1–17, 2020. 1.1.2, 2.1, 3.1, 3.3, 5.2.1, 5.2.2, 5.2.3, 5.3.1, E.1, E.2.2, E.4.1, E.4.1

[36] Kathleen M Carley, Michael K Martin, and Brian R Hirshman. The etiology of social

change. *Topics in Cognitive Science*, 1(4):621–650, 2009. E.2.1

[37] Kathleen M Carley, Guido Cervone, Nitin Agarwal, and Huan Liu. Social cyber-security. In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11*, pages 389–394. Springer, 2018. E.1

[38] KM Carley. Bend: a framework for social cybersecurity. *Future Force*, 6(2):22–27, 2020. 1.1, 5.2.3

[39] 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*, 2018. 1.3.3, 2.1, 3.1, 3.3, 4.2.2, 4.2.2, 5.2.3, 5.2.3, 5.3.2, D.1

[40] Arik Cheshin, Anat Rafaeli, and Nathan Bos. Anger and happiness in virtual teams: Emotional influences of text and behavior on others' affect in the absence of non-verbal cues. *Organizational behavior and human decision processes*, 116(1):2–16, 2011. 3.3, B.1

[41] CIRCLE. Youth voter turnout and impact in the 2022 midterm elections. Technical report, Tufts University, 2022. URL `https://circle.tufts.edu/2022-election-center`. 5.3.1

[42] Isobelle Clarke and Jack Grieve. Dimensions of abusive language on twitter. In *Proceedings of the first workshop on abusive language online*, pages 1–10, 2017. B.1

[43] COGSEC-Collaborative. COGSEC-Collaborative/AMITT: AMITT (adversarial misinformation and influence tactics and techniques) framework for describing disinformation incidents. includes TTPS and countermeasures. `https://github.com/cogsec-collaborative/AMITT`. Accessed: 2023-03-01. 1.1.2

[44] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960. doi: 10.1177/001316446002000104. URL `https://doi.org/10.1177/001316446002000104`. 3.2.2

[45] Warren Cornwall. Officials gird for a war on vaccine misinformation, 2020. 5.2.1

[46] Susanne E Craig and EeShan Bhatt. A short glossary of inclusive language. *Oceanography*, 34(2):6–9, 2021. B.1

[47] Daryl J Daley and David G Kendall. Stochastic rumours. *IMA Journal of Applied Mathematics*, 1(1):42–55, 1965. E.2.1

[48] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. Automated hate speech detection and the problem of offensive language. In *Proceedings of the international AAAI conference on web and social media*, volume 11, pages 512–515, 2017. B.1

[49] Daantje Derks, Arjan ER Bos, and Jasper Von Grumbkow. Emoticons and online message interpretation. *Social Science Computer Review*, 26(3):379–388, 2008. 3.3, B.1

[50] DISARM. DISARM is an open framework for those cooperating in the fight against disinformation. `https://www.disarm.foundation/`. Accessed: 2022-12-01. 1.1, 1.1.2

[51] Eve Dubé, Caroline Laberge, Maryse Guay, Paul Bramadat, Réal Roy, and Julie A Bettinger. Vaccine hesitancy: an overview. *Human vaccines & immunotherapeutics*, 9(8): 1763–1773, 2013. 5.2.1

[52] Owen Dyer. Vaccine safety: Russian bots and trolls stoked online debate, research finds. *BMJ: British Medical Journal (Online)*, 362, 2018. 1.3.1, 5.2.2

[53] Veikko Eranti and Markku Lonkila. The social significance of the facebook like button. *First Monday*, 20(6), 2015. 3.3

[54] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. *Commun. ACM*, 59(7):96–104, Jun 2016. ISSN 0001-0782. doi: 10.1145/2818717. URL `https://doi.org/10.1145/2818717`. 1.3.1, 5.2.2, 5.3.2

[55] Emilio Ferrara, Herbert Chang, Emily Chen, Goran Muric, and Jaimin Patel. Characterizing social media manipulation in the 2020 u.s. presidential election. *First Monday*, 25 (11), Oct. 2020. doi: 10.5210/fm.v25i11.11431. URL `https://journals.uic.edu/ojs/index.php/fm/article/view/11431`. 1.3.1, 5.3.1

[56] Christina Fleuriet, Megan Cole, and Laura K Guerrero. Exploring facebook: Attachment style and nonverbal message characteristics as predictors of anticipated emotional reactions to facebook postings. *Journal of Nonverbal Behavior*, 38:429–450, 2014. 3.3, B.1

[57] Camille Francois. Actors, behaviors, content: A disinformation ABC. `https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/ABC_Framework_TWG_Francois_Sept_2019.pdf`. Accessed: 2023-03-01. 1.1, 1.1.2, 4.1.1

[58] Kirby Goidel, Keith Gaddie, and Spencer Goidel. Rigged-election rhetoric: Coverage and consequences. *PS: Political Science amp; Politics*, 52(2):229–238, 2019. doi: 10.1017/S1049096518001646. 5.3.1

[59] Daniel Goleman. *Working with emotional intelligence*. Bantam, 1998. 2.3.3, 2.3.7, B.1

[60] Robert Gorwa and Douglas Guilbeault. Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet*, 12(2):225–248, 2020. 5.2.4

[61] Mark A Gunsch, Sheila Brownlow, Sarah E Haynes, and Zachary Mabe. Differential forms linguistic content of various of political advertising. *Journal of Broadcasting & Electronic Media*, 44(1):27–42, 2000. 3.6

[62] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008. E.3

[63] Mark A Hamilton. Message variables that mediate and moderate the effect of equivocal language on source credibility. *Journal of Language and Social Psychology*, 17(1):109–143, 1998. B.1

[64] Robert E Hamilton. Russia's attempts to undermine democracy in the west: Effects and causes. *Orbis*, 63(3):334–348, 2019. 1.1

[65] John J. Hamre. Remarks to Association of the United States Army Association of Old Crows. `https://cryptome.org/jya/dod081499.htm`, April 1999. Accessed:

2023-03-01. 1

[66] Karolina Hansen, Cindy Littwitz, and Sabine Sczesny. The social perception of heroes and murderers: Effects of gender-inclusive language in media reports. *Frontiers in psychology*, 7:369, 2016. B.1

[67] David R Heise. *Expressive order: Confirming sentiments in social actions*. Springer Science & Business Media, 2007. B.1, E.2.3

[68] Lawrence A Hosman. Language and persuasion. *The persuasion handbook: Developments in theory and practice*, pages 371–390, 2002. B.1

[69] Kristina Hristakieva, Stefano Cresci, Giovanni Da San Martino, Mauro Conti, and Preslav Nakov. The spread of propaganda by coordinated communities on social media. In *WebSci '22: 14th ACM Web Science Conference 2022, Barcelona, Spain, June 26 - 29, 2022*, pages 191–201. ACM, 2022. doi: 10.1145/3501247.3531543. URL https://doi.org/10.1145/3501247.3531543. 5.3.1

[70] Binxuan Huang and Kathleen M Carley. Disinformation and misinformation on twitter during the novel coronavirus outbreak. *arXiv preprint arXiv:2006.04278*, 2020. 1.3.1, 5.2.2

[71] Kathleen Hall Jamieson and Dolores Albarracin. The relation between media consumption and misinformation at the outset of the sars-cov-2 pandemic in the us. *The Harvard Kennedy School Misinformation Review*, 1(2):1–22, 2020. 5.2.1

[72] John Hopkins Coronavirus Research Center. Covid-19 dashboard. https://coronavirus.jhu.edu/map.html. Accessed: 2021-03-21. 5.2.1

[73] India R Johnson, Evava S Pietri, David M Buck, and Roua Daas. What's in a pronoun: Exploring gender pronouns as an organizational identity-safety cue among sexual and gender minorities. *Journal of Experimental Social Psychology*, 97:104194, 2021. B.1

[74] Neil F Johnson, Nicolas Velásquez, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud, Minzhang Zheng, Pedro Manrique, Stefan Wuchty, and Yonatan Lupu. The online competition between pro-and anti-vaccination views. *Nature*, 582 (7811):230–233, 2020. 5.2.2

[75] Joint Chiefs of Staff. Joint Publication 3-13: Information operations. *Chairman of the Joint Chief of Staff Publications*, November 2014. (document), 1.1.1, 1.1.2, 4.4.1, 4.4

[76] Joint Chiefs of Staff. Joint Publication 3-13: Public Affairs. *Chairman of the Joint Chief of Staff Publications*, August 2016. 1.1.2

[77] Johanna Kissler, Cornelia Herbert, Peter Peyk, and Markus Junghofer. Buzzwords: early cortical responses to emotional words during reading. *Psychological science*, 18(6):475–480, 2007. 3.3

[78] David Klepper. Thousands of pro-trump bots are attacking desantis, haley. *Associated Press*. URL https://apnews.com/article/trump-desantis-twitter-haley-presidential-election-4d61487294f9218855b 2.4.6

[79] Gokul S Krishnan, S Sowmya Kamath, and Vijayan Sugumaran. Predicting vaccine hes-

itancy and vaccine sentiment using topic modeling and evolutionary optimization. In *Natural Language Processing and Information Systems: 26th International Conference on Applications of Natural Language to Information Systems, NLDB 2021, Saarbrücken, Germany, June 23–25, 2021, Proceedings*, pages 255–263. Springer, 2021. 5.2.2

[80] Sumeet Kumar. *Social Media Analytics for Stance Mining: A Multi-Modal Approach with Weak Supervision*. PhD thesis, Carnegie Mellon University, 2020. Section 4.3.1. 1.3.2, 5.2.3, 5.3.2

[81] Per E Kummervold, Sam Martin, Sara Dada, Eliz Kilich, Chermain Denny, Pauline Paterson, and Heidi J Larson. Categorizing vaccine confidence with a transformer-based machine learning model: analysis of nuances of vaccine sentiment in twitter discourse. *JMIR medical informatics*, 9(10):e29584, 2021. 5.2.2

[82] Label Studio. Label studio: Open source data labeling platform. `https://labelstud.io/`. Accessed: 2023-04-09. 3.2.1

[83] Ashley Lopez. Turnout among young voters was the second highest for a midterm in past 30 years. *National Public Radio*, 2022. URL `https://www.npr.org/2022/11/10/1135810302/turnout-among-young-voters-was-the-second-highest-for-a-midterm-in-pas` 5.3.1

[84] Daniel P Maki and Maynard Thompson. Mathematical models and applications: with emphasis on the social life, and management sciences. Technical report, 1973. E.2.1

[85] Andrew Mathews, Anne Richards, and Michael Eysenck. Interpretation of homophones related to threat in anxiety states. *Journal of abnormal psychology*, 98(1):31, 1989. 3.3

[86] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001. E.3.1

[87] Shahan Ali Memon, Aman Tyagi, David R Mortensen, and Kathleen M Carley. Characterizing sociolinguistic variation in the competing vaccination communities. In *Social, Cultural, and Behavioral Modeling: 13th International Conference, SBP-BRiMS 2020, Washington, DC, USA, October 18–21, 2020, Proceedings 13*, pages 118–129. Springer, 2020. 5.2.2

[88] MITRE. MITRE/ATT&CK. `https://attack.mitre.org/`. Accessed: 2023-03-01. 1.1.2

[89] Karin Mogg, Brendan P Bradley, Tim Miller, Henry Potts, Joanna Glenwright, and John Kentish. Interpretation of homophones related to threat: Anxiety or response bias effects? *Cognitive Therapy and Research*, 18:461–477, 1994. 3.3

[90] Aleksandra Mojsilović, José Gomes, and Bernice Rogowitz. Semantic-friendly indexing and quering of images based on the extraction of the objective semantic cues. *International Journal of Computer Vision*, 56:79–107, 2004. 6.4

[91] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen Carley. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In *Proceedings of the international AAAI conference on web and social media*, volume 7, pages

400–408, 2013. E.1

[92] Luke Munn. More than a mob: Parler as preparatory media for the u.s. capitol storming. *First Monday*, 26(3), Feb. 2021. doi: 10.5210/fm.v26i3.11574. URL https://firstmonday.org/ojs/index.php/fm/article/view/11574. 5.3.1

[93] Engineering National Academies of Sciences, Medicine, et al. A decadal survey of the social and behavioral sciences: A research agenda for advancing intelligence analysis. 2019. 5.2.2

[94] Matthew L Newman, James W Pennebaker, Diane S Berry, and Jane M Richards. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5):665–675, 2003. B.1

[95] Lynnette Hui Xian Ng and K Carley. Flipping stance: Social influence on bot's and non bot's covid vaccine stance. 2021. 5.2.2

[96] Lynnette Hui Xian Ng and Kathleen M Carley. Botbuster: Multi-platform bot detection using a mixture of experts. *arXiv preprint arXiv:2207.13658*, 2022. 1.3.1, 4.3.2, 5.3.2

[97] 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. 1.3.1, 5.2.3, 5.3.2, 5.4.4

[98] Ben Nimmo. Anatomy of an info-war: How Russia's propaganda machine works, and how to counter it. https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-t
Accessed: 2023-03-01. 1.1, 1.1.2, 2.4.2

[99] Office of the Secretary of Defense for Public Affairs. DOD Instruction 5400.17: Official Use of Social Media for Public Affairs Purposes. *Department of Defense Instruction*, January 2023. 1.1.2

[100] Georgios Paltoglou, Mathias Theunis, Arvid Kappas, and Mike Thelwall. Predicting emotional responses to long informal text. *IEEE transactions on affective computing*, 4(1): 106–115, 2012. 3.3

[101] James Pamment. The EU's role in fighting disinformation: Crafting a disinformation framework. https://carnegieendowment.org/2020/09/24/eu-s-role-in-fighting-disinformation-crafting-disinformation-framework
Accessed: 2022-09-01. 1.1, 1.1.2, 4.1.1

[102] James W Pennebaker. Putting stress into words: Health, linguistic, and therapeutic implications. *Behaviour research and therapy*, 31(6):539–548, 1993. 3.3

[103] James W Pennebaker. The secret life of pronouns. *New Scientist*, 211(2828):42–45, 2011. 3.3, B.1

[104] James W Pennebaker and Martha E Francis. Cognitive, emotional, and language processes in disclosure. *Cognition & emotion*, 10(6):601–626, 1996. 3.3

[105] James W Pennebaker and Lori D Stone. Words of wisdom: language use over the life span. *Journal of personality and social psychology*, 85(2):291, 2003. B.1

[106] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001. 3.3

[107] James W. Pennebaker, Matthias R. Mehl, and Kate G. Niederhoffer. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1):547–577, 2003. ISSN 0066-4308. 3.3

[108] Pew Research Center: Internet, Science Tech. Pew Research Center: demographics of social media users and adoption in the United States. `https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/ABC_Framework_TWG_Francois_Sept_2019.pdf`, 2021. Accessed: 2021-04-26. E.3.2

[109] Krystle Phirangee and Jim Hewitt. Loving this dialogue!!!!: Expressing emotion through the strategic manipulation of limited non-verbal cues in online learning environments. In *Emotions, technology, and learning*, pages 69–85. Elsevier, 2016. 3.3, B.1

[110] Gregory A Poland and Ray Spier. Fear, misinformation, and innumerates: how the wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine*, 28(12): 2361–2362, 2010. 5.2.1

[111] Dorothy Porter and Roy Porter. The politics of prevention: anti-vaccinationism and public health in nineteenth-century england. *Medical history*, 32(3):231–252, 1988. 5.2.1

[112] Vladimir Putin. Transcript: Putin says Russia will protect the rights of Russians abroad. `https://www.washingtonpost.com/world/transcript-putin-says-russia-will-protect-the-rights-of-russians-abroa`
`2014/03/18/432a1e60-ae99-11e3-a49e-76adc9210f19_story.html`, 2014. Accessed: 2023-04-09. 5.4.2

[113] Vladimir Putin. On the historical unity of russians and ukrainians. *President of Russia*, 12, 2021. 5.4.2

[114] Paul Rozin, Loren Berman, and Edward Royzman. Biases in use of positive and negative words across twenty natural languages. *Cognition and Emotion*, 24(3):536–548, 2010. B.1

[115] Sam Schechner and Stacy Meichtry. How Zelensky and Putin are using online media in the war for ukraine; ukraine's president takes to social media to rally global support, as the kremlin restricts news and social media in russia, Feb 27 2022. URL `https://www.proquest.com/newspapers/how-zelensky-putin-are-using-online-media-war/docview/2633546489/se-2`. Copyright - Copyright 2022 Dow Jones Company, Inc. All Rights Reserved; Last updated - 2022-10-15. 5.3.1

[116] Ana Lucía Schmidt, Fabiana Zollo, Antonio Scala, Cornelia Betsch, and Walter Quattrociocchi. Polarization of the vaccination debate on facebook. *Vaccine*, 36(25):3606–3612, 2018. 5.2.2, 5.2.2

[117] Elena Schneider and Holly Otterbein. 'the central issue': How the fall of roe v. wade shook the 2022 election. *Politico*, 2022. URL `https://www.politico.com/news/`

2022/12/19/dobbs-2022-election-abortion-00074426. 5.3.1

[118] Emilio Serrano, Carlos Ángel Iglesias, and Mercedes Garijo. A novel agent-based rumor spreading model in twitter. In *Proceedings of the 24th International Conference on World Wide Web*, pages 811–814, 2015. E.1, E.3

[119] Shadi Shahsavari, Pavan Holur, Tianyi Wang, Timothy R Tangherlini, and Vwani Roychowdhury. Conspiracy in the time of corona: automatic detection of emerging covid-19 conspiracy theories in social media and the news. *Journal of computational social science*, 3(2):279–317, 2020. 5.2.1

[120] 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. 5.3.2

[121] Nate Silver. 2022 fivethirtyeight election forecast. *FiveThirtyEight*, 2022. URL https://projects.fivethirtyeight.com/2022-election-forecast/. 5.3.2

[122] Statista. Statista. http://https://www.statista.com/. Accessed: 2023-03-15. 1.1

[123] Erin M Sumner, Luisa Ruge-Jones, and Davis Alcorn. A functional approach to the facebook like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons. *New Media & Society*, 20(4):1451–1469, 2018. 3.3

[124] 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. ISSN 0261-927X. 3.3, 3.6, B.1

[125] Mike Thelwall, Kayvan Kousha, and Saheeda Thelwall. Covid-19 vaccine hesitancy on english-language twitter. *Profesional de la información (EPI)*, 30(2), 2021. 5.2.2

[126] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019. 1.2.2, 5.3.1

[127] Alina Trifan, Rui Antunes, Sérgio Matos, and Jose Luís Oliveira. Understanding depression from psycholinguistic patterns in social media texts. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42*, pages 402–409. Springer, 2020. B.1

[128] Twitter. Covid-19: Our approach to misleading vaccine information, Dec 2020. URL https://blog.twitter.com/en_us/topics/company/2020/covid19-vaccine.html. Accessed: 05-16-2021. 2.4.5, 5.2.1, 5.2.4

[129] Twitter. Updates to our work on covid-19 vaccine misinformation, Mar 2021. URL https://blog.twitter.com/en_us/topics/company/2021/updates-to-our-work-on-covid-19-vaccine-misinformation.html. (Last accessed on 05-16-2021). 2.4.5, 5.2.1

[130] Twitter. About different types of tweets, 2023. URL https://help.twitter.com/en/using-twitter/types-of-tweets. 1.2.1, 5.4.4

[131] Twitter. How to get the blue checkmark on Twitter, 2023. URL https://help.twitter.com/en/managing-your-account/

about-twitter-verified-accounts. 1.2.1, 4.3.2

[132] Joshua Uyheng, Thomas Magelinski, Ramon Villa-Cox, Christine Sowa, and Kathleen M Carley. Interoperable pipelines for social cyber-security: assessing Twitter information operations during NATO Trident Juncture 2018. *Computational and Mathematical Organization Theory*, 26:465–483, 2020. 4.3.4, 5.2.3

[133] Jose Antonio Vargas. Spring awakening: how an Egyptian revolution began on Facebook. *New York Times*, 2012. URL https://www.nytimes.com/2012/02/19/books/review/how-an-egyptian-revolution-began-on-facebook.html. 5.3.1

[134] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *science*, 359(6380):1146–1151, 2018. 5.4.2

[135] Joseph B Walther and Kyle P D'addario. The impacts of emoticons on message interpretation in computer-mediated communication. *Social science computer review*, 19(3): 324–347, 2001. 3.3, B.1

[136] Chao Wang, Zong Xuan Tan, Ye Ye, Lu Wang, Kang Hao Cheong, and Neng-gang Xie. A rumor spreading model based on information entropy. *Scientific reports*, 7(1):9615, 2017. E.1

[137] Weining Wang and Qianhua He. A survey on emotional semantic image retrieval. In *2008 15th IEEE International Conference on Image Processing*, pages 117–120. IEEE, 2008. 6.4

[138] Carol Waseleski. Gender and the use of exclamation points in computer-mediated communication: An analysis of exclamations posted to two electronic discussion lists. *Journal of Computer-Mediated Communication*, 11(4):1012–1024, 2006. B.1

[139] Lilian Weng, Alessandro Flammini, Alessandro Vespignani, and Fillipo Menczer. Competition among memes in a world with limited attention. *Scientific reports*, 2(1):335, 2012. E.3.1

[140] Evan M Williams and Kathleen M Carley. TSPA: Efficient target-stance detection on twitter. 2022. 1.3.2, 5.3.2

[141] Graham K Wilson. Brexit, Trump and the special relationship. *The British Journal of Politics and International Relations*, 19(3):543–557, 2017. doi: 10.1177/1369148117713719. URL https://doi.org/10.1177/1369148117713719. 5.3.1

[142] Stefan Wojcik and Adam Hughes. Sizing up Twitter users, 2019. URL https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/. 5.2.5, 5.3.3

[143] Yang Yang, Fumin Shen, Heng Tao Shen, Hanxi Li, and Xuelong Li. Robust discrete spectral hashing for large-scale image semantic indexing. *IEEE Transactions on Big Data*, 1(4):162–171, 2015. 6.4

[144] Xiaoyi Yuan, Ross J Schuchard, and Andrew T Crooks. Examining emergent communities and social bots within the polarized online vaccination debate in twitter. *Social media+ society*, 5(3):2056305119865465, 2019. 1.3.1, 5.2.2

[145] Damián H Zanette. Dynamics of rumor propagation on small-world networks. *Physical review E*, 65(4):041908, 2002. E.2.1

[146] Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: audio, visual and spontaneous expressions. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 126–133, 2007. 6.4

[147] Hengshun Zhou, Debin Meng, Yuanyuan Zhang, Xiaojiang Peng, Jun Du, Kai Wang, and Yu Qiao. Exploring emotion features and fusion strategies for audio-video emotion recognition. In *2019 International conference on multimodal interaction*, pages 562–566, 2019. 6.4

Appendix A

Data Set Information

A.1 List of terms for Data Collection

Table A.1 contains the list of terms used to collect the data for research regarding the 2022 U.S. Midterm Elections. They were collected from November 1-9, 2022 using the Twitter API. The terms are related to the Senate, House, and Gubernatorial races within seven swing states during the US midterm election.

A.2 List of Terms for Stance Detection

The terms in Table A.2 were assigned to each political party and used to seed the stance detection algorithms. The agents and hashtags were subsequently classified into Democrat, Republican, or unassigned.

Table A.1: 2022 U.S. Midterm Election (Swing States) Collection Terms

Swing State	Keywords
Arizona	(AZ02 OR AZ01 OR AZ06 OR AZSen OR AZGov) OR ((Kelly OR Blake OR Lake OR Hobbs OR Crane OR Halleran OR Hodge OR Schweikert OR Engel OR Ciscomani) (vote OR election OR elect OR race OR AZ OR Arizona OR democrat OR republican))
Georgia	(GA06 OR GASen) OR ((Warnock OR Walker OR Kemp OR Abrams OR McBath OR Handel) (vote OR election OR elect OR race OR GA OR Georgia OR democrat OR republican))
Nevada	(NVSen OR NVGov OR NV03 OR NV04 OR NV01) OR ((Mastro OR Laxalt OR Sisolak OR Lombardo OR Becker OR Lee OR Peters OR Hosford OR Robertson OR Titus) (vote OR election OR elect OR race OR NV OR Nevada OR democrat OR republican))
North Carolina	(Beasley OR Budd OR NCSen OR Nickel OR Hines OR NC13) AND (vote OR election OR elect OR race OR NC OR North Carolina OR democrat OR republican)
Ohio	(Vance OR Ryan OR OHSen OR DeWine OR Whaley OR OHGov OR Chabot OR Landsman OR OH01 OR Sykes OR Gilbert OR OH13 OR Kaptur OR Majewski OR OH09) AND (vote OR election OR elect OR race OR OH OR Ohio OR democrat OR republican)
Pennsylvania	(PASen OR PAGov OR PA07 OR PA08 OR PA17) OR ((Oz OR Fetterman OR Shapiro OR Mastriano OR Scheller OR Wild OR Bognet OR Cartwright OR Shaffer OR Deluzio) (vote OR election OR elect OR race OR PA OR Pennsylvania OR democrat OR republican))
Wisconsin	(OHSen OR OHGov OR OH01 OR OH13 OR OH09) OR ((Vance OR Ryan OR DeWine OR Whaley OR Chabot OR Landsman OR Sykes OR Gilbert OR Kaptur OR Majewski) (vote OR election OR elect OR race OR OH OR Ohio OR democrat OR republican))

Table A.2: 2022 U.S. Midterm Election (Democrat/Republican) Agent and Hashtag Stance Detection Terms

Political Stance	Keywords
Democrat	VoteBlueForDemocracy, VoteBlueToProtectYourRights, VoteBlue, BlueTsunami, wtpBLUE, DemVoice1, VoteBlueToSaveDemocracy, VoteBlueIn2022, BlueWave, TruBlue, OurBlueVoice, PuppetVance, Dems4USA, BlueIn22, Fetterman, ProudBlue22, BlueVoices, Warnock, GeorgiaNeedsStaceyAbrams, voteblue2022, LawlessLaxalt, NotYourBudd, NoToBo, VoteBlueIn22, AntiJobsJohnson, VoteBlueTomorrow
Republican	Doug4Gov, VoteRed, RedWave, RepublicansForKatieHobbs, DrOz, VoteRed2022, Walker, Vance, VoteRedToSaveAmerica2022, HerschelWalker, RedWave2022, LeadRight, GoRedStateByState, VoteredtosaveAmerica, Laxalt, RedTsunami2022, voterepublican, JDVance, RepublicansForRyan, Budd, VoteRedForFreedom, RedTsunami, redstatebystate, VoteRedTomorrow, RedOrAmericaIsDEAD, VoteREDToSaveThisCountry

Appendix B

List of CUES and Descriptions

Table B.1 list the current CUES and descriptions used in the Netmapper software.

Table B.1: CUES indicators descriptions and BEND maneuvers (*Indicators not CUES terms)

CUES/indicators	Description	Maneuver	References
named entity	Number of named entities	explain	
abusive	Number of abusive terms	nuke, neutralize, narrow, dismay, distract	[42, 48]
expletive	Number of expletive terms	nuke, neutralize, dismay, distract	[48]
exclusive	Number of exclusive terms	narrow	
inclusive	Number of inclusive terms	boost, bridge	[46, 66]
absolutist	Number of absolutist terms	build, back, excite, nuke, neutralize, dismay, distract	[1, 14, 127]
equivocal	Number of equivocal terms	explain, neutralize, distort	[7, 63, 68]
connective	Number of connective terms	enhance	

Continued on next page

Table B.1 – continued from previous page

CUES/indicators	Description	Maneuver	References
numbers	Number of numbers	explain	
# exclamation points	Number of exclamation points	excite, distort	[109, 138]
# question marks	Number of question marks	distort, distract	
Is in all caps	Number of capital letters	-	[56]
positive	Number of positive terms	build, back, boost, engage, excite, distract	[114]
negative	Number of negative terms	nuke, neutralize, distract	[114]
# happy emots/emojis (emoticons_happy)	Number of happy emoticons/emojis	back, boost, engage, excite	[49, 56, 109, 135]
# sad emots/emojis (emoticons_sad)	Number of sad emoticons/emojis	nuke, neutralize	[49, 56, 109, 135]
# angry emots/emojis (emoticons_angry)	Number of angry emoticons/emojis	nuke, neutralize	[49, 56, 109, 135]
# embarrassed emots/emojis (emoticons_embarrassed)	Number of embarrassed emoticons/emojis	nuke, neutralize	[49, 56, 109, 135]
# positive emoticons	Number of positive emoticons	-	[49, 56, 109, 135]
# positive emoji	Number of positive emojis	back, boost, engage, excite, distract	[49, 56, 109, 135]
# neutral emoticons	Number of neutral emoticons	-	[49, 56, 109, 135]
# neutral emoji	Number of neutral emojis	-	[49, 56, 109, 135]
# negative emoticons	Number of negative emoticons	-	[49, 56, 109, 135]
# negative emoji	Number of negative emojis	nuke, neutralize, distract	[49, 56, 109, 135]
poweranger	Number of power anger terms	nuke, neutralize, dismay	[40, 59, 67]
powerencourage	Number of power encourage terms	back, enhance, excite	[59, 67]
powerfear	Number of power fear terms	neutralize, dismay	[59, 67]
powerforbidden	Number of power forbidden terms	narrow, dismay	[59, 67]
powergreed	Number of power greed terms	excite	
powerlust	Number of power lust terms	excite	

Continued on next page

Table B.1 – continued from previous page

CUES/indicators	Description	Maneuver	References
powersafety	Number of power safety terms	enhance	
1st person	Number of 1st person terms	excite	[73, 103, 105, 124]
2nd person	Number of 2nd person terms	-	[73, 103, 105, 124]
3rd person	Number of 3rd person terms	-	[73, 103, 105, 124]
# mv_care_virtue	Number of moral value for care terms	excite	[27, 28, 29, 30, 31]
# mv_care_vice_harm	Number of moral value for harm terms	dismay	[27, 28, 29, 30, 31]
# mv_fairness_virtue	Number of moral value for fairness terms	back	[27, 28, 29, 30, 31]
# mv_fairness_vice_cheating	Number of moral value for cheating terms	neutralize, dismay	[27, 28, 29, 30, 31]
# mv_loyalty_virtue	Number of moral value for loyalty terms	back, boost	[27, 28, 29, 30, 31]
# mv_loyalty_vice_betrayal	Number of moral value for betrayal terms	nuke	[27, 28, 29, 30, 31]
# mv_authority_virtue	Number of moral value for authority terms	back	[27, 28, 29, 30, 31]
# mv_authority_vice_subversion	Number of moral value for subversion terms	neutralize	[27, 28, 29, 30, 31]
# mv_sanctity_virtue	Number of moral value for sanctity terms	enhance	[27, 28, 29, 30, 31]
# mv_sanctity_vice_degradation	Number of moral value for degradation terms	distract	[27, 28, 29, 30, 31]
# mv_liberty_virtue	Number of moral value for liberty terms	-	[27, 28, 29, 30, 31]
# mv_liberty_vice_oppression	Number of moral value for oppression terms	-	[27, 28, 29, 30, 31]

Continued on next page

Table B.1 – continued from previous page

CUES/indicators	Description	Maneuver	References
# back	Number of back terms	back	
# build	Number of build terms	build	
# bridge	Number of bridge terms	bridge	
# boost	Number of boost terms	boost	
# engage	Number of engage terms	engage	
# explain	Number of explain terms	explain	
# excite	Number of excite terms	excite	[59, 67]
# enhance	Number of enhance terms	enhance	
# neutralize	Number of neutralize terms	neutralize	
# nuke	Number of nuke terms	nuke	
# narrow	Number of narrow terms	narrow	
# neglect	Number of neglect terms	neglect	
# dismiss	Number of dismiss terms	dismiss	
# distort	Number of distort terms	distort	[2, 94]
# dismay	Number of dismay terms	dismay	[59, 67]
# distract	Number of distract terms	distract	
Agent-references count*	Number of agent references	build, boost	
References-exactly-one-agent*	TRUE if has exactly one agent reference	back, neutralize	
Completely-off-topic concepts-count*	Number of terms not used by any other documents in the corpus	bridge, engage, distract, narrow	
References-at-least-two-agents*	TRUE if references at least two agents	bridge	
References-at-least-two-communities*	TRUE if references at least two communities	bridge	
On-topic-concepts-count*	Number of concepts similar to other documents	engage, explain, excite, dismay	
images-count*	Number of images	engage	

Continued on next page

Table B.1 – continued from previous page

CUES/indicators	Description	Maneuver	References
propagates-count (propagation-count)*	Number propagations	engage, excite	
urls-count*	Number of URLs	engage, explain	
govt-agent-reference count*	Number of references to government agents	explain	
has-less-than-two-negatives*	TRUE if has less than two negative terms	explain	
has-less-than-two-positives*	TRUE if has less than two positive terms	explain	
media-agency-reference count*	Number of media agencies referenced	explain	
is-propagator*	TRUE if is a propagator (retweet, reply, quote)	explain	
off-topic-concepts count*	Number of terms used by relatively few other documents in the corpus	enhance	
negative-sarcasm*	TRUE if has negative sarcasm	excite, dismiss	
positive-sarcasm*	TRUE if has positive sarcasm	excite, dismiss, dismay	
concept-specialization*	Number terms used by relatively few other documents in the corpus	narrow, distract	
has-topic-concepts-and-strong-cues*	Number of terms used by relatively many other documents and number of CUES that appear at least twice in the document	dismiss	
has-topic-words-and-laughter-emos*	Number of terms used by relatively many other documents and number of laughter emojis/emoticons	dismiss	

Appendix C

CUES Usage Correlation by Maneuver

The following graphs depict the correlations of CUES usage within messages for each of the BEND maneuvers. The brown boxes indicate uncollected data.

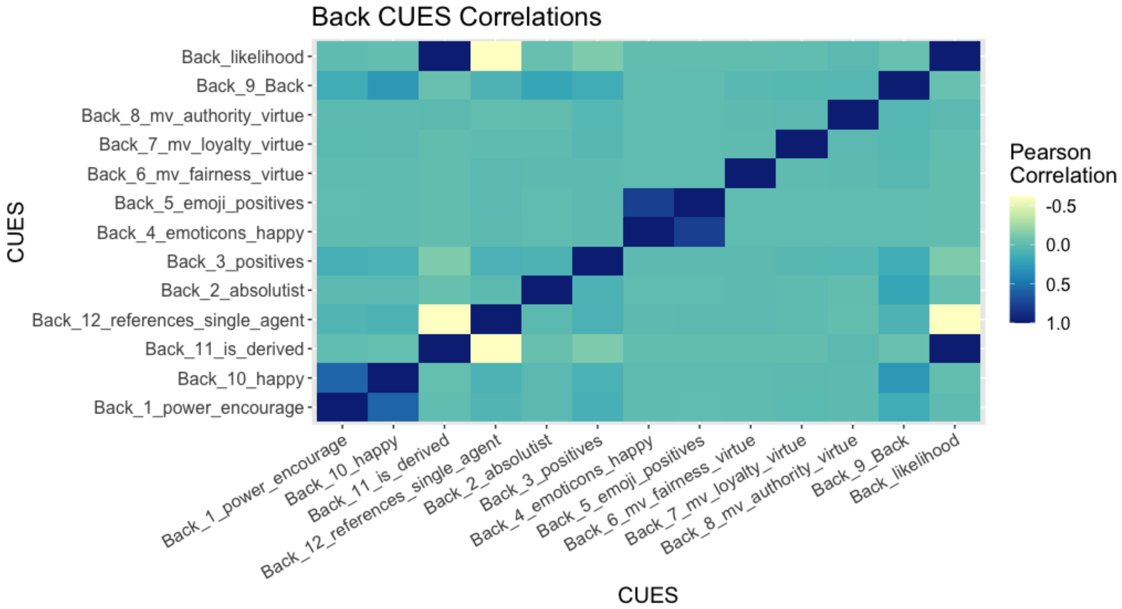


Figure C.1: Correlations between *back* CUES

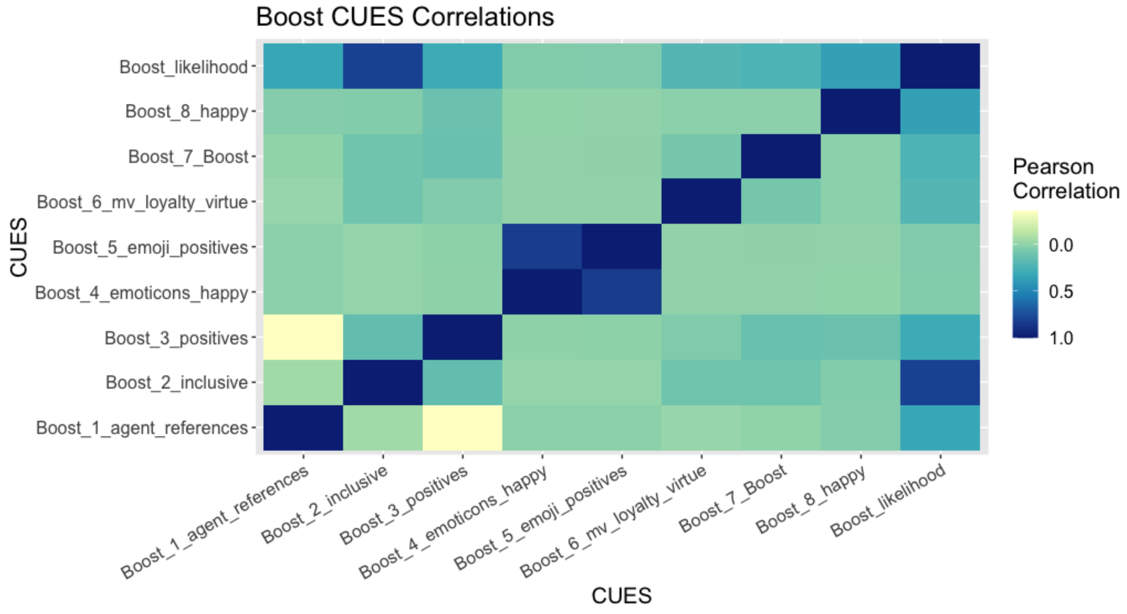


Figure C.2: Correlations between *boost* CUES

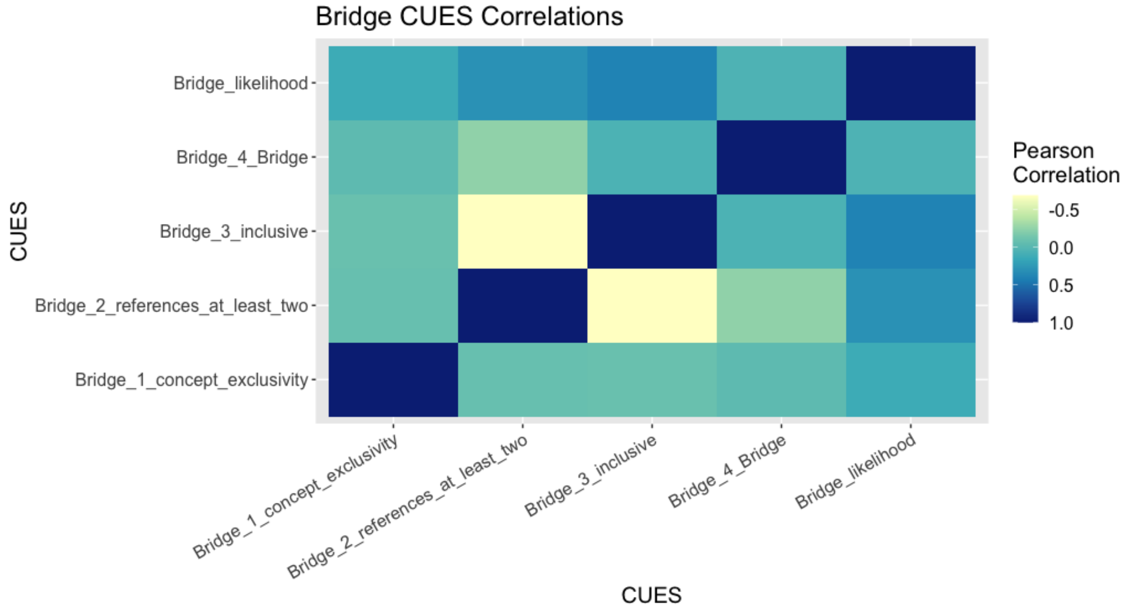


Figure C.3: Correlations between *bridge* CUES

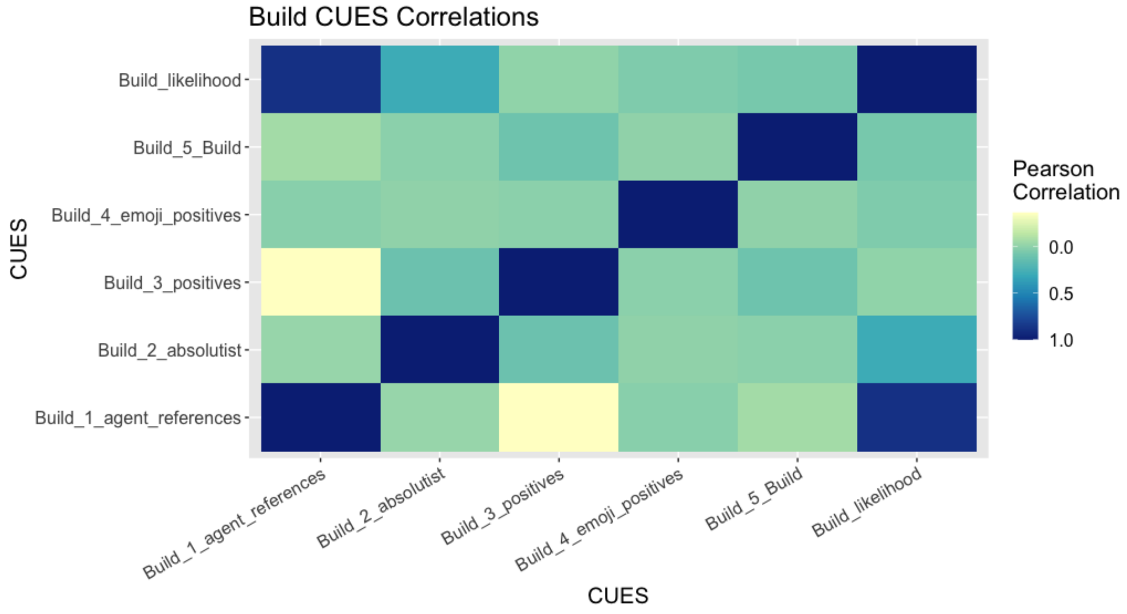


Figure C.4: Correlations between *build* CUES

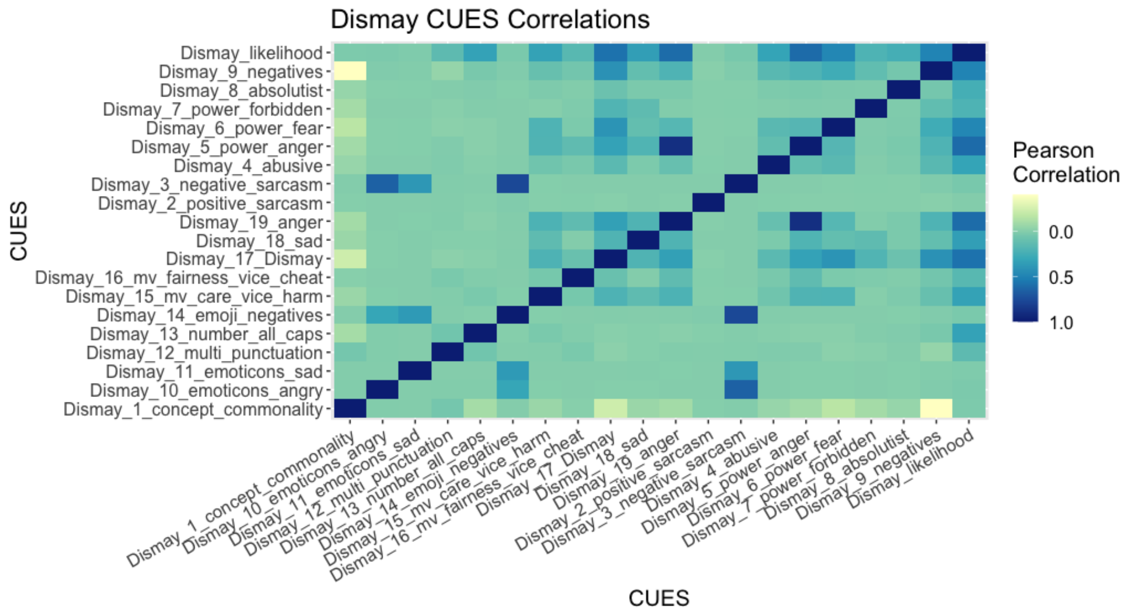


Figure C.5: Correlations between *dismay* CUES

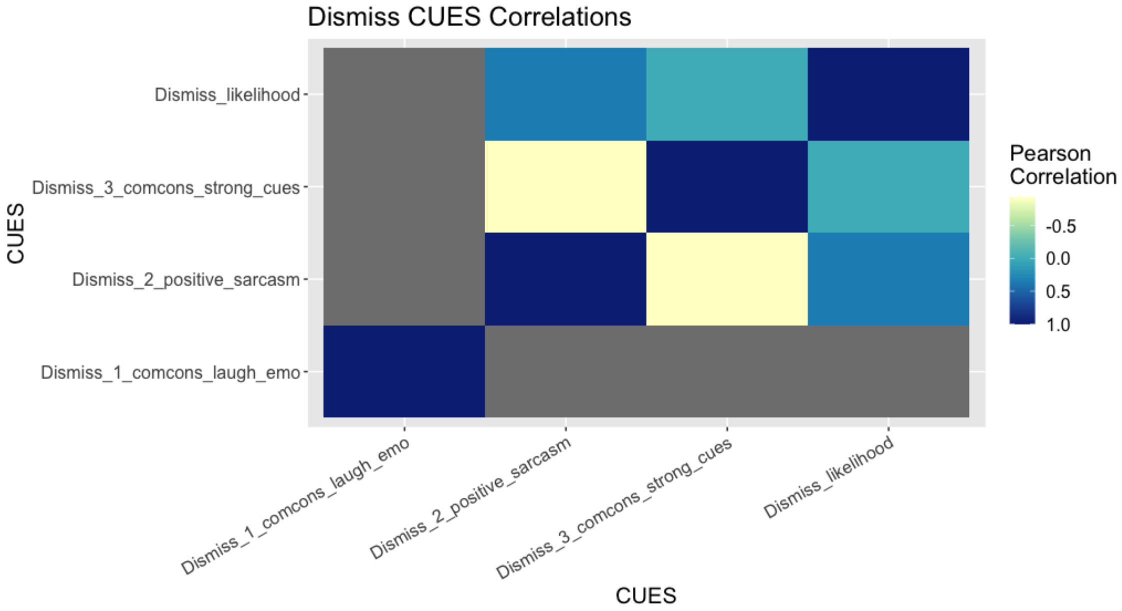


Figure C.6: Correlations between *dismiss* CUES

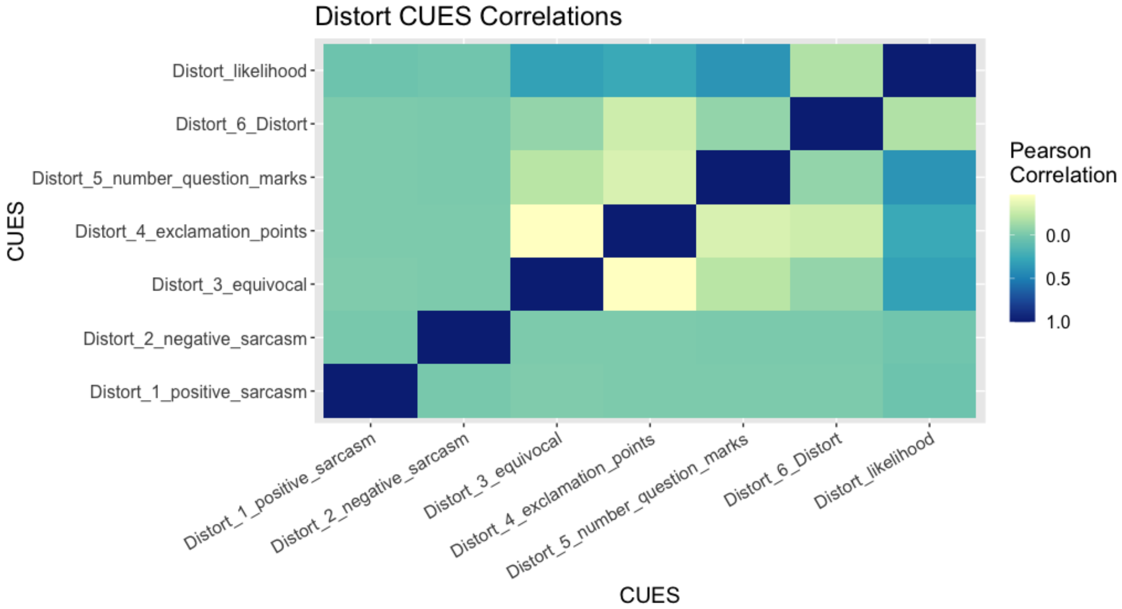


Figure C.7: Correlations between *distort* CUES

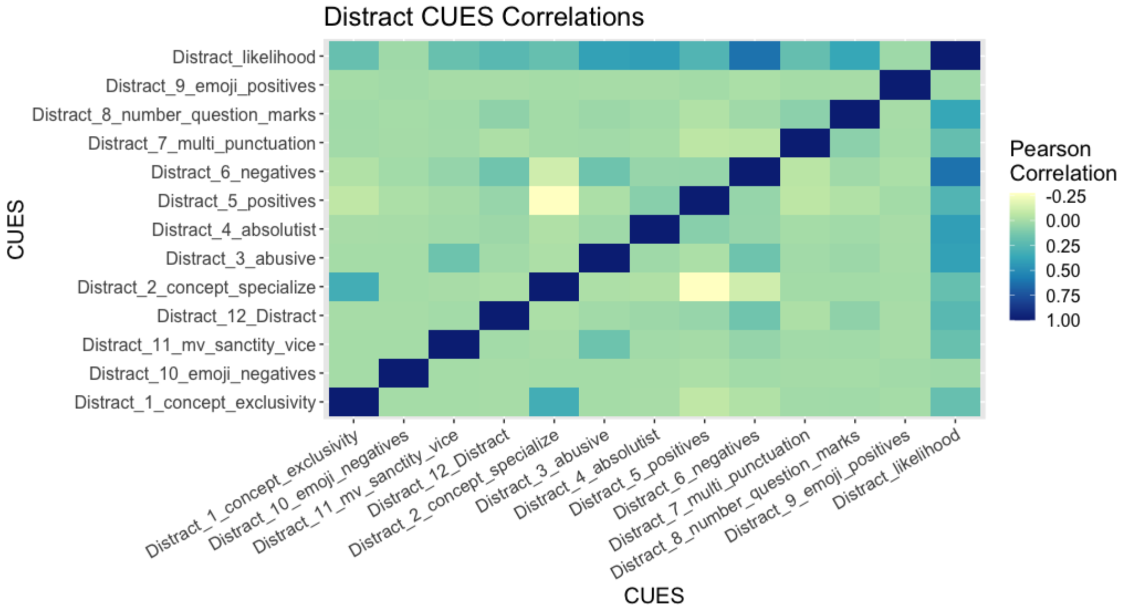


Figure C.8: Correlations between *distract* CUES

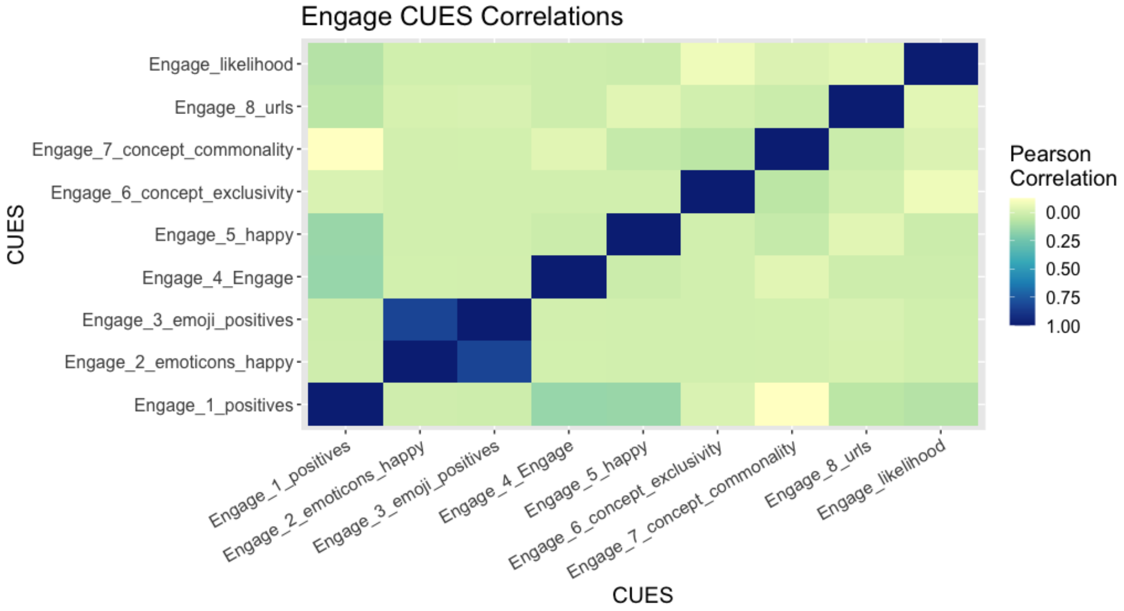


Figure C.9: Correlations between *engage* CUES

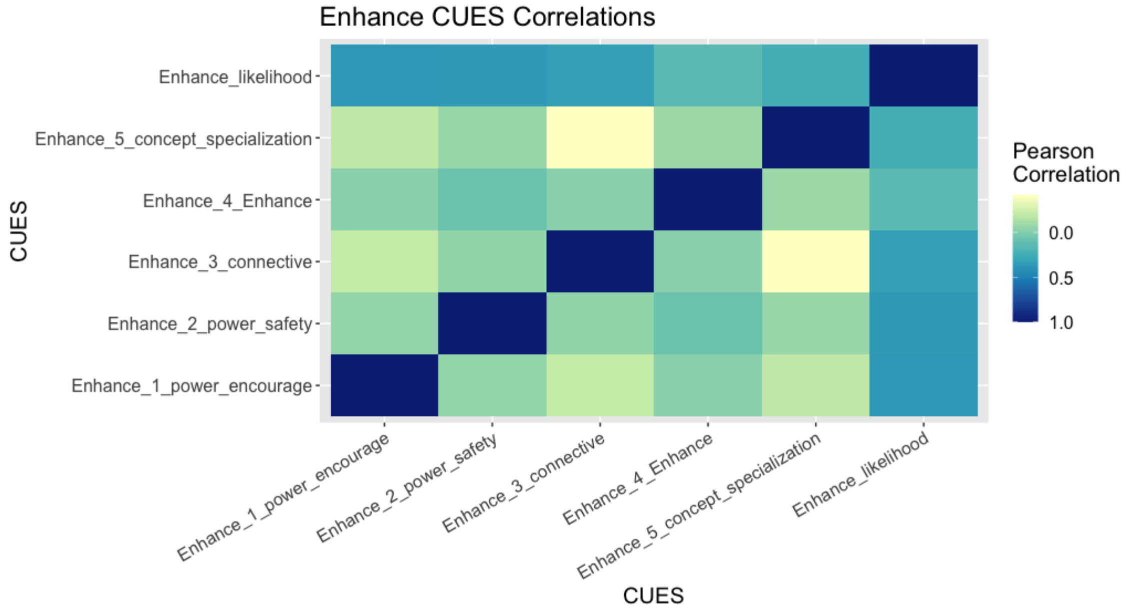


Figure C.10: Correlations between *enhance* CUES

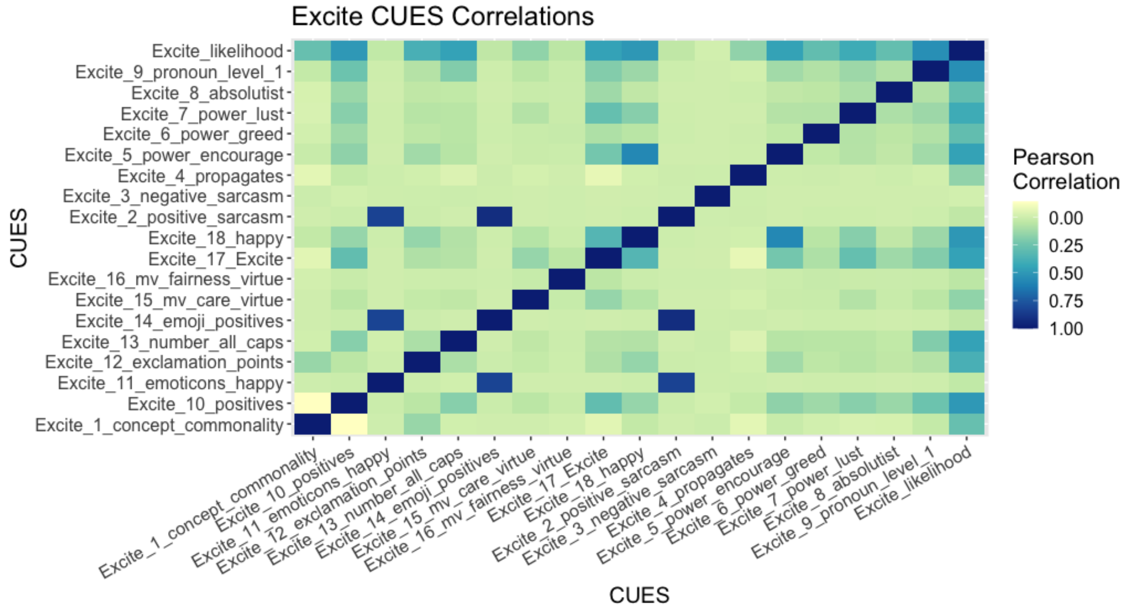


Figure C.11: Correlations between *excite* CUES

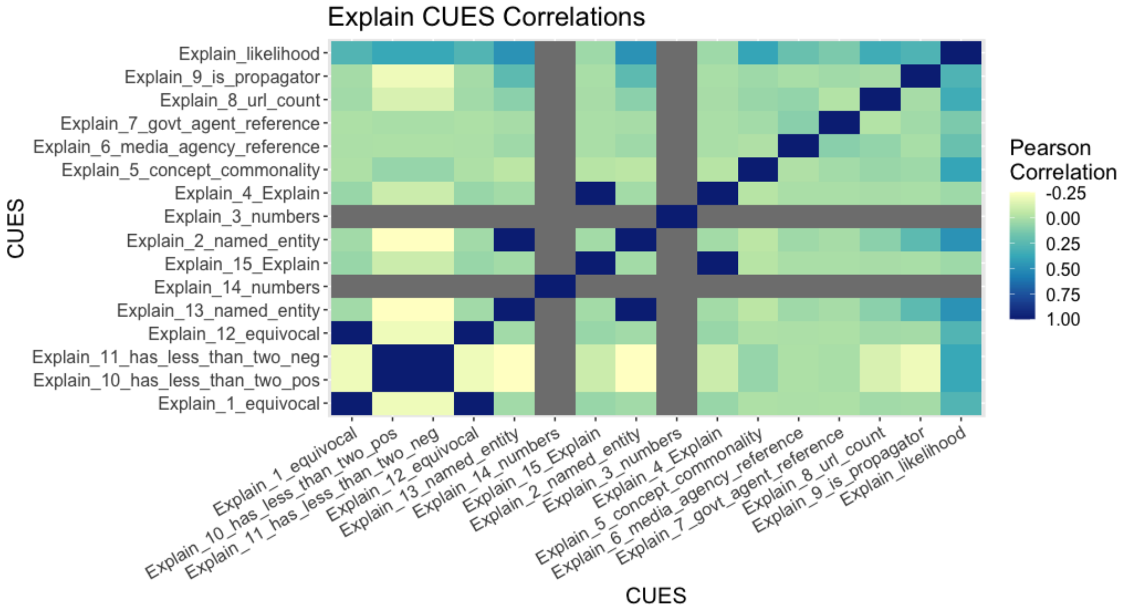


Figure C.12: Correlations between *explain* CUES

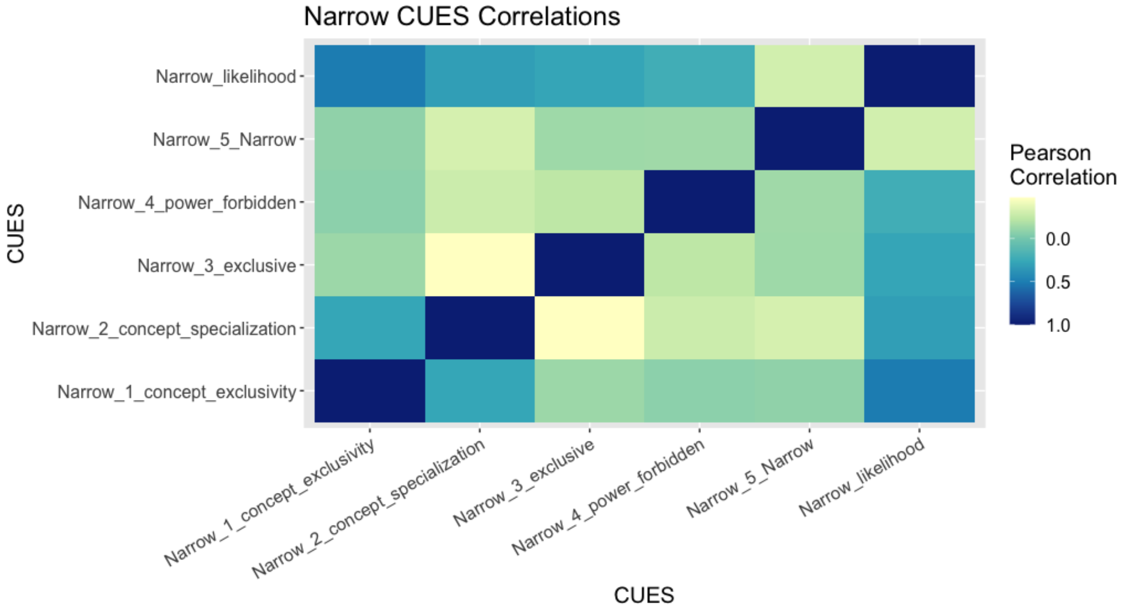


Figure C.13: Correlations between *narrow* CUES

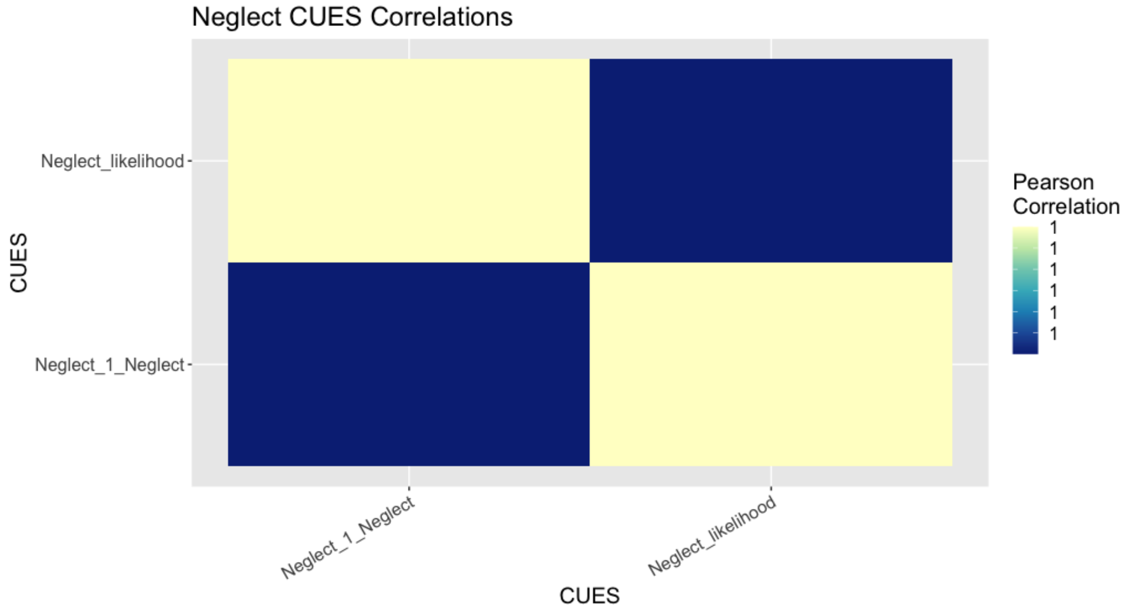


Figure C.14: Correlations between *neglect* CUES

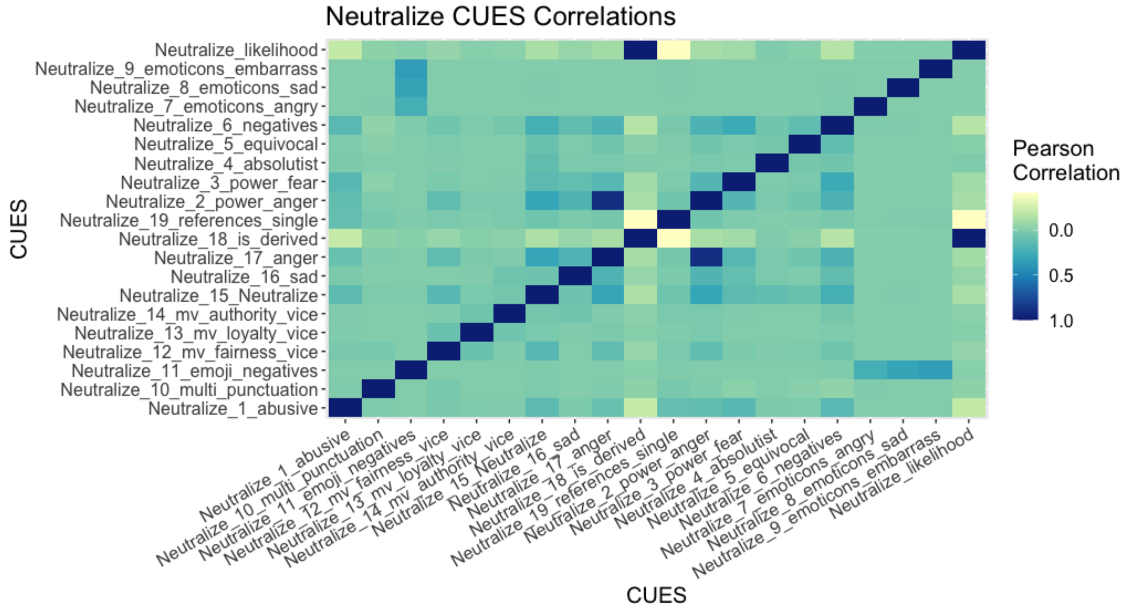


Figure C.15: Correlations between *neutralize* CUES

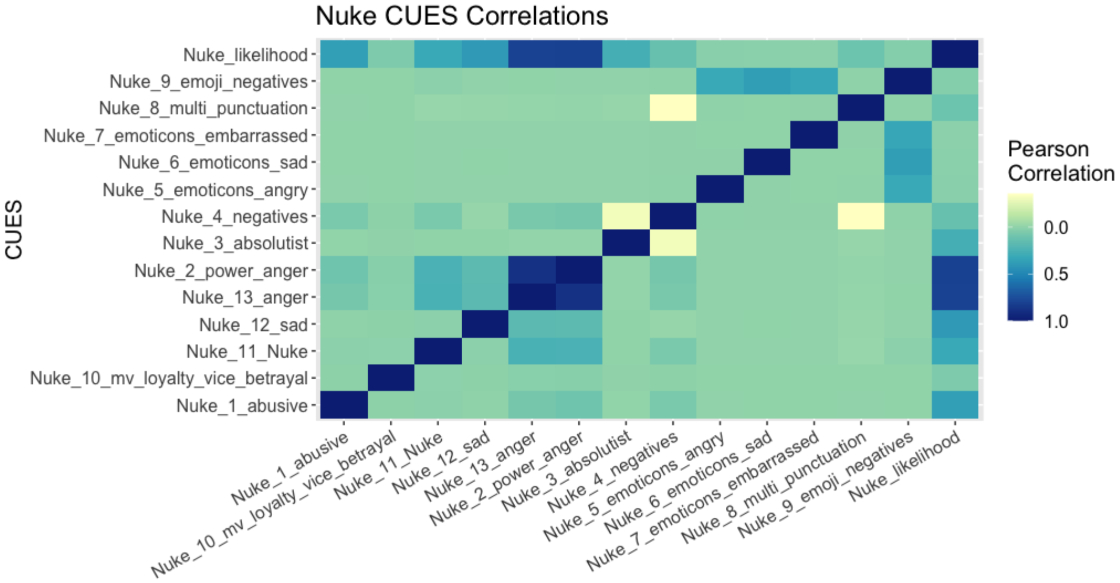


Figure C.16: Correlations between *nuke* CUES

Appendix D

BEND Framework Tutorials

D.1 How to Import CUES

CUES are calculated using the Netmapper software [39]. The CUES document is a .tsv file that can be pipelined into ORA-PRO as input for calculating the BEND maneuvers within a data set. This section outlines the process to import the values as tweet attributes for a given data set. Options not discussed within this tutorial should use the ORA-PRO defaults.

D.1.1 Select *Import Attributes* function (Figure D.1)

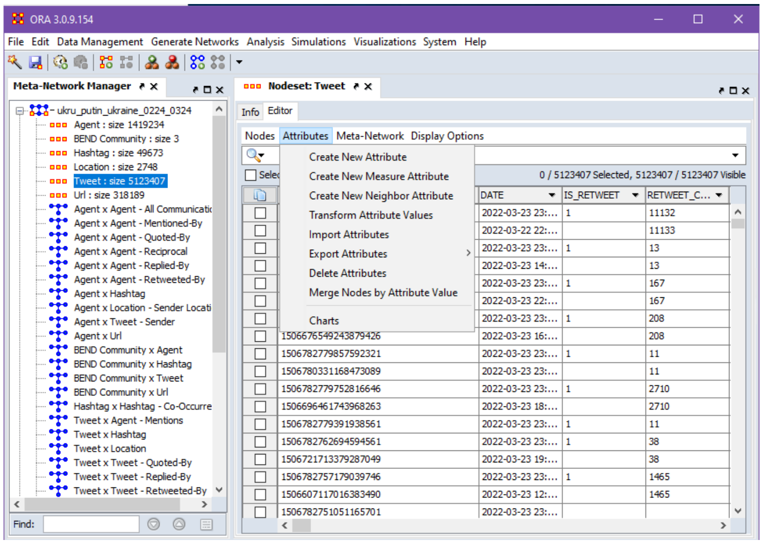
1. Select the *Tweet nodeset* in *Meta-Network Manager*
2. Select *Editor* tab
3. Select *Attributes* menu
4. Click on *Import attributes*

D.1.2 Import Attributes Parameters (Figure D.2)

1. Click *Browse* to select the CUES document. This should be .tsv file
2. Match *Node ID* with file column *twitter_id*. This ensures that the CUES from the tweets detected using Netmapper are added as attributes to the corresponding tweets in the data set
3. Use the *Actions* button to *Select All* and then de-select the *Author* and *Date* checkboxes. These values typically already exist in the data set

D.1.3 Import Successful (Figure D.3)

1. An indication that the CUES were successfully imported are the correct listings of CUES in the *Attribute* column and numerical values in the *# Imported Values* column
2. Select OK

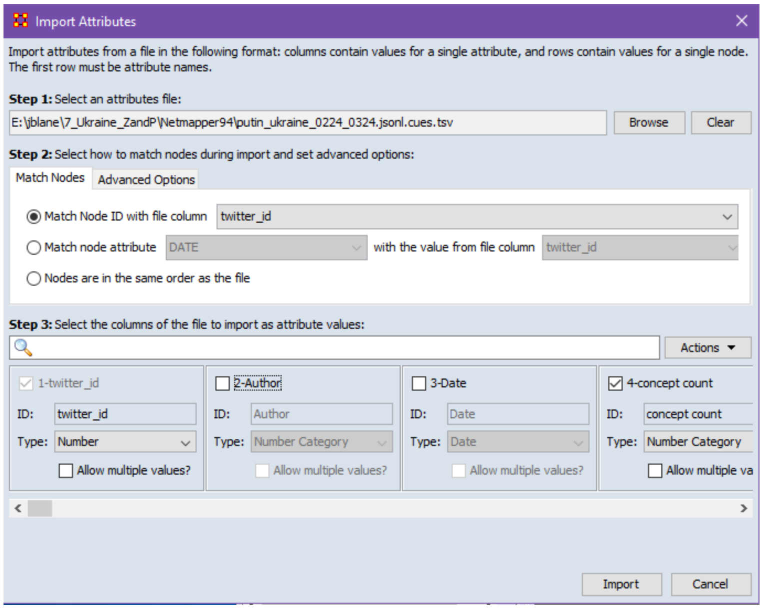


ORA 3.0.9.154
File Edit Data Management Generate Networks Analysis Simulations Visualizations System Help
Meta-Network Manager
ukru_putin_ukraine_0224_0324
Agent : size 1419234
BEND Community : size 3
Hashtag : size 49673
Location : size 2748
Tweet : size 5123407
Url : size 318189
Agent x Agent - All Communicatio
Agent x Agent - Mentioned-By
Agent x Agent - Quoted-By
Agent x Agent - Reciprocal
Agent x Agent - Replied-By
Agent x Agent - Retweeted-By
Agent x Hashtag
Agent x Location - Sender Locati
Agent x Tweet - Sender
Agent x Url
BEND Community x Agent
BEND Community x Hashtag
BEND Community x Tweet
BEND Community x Url
Hashtag x Hashtag - Co-Occurre
Tweet x Agent - Mentions
Tweet x Hashtag
Tweet x Location
Tweet x Tweet - Quoted-By
Tweet x Tweet - Replied-By
Tweet x Tweet - Retweeted-By
Find:
Nodeset: Tweet
Info Editor
Nodes Attributes Meta-Network Display Options
Create New Attribute
Create New Measure Attribute
Create New Neighbor Attribute
Transform Attribute Values
Import Attributes
Export Attributes
Delete Attributes
Merge Nodes by Attribute Value
Charts
Selec

0 / 5123407 Selected, 5123407 / 5123407 Visible

		DATE	IS_RETWEET	RETWEET_C...
□		2022-03-23 23:...	1	11132
□		2022-03-22 22:...		11133
□		2022-03-23 23:...	1	13
□		2022-03-23 14:...		13
□		2022-03-23 23:...	1	167
□		2022-03-23 22:...		167
□		2022-03-23 23:...	1	208
□	1506676549243879426	2022-03-23 16:...		208
□	1506782779857592321	2022-03-23 23:...	1	11
□	1506780331168473089	2022-03-23 23:...		11
□	1506782779752816646	2022-03-23 23:...	1	2710
□	1506696461743968263	2022-03-23 18:...		2710
□	1506782779391938561	2022-03-23 23:...	1	11
□	1506782762694594561	2022-03-23 23:...	1	38
□	1506721713379287049	2022-03-23 19:...		38
□	1506782757179039746	2022-03-23 23:...	1	1465
□	1506607117016383490	2022-03-23 12:...		1465
□	1506782751051165701	2022-03-23 23:...		

Figure D.1: Select Import Attributes



Import attributes from a file in the following format: columns contain values for a single attribute, and rows contain values for a single node. The first row must be attribute names.

Step 1: Select an attributes file:

E:\jblane\7_Ukraine_ZandP\Netmapper94\putin_ukraine_0224_0324.jsonl.cues.tsv

Step 2: Select how to match nodes during import and set advanced options:

● Match Node ID with file column twitter_id

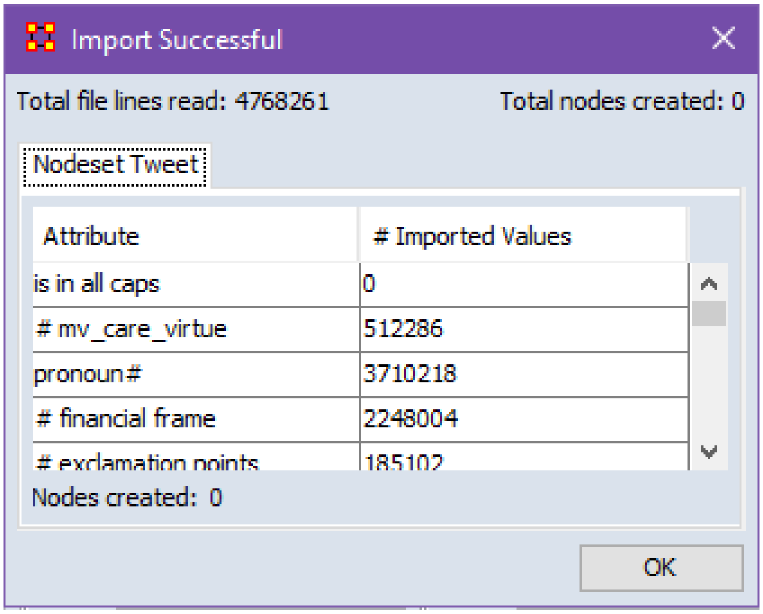
○ Match node attribute DATE with the value from file column twitter_id

○ Nodes are in the same order as the file

Step 3: Select the columns of the file to import as attribute values:

☑ 1-twitter_id	□ 2-Author	□ 3-Date	☑ 4-concept count
ID: twitter_id	ID: Author	ID: Date	ID: concept count
Type: Number	Type: Number Category	Type: Date	Type: Number Category
□ Allow multiple values?	□ Allow multiple values?	□ Allow multiple values?	□ Allow multiple va

Figure D.2: Import Attributes Parameters



Import Successful

Total file lines read: 4768261

Total nodes created: 0

Attribute	# Imported Values
is in all caps	0
# mv_care_virtue	512286
pronoun#	3710218
# financial frame	2248004
# exclamation points	185102

Nodes created: 0

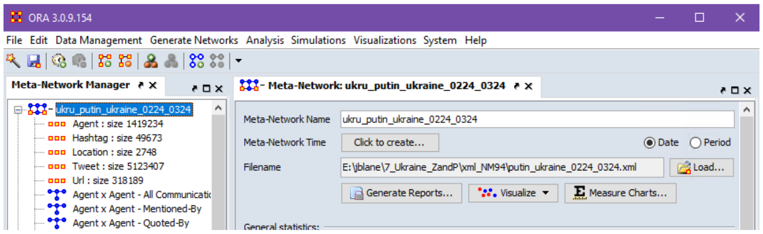
Figure D.3: Import Successful

D.2 How to Run a BEND Report

This section outlines the tutorial for running a basic BEND report with explanations for why some features are selected. Options not discussed within this basic tutorial should use the ORA-PRO defaults. However, other features may be used depending on the data set and the desired analysis.

D.2.1 Generate Report (Figure D.4)

1. Select the *meta-network* in *Meta-Network Manager*
2. Click *Generate Reports...*


ORA 3.0.9.154
File Edit Data Management Generate Networks Analysis Simulations Visualizations System Help
Meta-Network Manager
Meta-Network: ukru_putin_ukraine_0224_0324
Generate Reports...
Visualize
Measure Charts...
Load...
General statistics:

ukru_putin_ukraine_0224_0324
- Agent : size 1419234
- Hashtag : size 49673
- Location : size 2748
- Tweet : size 5123407
- Url : size 318189
- Agent x Agent - All Communicatic
- Agent x Agent - Mentioned-By
- Agent x Agent - Quoted-By

Meta-Network Name: ukru_putin_ukraine_0224_0324

Meta-Network Time: Click to create... ● Date ○ Period

Filename: E:\jblane\7_Ukraine_ZandP\xml_NM94\putin_ukraine_0224_0324.xml

Figure D.4: Generate Report

D.2.2 Select BEND & Community Assessment Report (Figure D.5)

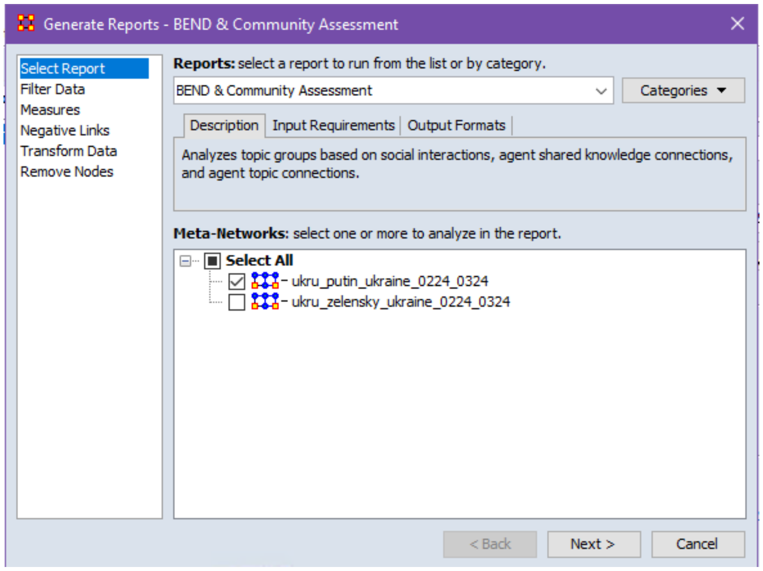
1. Select the BEND & Community Assessment Report
2. Ensure the appropriate data set(s) are selected
3. Click *Next*

D.2.3 Select Nodesets (Figure D.6)

1. In the menu, select *Select Nodesets*
2. Check *Agent nodeset* checkbox. Select the *Agent* nodeset in the drop-down menu to indicate who the agents are interacting within the data set
3. Check *Document nodeset* checkbox. Select the *Tweet* nodeset in the drop-down menu to indicate the individual documents that are being analyzed in this report
4. Check only the *Hashtag* checkbox. For the purposes of this tutorial, hashtags will serve the function of representing *concepts* or single ideologies within a message. Concepts are used to indicate a main idea, theme, or topic within a message.

D.2.4 Agent attributes (Figure D.7)

1. In the menu, select *Agent Attributes*



Reports: select a report to run from the list or by category.

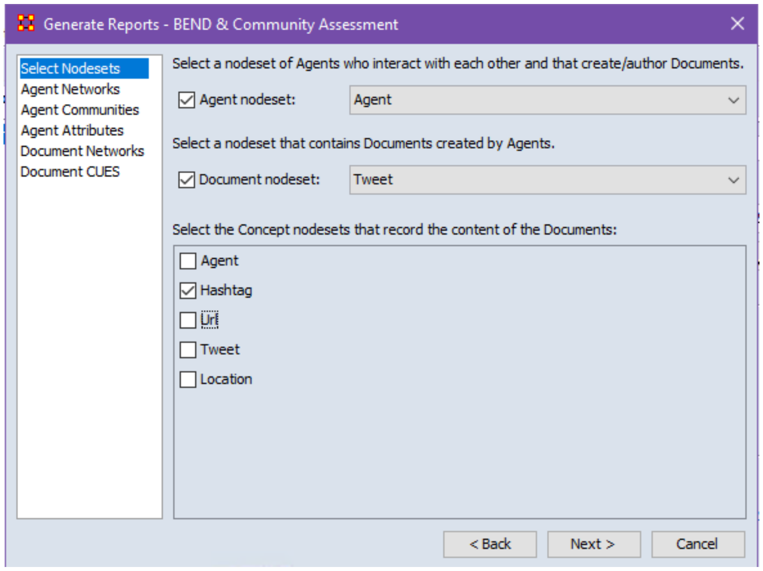
BEND & Community Assessment

Analyzes topic groups based on social interactions, agent shared knowledge connections, and agent topic connections.

Meta-Networks: select one or more to analyze in the report.

- ■ **Select All**
  - ☑ - ukru_putin_ukraine_0224_0324
  - □ - ukru_zelensky_ukraine_0224_0324

Figure D.5: Select BEND & Community Assessment report



Select a nodeset of Agents who interact with each other and that create/author Documents.

☑ Agent nodeset: Agent

Select a nodeset that contains Documents created by Agents.

☑ Document nodeset: Tweet

Select the Concept nodesets that record the content of the Documents:

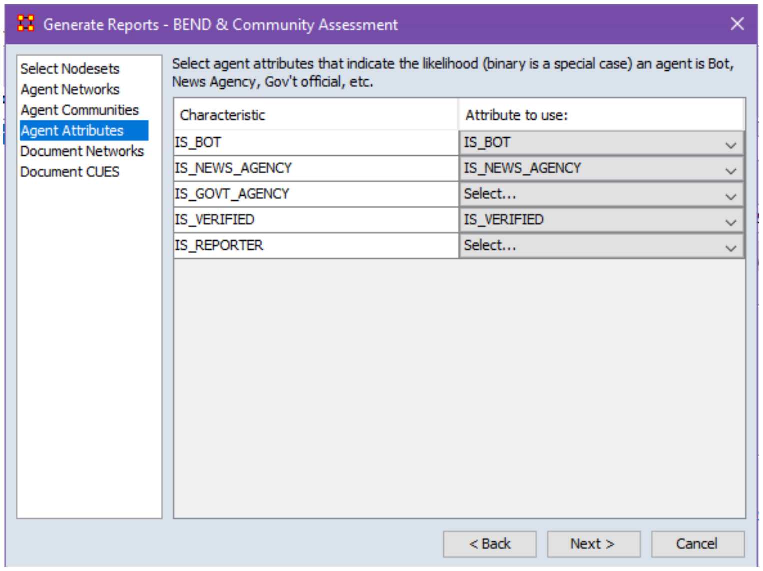
□ Agent
☑ Hashtag
□ Url
□ Tweet
□ Location

Figure D.6: Select Nodesets

2. If there is an attribute within the data set that identifies bots or other agent characteristics (the attribute must be a *category* type), select the attribute that identifies the items in the drop-down menus.

D.2.5 Document CUES (Figure D.8)

1. In the menu, select *Document CUES*. These should reflect the attributes imported into the Tweet nodeset from the CUES .tsv file.

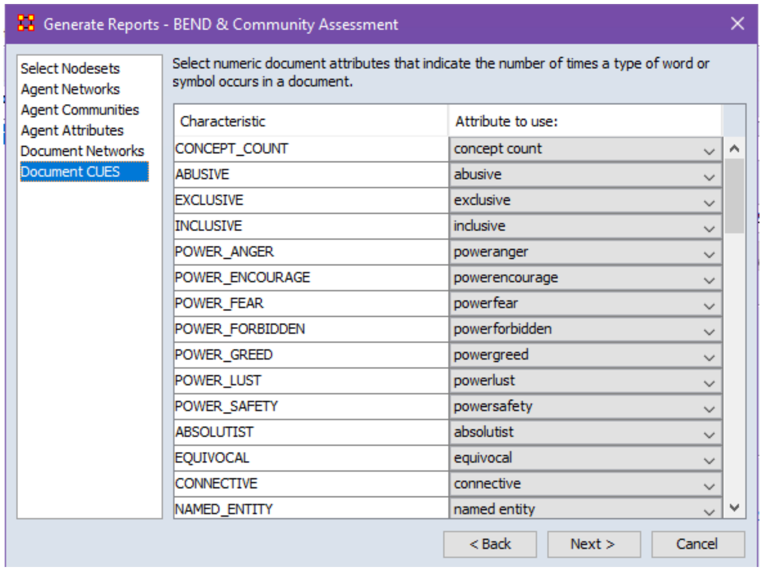


Select agent attributes that indicate the likelihood (binary is a special case) an agent is Bot, News Agency, Gov't official, etc.

Characteristic	Attribute to use:
IS_BOT	IS_BOT
IS_NEWS_AGENCY	IS_NEWS_AGENCY
IS_GOVT_AGENCY	Select...
IS_VERIFIED	IS_VERIFIED
IS_REPORTER	Select...

Figure D.7: Select Agent Attributes

2. Click *Next*



Select numeric document attributes that indicate the number of times a type of word or symbol occurs in a document.

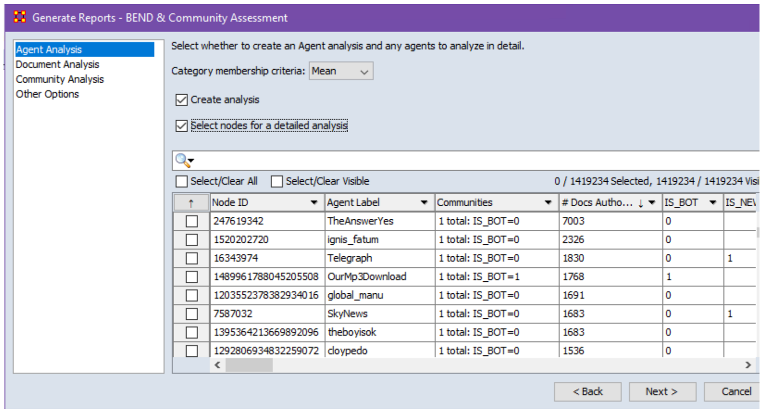
Characteristic	Attribute to use:
CONCEPT_COUNT	concept count
ABUSIVE	abusive
EXCLUSIVE	exclusive
INCLUSIVE	inclusive
POWER_ANGER	poweranger
POWER_ENCOURAGE	powerencourage
POWER_FEAR	powerfear
POWER_FORBIDDEN	powerforbidden
POWER_GREED	powergreed
POWER_LUST	powerlust
POWER_SAFETY	powersafety
ABSOLUTIST	absolutist
EQUIVOCAL	equivocal
CONNECTIVE	connective
NAMED_ENTITY	named entity

Figure D.8: Select DocumentCUES

D.2.6 Agent Analysis (Figure D.9)

1. Select *Mean* for the *Category membership criteria*
2. Check Create analysis

3. If desired, check *Select nodes for detailed analysis*. There may be key actors or other interesting agents that may need further inspection. Some interesting features for agents to examine include those with high *# Docs Authored* or containing interesting attributes or concepts. Select these actors to create a focused BEND analysis on the individual account.



Select whether to create an Agent analysis and any agents to analyze in detail.

Category membership criteria: Mean

☑ Create analysis

☑ Select nodes for a detailed analysis

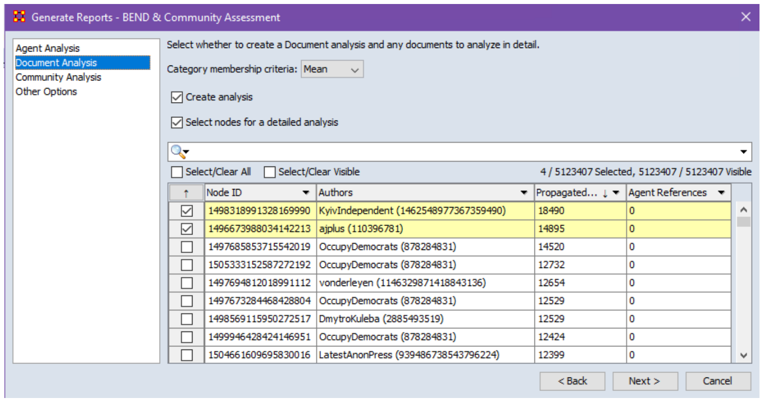
□ Select/Clear All □ Select/Clear Visible 0 / 1419234 Selected, 1419234 / 1419234 Visi

↑	Node ID	Agent Label	Communities	# Docs Autho... ↓	IS_BOT	IS_NEV
□	247619342	TheAnswerYes	1 total: IS_BOT=0	7003	0	
□	1520202720	ignis_fatum	1 total: IS_BOT=0	2326	0	
□	16343974	Telegraph	1 total: IS_BOT=0	1830	0	1
□	1489961788045205508	OurMp3Download	1 total: IS_BOT=1	1768	1	
□	1203552378382934016	global_manu	1 total: IS_BOT=0	1691	0	
□	7587032	SkyNews	1 total: IS_BOT=0	1683	0	1
□	1395364213669892096	theboyisok	1 total: IS_BOT=0	1683	0	
□	1292806934832259072	cloypedo	1 total: IS_BOT=0	1536	0	

Figure D.9: Agent Analysis

D.2.7 Document Analysis (Figure D.10)

1. Select *Mean* for the *Category membership criteria*
2. If desired, check *Select nodes for detailed analysis*. There may be specific documents (Tweets) that may require further inspection. Some interesting features for documents may be those that were propagated the most or have the most agent references. Select these documents to create a focused BEND analysis of the individual message.
3. Click *Next*



Select whether to create a Document analysis and any documents to analyze in detail.

Category membership criteria: Mean

☑ Create analysis

☑ Select nodes for a detailed analysis

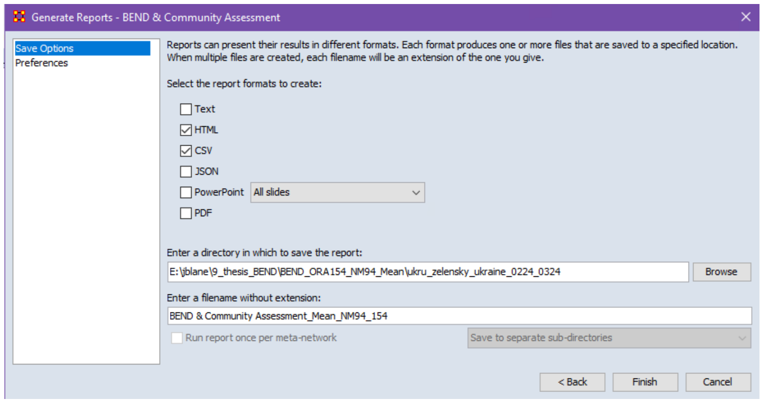
□ Select/Clear All □ Select/Clear Visible 4 / 5123407 Selected, 5123407 / 5123407 Visible

↑	Node ID	Authors	Propagated... ↓	Agent References
☑	1498318991328169990	KyivIndependent (1462548977367359490)	18490	0
☑	1496673988034142213	ajplus (110396781)	14895	0
□	1497685853715542019	OccupyDemocrats (878284831)	14520	0
□	1505333152587272192	OccupyDemocrats (878284831)	12732	0
□	1497694812018991112	vonderleyen (1146329871418843136)	12654	0
□	1497673284468428804	OccupyDemocrats (878284831)	12529	0
□	1498569115950272517	DmytroKuleba (2885493519)	12529	0
□	1499946428424146951	OccupyDemocrats (878284831)	12424	0
□	1504661609695830016	LatestAnonPress (939486738543796224)	12399	0

Figure D.10: Document Analysis

D.2.8 Finish BEND report (Figure D.5)

1. Check HTML for a view of the reports in a web browser. This form is an interactive document with links between each of the supporting documents within the BEND report.
2. If desired, check CSV for the supporting numerical data that is used to populate the overall report. These documents contain information to create plots or analyses that are not automatically calculated within the BEND report.
3. Enter the directory to save the report
4. Enter the name of the filename. In the filename, indicate the version of Netmapper and ORA-PRO used in the calculations of the report as both software are regularly updated.
5. If multiple data sets were selected, check *Ren report once per meta-network* to run a report on each of the data sets separately.
6. If multiple data sets were selected, select *Save to separate subdirectories*
7. Click *Finish*
8. Once completed, an HTML document will open in a web browser with the report. Depending on the size of the data set, the report may take a couple of seconds or sometimes up to an hour or more



Reports can present their results in different formats. Each format produces one or more files that are saved to a specified location. When multiple files are created, each filename will be an extension of the one you give.

Select the report formats to create:

☐ Text
☑ HTML
☑ CSV
☐ JSON
☐ PowerPoint All slides
☐ PDF

Enter a directory in which to save the report:
E:\jblane\9_thesis_BEND\BEND_ORA154_NM94_Mean\ukru_zelensky_ukraine_0224_0324

Enter a filename without extension:
BEND & Community Assessment_Mean_NM94_154

☐ Run report once per meta-network
Save to separate sub-directories

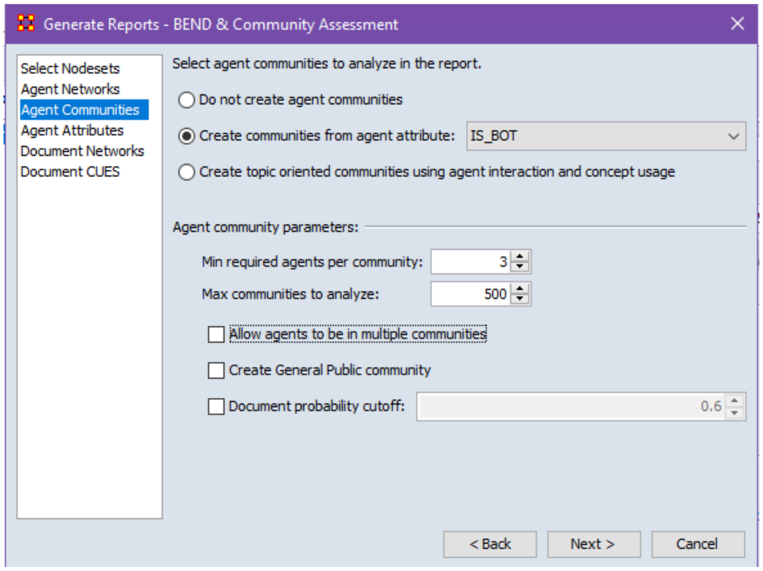
Figure D.11: Finish BEND Report

D.3 How to Create Communities for BEND Analysis

This tutorial outlines how to use the community features in the BEND & Community Assessment report. The process is the same except for the sections specifically designed for the community analysis. This tutorial specifically shows how to create the BEND analysis for both attribute-based and topic-oriented communities as a supplement to the BEND tutorial in Section D.2. Only one type of community can be analyzed per report.

D.3.1 Option 1: Agent Communities (attribute-based) Figure D.12)

1. To analyze communities by attributes, select the radial button for *Create communities from agent attributes*. In the example, communities based on whether an agent is a bot or not a bot are created for analysis
2. Select the *Min required agents per community*
3. Select the *Max communities to analyze*
4. Select *Allow agents to be in multiple communities* if agents should be assigned to multiple groups. In the example with bots or not bots, since agents can only fall in one of the categories, this option is unnecessary
5. Select *Create General Public community* if it is desired to combine all agents that do not fit into a community into a single community. This should only be checked if there is value in analyzing the leftover agents as a community. Otherwise, the community may be unnecessarily large and meaningless for analysis



Select agent communities to analyze in the report.

○ Do not create agent communities

● Create communities from agent attribute: IS_BOT

○ Create topic oriented communities using agent interaction and concept usage

Agent community parameters:

Min required agents per community: 3

Max communities to analyze: 500

□ Allow agents to be in multiple communities

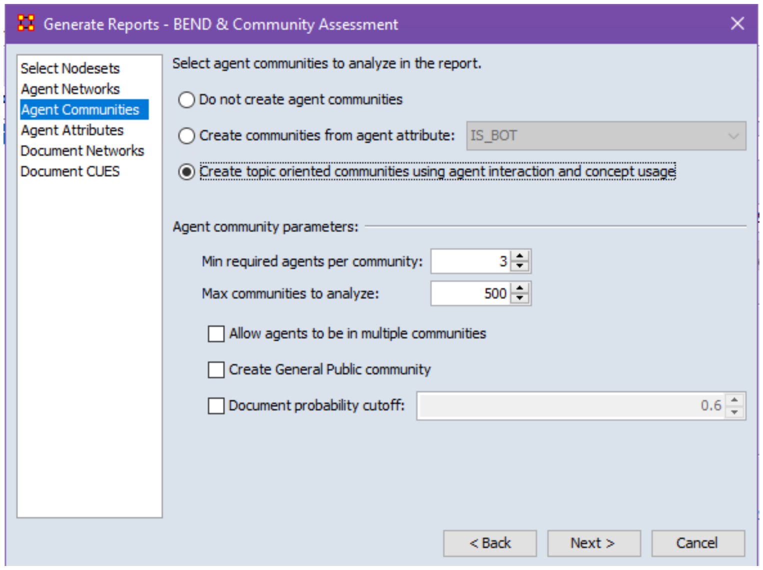
□ Create General Public community

□ Document probability cutoff: 0.6

Figure D.12: Agent Communities (attribute-based)

D.3.2 Option 2: Agent Communities (topic-oriented) Figure D.13)

1. This report uses the Leiden clustering algorithm based on an all-communications network and the *concepts* from the messages to create topic-oriented communities. Select the radial button for *Create topic oriented communities using agent interaction and concept usage* to create topic-oriented communities for analysis.
2. Select the *Min required agents per community*
3. Select the *Max communities to analyze*
4. Select *Allow agents to be in multiple communities* if agents should be assigned to multiple groups.
5. Select *Create General Public community* if it is desired to combine all agents that do not fit into a community into a single community. This should only be checked if there is value in analyzing the leftover agents as a community. Otherwise, the community may be unnecessarily large and meaningless for analysis



Select agent communities to analyze in the report.

○ Do not create agent communities

○ Create communities from agent attribute: IS_BOT

● Create topic oriented communities using agent interaction and concept usage

Agent community parameters:

Min required agents per community: 3

Max communities to analyze: 500

□ Allow agents to be in multiple communities

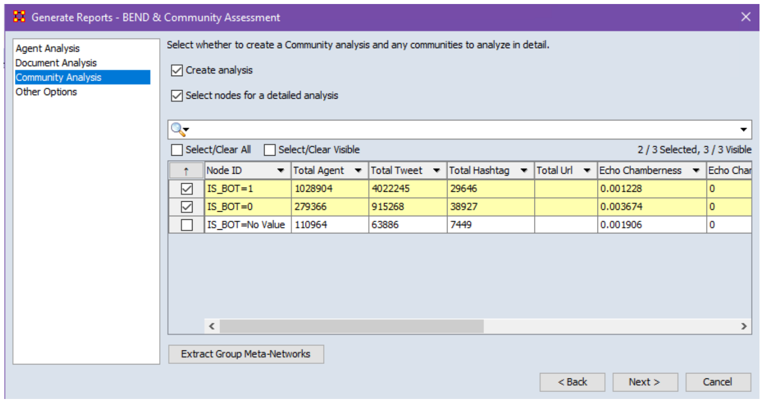
□ Create General Public community

□ Document probability cutoff: 0.6

Figure D.13: Agent Communities (topic-oriented)

D.3.3 Community Analysis Figure D.14)

1. In the menu, select *Community Analysis*
2. Check *Create analysis*
3. Select the communities to conduct a detailed BEND analysis. This can be either the communities created by attributes or the topic-oriented communities. Information such as attributes, number of agents and hashtags, community network calculations such as echo chamberness, and top concepts are features to consider when selecting communities to analyze



Select whether to create a Community analysis and any communities to analyze in detail.

☑ Create analysis

☑ Select nodes for a detailed analysis

□ Select/Clear All □ Select/Clear Visible 2 / 3 Selected, 3 / 3 Visible

↑	Node ID	Total Agent	Total Tweet	Total Hashtag	Total Url	Echo Chamberness	Echo Char
☑	IS_BOT=1	1028904	4022245	29646		0.001228	0
☑	IS_BOT=0	279366	915268	38927		0.003674	0
□	IS_BOT=No Value	110964	63886	7449		0.001906	0

Figure D.14: Community Analysis

Appendix E

BEND Maneuver Simulation

We develop an agent-based model of a Twitter environment to simulate using social-cyber (BEND) maneuvers to deter a disinformation campaign. We explore the use of the network maneuvers of *back*, *build*, and *neutralize* to manipulate the network and the information maneuvers of *excite*, *dismay*, *explain*, and *dismiss* to control the narrative. Using belief as a measure of effectiveness, we explore the changes in user behavior and the resulting network. We demonstrate that *build* is the most effective network maneuver countermeasure for deterrence. The results also show that affecting a tweet's emotional and logical values through information maneuvers effectively controls the overall network belief.

E.1 Introduction

The lines between real and digital life have become indistinguishable. Online platforms have gained distinction as the primary source for news, discussion, sharing ideas, and community building. Maligned actors continuously develop ways to exploit this space to conduct cyber-mediated attacks on civil society. The emerging, transdisciplinary field of social cybersecurity provides theories and tools to help study and counter such cyber-mediated attacks [35].

While recent research in the field of social cybersecurity focuses on establishing a framework to discuss information maneuver [37], social cyber-forensics [16], and diffusion [91] [118] [136], the current pandemic and US election have demonstrated the need for greater understanding of social-cyber maneuvers and how to mitigate them. However, studying social-cyber maneuvers poses two problems: there is no exact way to measure the effects of information operations (open research problem). Two, it is not ethically tractable to manipulate human test subjects' beliefs thoughts. Agent-based modeling (ABM) is a tool that can help researchers overcome the latter of the two problems. ABM provides a powerful method for representing complex and dynamic real-world environments. Until recently, modeling efforts in this domain focused on information diffusion and rumor propagation. In most cases, studies abstract away platform-specific mechanics; `twitter_sim` is the first model our team is aware of that evaluates forms of social-cyber maneuver [18] through simulation.

Our principal research question is: can anti-disinformation maneuvers mitigate disinformation campaigns? This paper explores a social network experiencing a disinformation campaign leveraging the *back* disinformation maneuver to sway network belief. We then implement information and network maneuvers to counter the *back* maneuver using belief to measure the effect of disinformation and mitigating strategies. Our study sets out to improve `twitter_sim` and incorporate more realistic mechanics to expand its ability to model social-cyber maneuvers to counter the spread of disinformation.

E.2 Related Works

E.2.1 Modeling Information and Beliefs

Many models have been used to simulate the spread of information as well as the dynamics on social networks; e.g., [13] [47] [84] and [145]. Construct, an agent-based social-network model, simulates information and belief diffusion using a turn-based social influence approach[34]. Agents have bounded rationality where they use homophily and expertise to help decide with whom they will have interactions. Additionally, agents have general memory of task knowledge, social memory of with whom they are interacting, and a transactive memory of who knows what and whom [36]. This paper builds from Beskow and Carley's `twitter_sim` model [18] and Construct. We further limit the agent's attention by limiting the number of read tweets and exploring more BEND maneuvers. These models are compared in Table 1.

Table E.1: Docking Lite Comparison of `twitter_sim2.0`, `twitter_sim`, and Construct.

Features	`twitter_sim2.0`	`twitter_sim`	Construct
General Population	x	x	
Media Agents	x	x	x
Opinion Leaders	x	x	x
Information Access	x	x	x
General Memory	x	x	x
Transactive Memory			x
Homophily	x	x	x
Limited Attention	x	x	
Dynamic Network	x	x	x
Emotional Response	x		

E.2.2 BEND Framework

The BEND framework provides a way of understanding information/influence operations [35]. The framework provides two types of social-cyber maneuvers for manipulating information, ideas, and beliefs: information (or narrative) maneuvers and network maneuvers. BEND describes 16 maneuvers (which contribute to the acronym for ”BEND”); we incorporate seven to extend the `twitter_sim` model to explore both information and network forms of maneuver.

E.2.3 Modeling Emotions and Reason

Several theories discuss how emotions and cognition affect an individual's behavior. In Heise's Affect Control Theory, people create events or conduct actions that confirm their fundamental sentiments of themselves in a particular situation. Emotions then reflect a person's sentiment about themselves and the validations or invalidations of self-created by the situation at the moment [67]. Behavior is thus related to a combination of the state of the person and the situation.

E.3 Model Description

In `twitter_sim2.0`, we made the discrete event distribution better represent Twitter, modified agent behavior to facilitate BEND maneuvers, and introduced four additional BEND maneuvers for testing. The model is Python 3.7 based. We used the NetworkX package [62] to generate scale-free networks. `Twitter_sim2.0` is a discrete event agent-based model that attempts to replicate individual user behavior and Twitter platform mechanics. The simulation consists of three types of agents- normal users, spreaders (facilitate disinformation operations), and beacons (trusted community leaders or subject matter experts) [118]. Spreaders and beacons both conduct social-cyber maneuvers to change the belief of the simulated network.

E.3.1 Review of `twitter_sim` Features

Homophily (Similarity):

Homophily is the tendency for people to seek out connections with those that are similar to themselves [86]. In this simulation, homophily (similarity) is represented by an adjacency matrix of all nodes in the network; the link values are the Jaccard similarity coefficient between nodes ($similarity = \frac{successors_A \cap successors_B}{successors_A \cup successors_B}$).

Influence:

In graph theory, the indegree of a node (vertex) in a directed graph is the number of edges directed into the node. In this model, influence is represented by a normalized value of an agent's indegree.

Tweet:

We simulate the Twitter's content sharing functions for tweeting, retweeting, replies, and mentions ($Tweet_{value} = type \times similarity_{ij} \times influence_i$). A tweet's value is a function of tweet type (0: normal, 1: disinformation, -1: anti-disinformation), similarity, and influence [18].

Limited Attention and Changing Beliefs:

In our model, agents are influenced only by the content they see. Limited attention [139] constrains how many tweets a user reads each active period.

We measure the effectiveness of disinformation operations by the change in belief using a continuous value between zero and one. An agent with a belief value less than 0.5 believes disinformation, and a value greater than 0.5 does not believe disinformation. Belief is calculated as $belief_t = belief_{t-1} + (mean(tweets_{read}) + global_{perc})) \times (1 - belief_{t-1})$.

E.3.2 Model Changes Introduced for `twitter_sim2.0`

Discrete Event Simulation:

As mentioned, a user is influenced only by seen content. A key component to the content a user sees is their activity level or how often they engage on Twitter's platform. The original `twitter_sim` model simulates user activity ranging from every two months up to eighteen times per day. Based on recent research from Pew's Internet & Technology center, we adjusted the distribution of activity in our model [108]. In the simulation, agents are assigned a daily, weekly, or less often activity attribute in the appropriate percentages. Each time an agent wakes, their next wake time is stored by randomly sampling awake time from 0-24 hours daily, 25-168 hours weekly, and 169-1081 hours for less than weekly.

Model Agents and their Twitter Actions:

In `twitter_sim2.0`, normal users, spreaders, and beacons tweet, send mentions, and retweet. Beacons use replies to counter disinformation. All agents retweet with a given probability a portion of their read tweets. Normal users do not create disinformation but do spread it through retweeting. Beacons will not retweet messages that contain disinformation, but spreaders will.

In `twitter_sim2.0`, agents generate mentions with a given probability instead of every time step as in `twitter_sim`. If an agent-i mentions agent-j, and an edge does not exist between the two agents, agent-i will be added to a recommendation queue for agent-j to form an edge between the two. BEND network maneuvers can exploit the mechanics of mentions [18].

Beacons counter disinformation by replying with anti-disinformation but can only counter disinformation they see. Additionally, beacons conduct BEND maneuvers to counter disinformation. We set the percentage of beacons in the simulation as an independent variable ranging from 0-15$\%$.

Spreaders send messages with a mix of noise and disinformation. Spreaders start the simulation with one link and gradually build connections and work their way into the network. Spreaders use the *back* BEND maneuver and link to influential agents aligned with their disinformation campaign.

Impacts of Emotion and Logic:

To facilitate impacts of emotion and logic in our simulations, we adjust the original tweet value calculation to include a multiplier for emotion and a multiplier for logic ($Tweet_{value} = type \times similarity_{ij} \times influence_i \times emotion \times logic$).

Emotion maneuvers represent both the *excite* and *dismay* BEND maneuvers, where *excite* is discussion that brings joy or happiness, and *dismay* is discussion that brings sadness or anger. A very emotional tweet can be either very exciting or very dismaying. In our model, both the tweet

Table E.2: Experimental Design for twitter sim2.0

Variable	Number Variants	Variant Values
Spreaders (% of Network)	2	5, 10
Beacons (% of Network)	3	0, 10, 15
Type of Network	1	normal
Spreader BEND	1	BACK
Beacon BEND (Network Maneuver)	3	BACK, BUILD, NEUTRALIZE
Beacon BEND (Information Maneuver)	3	EMOTION, LOGIC, COMBINED

and the user have an emotional state. The closer the tweet's and user's emotional state, the higher the emotional strength and impact on the user's actions. Users reinforce their beliefs based on the matched sentiment.

The logic maneuvers also represent two similar and opposite BEND maneuvers. *Explain* are actions that provide more details on a topic. *Dismiss* are actions that provide details on the unimportance of a topic. We model this by assigning tweets a value for how logically persuasive it is. This can range from innocuous cat videos to a scientific paper. Likewise, users have a reasoning level, which is a measure of how open they are to new ideas or their ability to listen and comprehend the logic of a tweet. Our model compares the two values, and the resulting logical strength of the tweet is a function of combining the two values. Therefore, a highly logical tweet sent to a user with a high reasoning level is more affected by the tweet, resulting in a high logical strength for the tweet value.

E.4 Experiments

This section explores emergent behavior when beacons use BEND maneuvers to counter a spreader disinformation campaign. See Table E.2 for a detailed view of the independent and control variables for the experiments. The percentages of spreaders and beacons were selected to ensure three ratios: more spreaders than beacons, parity, and more beacons than spreaders. As a baseline, spreaders conduct disinformation maneuvers (*back*), and beacons do not conduct anti-disinformation maneuvers.

E.4.1 Network Based Maneuvers

We selected *back*, *build*, and *neutralize* to provide a diversified representation of network-based BEND maneuvers. These maneuvers affect who is talking and listening to whom in Twitter [35]. Each combination of variables was run 100 times on a 100 node network, and we averaged the results for reporting. Each run simulated 1,680 hours of activity or approximately 2.5 months.

Maneuvers

A *back* maneuver increases the importance of opinion leaders [35]. On Twitter, *back* is conducted through following and retweeting. The greater the scale, the greater the impact. Beacons execute *back* by finding opinion leaders with beliefs between 0 and 0.49 (representing unbelief in disinformation) and following them.

The *build* maneuver involves creating a group, or the appearance of a group [35]. This maneuver focuses on building a group around a common bond and then injecting a directed narrative into the group. Beacon agents conduct *build* by co-mentioning other agents in the network. When a beacon co-mentions (*Agent X* and *Agent Y*), the beacon appears in both of the agents' suggested follow queues, and *Agent X* will appear in *Agent Y*'s suggested follow queue and vice versa.

A *neutralize* maneuver involves actions that limit the effectiveness of opinion leaders [35]. Reducing the number of users that follow an opinion leader achieves a *neutralize* effect. *Neutralize* is similar to *back* but with the opposite desired effect. Beacon agents conduct *neutralize* by finding opinion leaders that propagate disinformation and then follow them. The beacon will send anti-disinformation replies by linking to the opinion leader, possibly directing their followers away from disinformation.

E.4.2 Information Based Maneuvers

We represented *excite*, *dismay*, *explain*, and *dismiss* maneuvers by simulating differences in the emotional and logical aspects of a tweet in conjunction with varying user emotional states and reasoning levels. The emotion simulations represent *excite* and *dismay*, and the logic simulations represent *explain* and *dismiss*. The combined experiment combines both simulations. Each experiment executed 100 runs, resulting in 2400 replications.

Maneuvers

For simulating emotional behavior, the beacons send anti-disinformation tweets and retweets at a minimum of 80% emotional strength. This simulates tweets that attempt to persuade by appealing to the users' emotions to either *excite* or *dismay*, understanding that not all tweets can map to every person's emotional state at 100%. When simulating emotional behavior independent of reason, the logic of the tweet and user reasoning levels result in a 50% average logic strength.

As logical behavior is a function of the logic of the tweet and the user's reasoning level, beacons attempt to deter the spread of disinformation by creating very logical tweets in favor of anti-disinformation. To simulate the logic maneuvers, we set all tweets sent by beacons to the maximum logic level. The beacons simulate *explain* when sending very logical tweets in favor of anti-disinformation or *dismiss* when sending against disinformation. This value is combined with the user reasoning level to determine the logic strength. When simulating logic independent of emotion, we maintain an emotional strength of 50%. In the combined emotion and logic maneuver experiment, we execute both methods simultaneously. Beacons send tweets at 80% emotional strength and tweets that are 100% logical.

E.5 Results and Discussion

Table E.3: BEND Maneuver Results

Network Maneuver				Information Maneuver			
		5% Spreaders	10% Spreaders			5% Spreaders	10% Spreaders
BEND Maneuver	% of Beacons	Change in Network Belief	Change in Network Belief	BEND Maneuver	% of Beacons	Change in Network Belief	Change in Network Belief
None	0	0.00257	0.00477	None	0	0.00257	0.00477
Back	10	-0.00112	0.00079	Emotion	10	-0.00142	-0.00083
	15	-0.00220	-0.00056		15	-0.00243	-0.00167
Build	10	**-0.00780**	**-0.00665**	Logic	10	-0.00182	-0.00084
	15	**-0.00913**	**-0.00969**		15	-0.00270	-0.00202
Neutralize	10	-0.00321	-0.00227	Combined	10	**-0.00294**	**-0.00190**
	15	-0.00472	-0.00376		15	**-0.00392**	**-0.00309**

E.5.1 Baseline

From the results in Table E.3, we see that when there are no beacons in the network, the overall belief grows in favor of disinformation (0.003). When we double the number of spreaders, we find that the overall belief almost doubles in favor of disinformation (0.005). This result indicates that an increased mass of accounts conducting *back* increases the effectiveness of the maneuver.

E.5.2 Network Maneuver

Maneuvers

We see from the results displayed in Table E.3 that the *back* is the least effective BEND maneuver when attempting to counter a *back* disinformation maneuver. Anti-disinformation forces employing *back* to counter *back* must have a greater ratio of agents to succeed because, at parity and under-match, disinformation forces will achieve their desired effects. Using the *build* BEND maneuver to counter a *back* disinformation campaign was the most effective countermeasure across all percentages of spreaders and beacons. These results may be evidence of the importance of creating or reinforcing strong, knowledgeable, and resilient communities in combating disinformation. Further, *neutralize* is an effective countermeasure against the *back* disinformation maneuver for all percentages of beacons and spreaders. However, it is not as effective as the *build* BEND maneuver. In the presence of 5% spreaders, *neutralize* is less than half as effective, and as spreaders increase, its effectiveness decreases compared to *build*.

Sensitivity to Network Size:

We examined how network size impacted the most effective maneuver, *build*. We keep the percentage of spreaders at 10% of the network. We find that *back* remains an effective countermeasure against the *back* disinformation maneuver. Further studies should explore whether the maneuver's degree of effectiveness decreases as network size increases.

Overall

We discover that how the network belief is changed varies by network BEND maneuver. *build* and *neutralize* have a more significant impact on changing the belief of users who believe disinformation. Whereas, *back* has more of an effect on users who do not believe disinformation.

The results for *neutralize* compared to *build* may further strengthen the importance of building or reinforcing strong, resilient communities as the best policy for fighting disinformation. In some cases, the use of *neutralize* could have the opposite desired effect on disinformation. An example of this could be the disinformation opinion leader's Twitter account is suspended (neutralized), strengthening the opinion leader's position as evidence for whatever narrative they are pushing.

E.5.3 Information Maneuver

Maneuvers

When we independently simulated the emotional values for *excite* and *dismay* to beacon-sent tweets, the tweet's belief changed in favor of anti-disinformation by -0.00083 and 0.-0.00167 for beacons at 10% and 15%, respectively.

For *explain* and *dismiss*, when we independently simulated the logic maneuvers without beacons, the total network belief grows linearly to 0.00293. At 10% beacons, the total network belief grows slightly until halfway through the period before it declines in favor of anti-disinformation, ending at -0.00084. Finally, at 15% beacons, the total network belief declines more immediately, ending at -0.000202.

The combined emotion and logic maneuvers simulated the effects of both maneuvers acting simultaneously. The combined maneuver caused an immediate and consistent shift in belief towards anti-disinformation for all beacon levels in the network. Table E.3 displays the results for these maneuvers.

Overall

Beacons executing emotion and logic maneuvers changed the direction of the overall total network belief. They performed in the same general manner given varying beacon and spreader values. If the number of spreaders were greater than the number of beacons, then the belief value would grow over time in favor of disinformation. In the opposite case, the belief value would decrease over time in favor of anti-disinformation. The final values varied depending on the beacon to spreader ratio.

Compared to the baseline at 10% for beacons and spreaders, the logic maneuver results in approximately the same belief value as the emotion maneuver. They differ because the emotional information effect is immediate but has a gradual decrease, where the logical reasoning effect is delayed with a steeper decrease. This trend suggests that logical reasoning is less effective than emotional information in the short run, but it accumulates a greater effect over time. Combining both the emotion and logic maneuvers proved that there are compounding effects on the total network belief by manipulating both aspects of the message. Table E.3 shows the average total change in network belief for the different information maneuvers experiments.

E.6 Validation

This model is a simulation of personal beliefs, which are difficult to measure, making it difficult to validate the model. Much of the validation for our model is carried forward validation assumptions from the `twitter_sim`[18]. The simulation has face validity as it represents a Twitter environment with agents who behave like Twitter users. Additional validation relies on statistical and stylized facts using collected tweet artifacts such as the daily volume of tweets and the distribution of tweets, retweets, replies, mentions, and quotes. User behavior is based on Pew Research to accurately define the percentage of users who check Twitter daily, weekly, and greater weekly. Our analysis of 2.2 million tweets from COVID-19 caused us to adjust the model to reflect that about 8% of tweets contain mentions. Though the model cannot accurately reflect true emotional strength or user reason, the model does apply previous research to simulate user actions in response to emotional and logical tweets.

E.7 Conclusion

Our results demonstrate that using anti-disinformation BEND maneuvers as countermeasures against BEND disinformation maneuvers may be viable for fighting disinformation campaigns. Our results suggest that building and strengthening communities with truth and trust is a far more effective strategy than offensive actions against disinformation opinion leaders.

We demonstrated how to use emotion and logic within a social network as countermeasures. Beacons play an essential role in sending tweets to negate the direction of the total network belief caused by the *backing* maneuver. We found the ratio between beacons and spreaders and beacon tweet characteristics as two critical factors in reducing disinformation diffusion. When the counter maneuvers were combined, the effects on the overall belief were more pronounced, showing the advantage of a well-devised narrative.

There are several limitations. First, we limit our measure of effectiveness to change in belief. Second, we assume that *excite* and *dismay* as well as *explain* and *dismiss* are complementary to each other. In reality, user narratives may be more complex. Further research is necessary to refine the model to reflect human emotion and logic activity better. Finally, because of the complexity of the network, the robustness of the tweet attributes, and the number of agents and agent rules, we were limited by computing power to simulating networks much smaller than in the real world.

Future work should incorporate edge lists of actual tweet networks to seed the network rather than synthetic scale-free networks allowing the results to be directly compared with real-world events. Future studies must include a complete list of BEND maneuvers and maneuver combinations to simulate influence campaigns better. Finally, incorporating more comprehensive tweet attributes to differentiate between types and intensity of emotions and logic will improve information maneuver simulation leading to better maneuver evaluations.