Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

2017-12

Understanding violence through social media

Frost, Harold G.; Evans, Anthony W.; Hodges, Robert H., Jr.

Monterey, California: Naval Postgraduate School

http://hdl.handle.net/10945/56920

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

UNDERSTANDING VIOLENCE THROUGH SOCIAL MEDIA

by

Harold G. Frost
Anthony W. Evans
Robert H. Hodges Jr.

December 2017

Thesis Advisor:	T. Camber Warren
Second Reader:	Robert Burks

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** December 2017	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE UNDERSTANDING VIOLENCE THROUGH SOCIAL MEDIA			**5. FUNDING NUMBERS**
6. AUTHOR(S) Harold G. Frost, Anthony W. Evans, and Robert H. Hodges Jr			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB number ____N/A____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			**12b. DISTRIBUTION CODE**
13. ABSTRACT (maximum 200 words) While social media analysis has been widely utilized to predict various market and political trends, its utilization to improve geospatial conflict prediction in contested environments remains understudied. To determine the feasibility of social media utilization in conflict prediction, we compared historical conflict data and social media metadata, utilizing over 829,537 geo-referenced messages sent through the Twitter network within Iraq from August 2013 to July 2014. From our research, we conclude that social media metadata has a positive impact on conflict prediction when compared with historical conflict data. Additionally, we find that utilizing the most extreme negative terminology from a locally derived social media lexicon provided the most significant predictive accuracy for determining areas that would experience subsequent violence. We suggest future research projects center on improving the conflict prediction capability of social media data and include social media analysis in operational assessments.			
14. SUBJECT TERMS social media, sentiment, Twitter, ISIS, Iraq, regression analysis, conflict prediction, defense			**15. NUMBER OF PAGES** 65
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

THIS PAGE INTENTIONALLY LEFT BLANK

UNDERSTANDING VIOLENCE THROUGH SOCIAL MEDIA

Harold G. Frost
Major, United States Army
B.S., Middle Tennessee State University, 2006

Anthony W. Evans
Major, United States Army
B.A., University of California-Davis, 2006

Robert H. Hodges Jr.
Lieutenant Commander, United States Navy
B.A., University of Montana, 2002

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN DEFENSE ANALYSIS

from the

NAVAL POSTGRADUATE SCHOOL
December 2017

Approved by: T. Camber Warren, Ph.D.
Thesis Advisor

Robert Burks, Ph.D.
Second Reader

John Arquilla, Ph.D.
Chair, Department of Defense Analysis

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

While social media analysis has been widely utilized to predict various market and political trends, its utilization to improve geospatial conflict prediction in contested environments remains understudied. To determine the feasibility of social media utilization in conflict prediction, we compared historical conflict data and social media metadata, utilizing over 829,537 geo-referenced messages sent through the Twitter network within Iraq from August 2013 to July 2014. From our research, we conclude that social media metadata has a positive impact on conflict prediction when compared with historical conflict data. Additionally, we find that utilizing the most extreme negative terminology from a locally derived social media lexicon provided the most significant predictive accuracy for determining areas that would experience subsequent violence. We suggest future research projects center on improving the conflict prediction capability of social media data and include social media analysis in operational assessments.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

THIS PAGE INTENTIONALLY LEFT BLANK

I.   INTRODUCTION

Social media continues to evolve as a means of sentiment sharing, communication, and social interaction. The interconnectedness of population groups continues to grow with the advancement of mobile device technology and the accessibility of Internet communication infrastructures.[1] This medium provides a constant flow of social tendencies throughout 37 percent of the world's population, which are not limited to benign subjects but also include radical leanings and other forms of social unrest.[2] Analyzing social media trends in key locations of strategic concern could provide an additional tool for conflict prediction.[3] Additionally, examining the relationship between social media sentiments and violent events could allow decision makers to be proactive and less reactionary.

Why is this important? Human domain dynamics constantly shift, so the requirement to seek new and inventive ways to gather intelligence on those shifts is imperative. Moreover, the environments where human intelligence is needed most are often the hardest to access physically. The lack of real-time human intelligence in locations with limited access is therefore a serious problem. However, advances in communication technology have also produced new means for maintaining situational awareness. This research seeks to provide insight into human sentiment through social media analysis as a viable solution to this problem. Once we understand the relationship between sentiment and violence in a particular conflict, we can operationalize solutions to curb the occurrence of violence through conflict resolution.

---

[1] National Intelligence Council, *Global Trends 2030: Alternative Worlds* (Washington, DC: Office of the Director of National Intelligence, December 2012).

[2] "Acting Social Network Penetration in Selected Countries as of January 2017," Statista, 2017, https://www.statista.com/statistics/269615/social-network-penetration-by-region/.

[3] T. Camber Warren, "Mapping the Rhetoric of Violence: Political Conflict Discourse and the Emergence of Identity Radicalization in Nigerian Social Media" (presentation, Annual Meeting of the American Political Science Association, San Francisco, CA, Sept. 3–6, 2015), 6.

Previous thesis research at the Naval Postgraduate School (NPS) revealed a correlation between polling data and Twitter semantic analysis within Yemen.[4] This research discovered that "social media data, when combined with polling, has a positive impact on analysis" and could be a "reliable source of stand-alone data for evaluating popular support under certain conditions."[5] To build on this research area and operationalize this data analysis method, further research should specifically focus on whether semantic trends occurred before significant violent events. Additionally, this research breaks new ground by seeking to predict violence in an active conflict zone, through the use of locally-derived lexicons. We propose that the inclusion of locally-derived lexicons will likely provide a more accurate method for future research, especially in areas where Arabic is the dominant language.

Thus, this study focuses on answering the research question, *How can social media analysis be used as a tool for predicting violent events and understanding social sentiment during violent events?* Through the analysis of social media metadata and historical databases of violent events, this paper examines the relationship between population sentiments and violent events in Iraq from August 1, 2013, through July 31, 2014. We hypothesize that shifts in social media content tend to occur before violent acts take place. If a predictive relationship is found through analysis between social media metadata and historical databases of violent acts in Iraq, the study could provide political and military leaders a new method of conflict prediction and therefore increase situational awareness and response time at the strategic, operational, and tactical levels. Accurate knowledge of an operational environment is an intelligence requirement for every military commander. If this knowledge can be used to predict friendly and enemy actions, this can enable forces to posture more effectively in preparation for those actions, and allow commanders to respond more appropriately.

---

[4] Andrew K. Bourret, Joshua D. Wines, and Jason M. Mendes, "Assessing Sentiment in Conflict Zones through Social Media" (master's thesis, Naval Postgraduate School, 2016), http://calhoun.nps.edu/handle/10945/51650.

[5] Ibid.

II.   LITERATURE REVIEW

A.   SOCIAL MEDIA ANALYSIS AND MILITARY DOCTRINE

Social media is increasingly spreading to every part of the world and, by 2015, accounted for 31 percent of smartphone data consumption.[6] Twitter, Facebook, Instagram and other Internet based companies are experiencing record growth outside of Western countries. The attractiveness of this capability is due to many things, but surely due to the freedom of voicing an opinion in relative safety and anonymity. Despite locations where government control extends into the social media realm, personal and group leanings through this medium have become relatively common throughout the world.

The commonality in social expression and the desire to express personal leanings and be connected to a larger social group through the cellular medium continue to expand. For example, both Twitter and Instagram are in the top 20 most-visited websites in the world according to the Alexa ranking.[7] Twitter currently has 313 million monthly users, and over 75 percent of those users exist outside the continental United States.[8] These broad based services provide a significant cross section of the "wired" planet and can be analyzed for the purposes of gathering a greater understanding of human interest in a geographical space.[9]

The predictive capacity of social media has been researched for some time now. Many researchers have been able to show that trends can be used to predict real-world

---

[6] Ian Tunnicliffe and Steve Tatham, *Social Media—The Vital Ground: Can We Hold It?* (Carlisle, PA: Strategic Studies Institute and U.S. Army War College Press, 2017), 8.

[7] "The Top 500 Sites on the Web," Alexa Internet, Inc., accessed February 14, 2017, http://www.alexa.com/topsites.

[8] "#numbers," Twitter, accessed February 14, 2017, http://about.twitter.com.

[9] Sheng Yu and Subhash Kak, *A Survey of Prediction Using Social Media* (Stillwater, OK: Department of Computer Science, Oklahoma State University, 2012), 1.

outcomes.[10] These studies were primarily conducted in Western/First World markets with predictive analysis focused on market trends or election results. There has been relatively little work centered in the third world populations, who experienced the highest rates of smartphone and Internet growth in recent years, and even less with a focus on conflict prediction.[11] However, there has been some research on this very topic, and it is on this research we wish to build. A 2016 Naval Postgraduate thesis titled "Assessing Sentiment In Conflict Zones Through Social Media" was published comparing Yemeni polling data to social media trends.[12] The authors compared "geographically anchored polling and social media data, utilizing over 1.4 million geo-referenced messages sent through the Twitter network from Yemen over the period from October 2013 to January 2014."[13] Once compiled and compared, the data was used to determine popular support for extremist groups and support for the Yemeni government and concluded that social media data can be a used for evaluating popular support under certain conditions.[14]

There are many different ways to measure local sentiment ranging from keyword spotting to lexical analysis.[15] In addition to these, the U.S. military has its own methods of analysis to provide intelligence to decision makers. Within the U.S. Army, the primary role of intelligence is to provide, "Commanders and decision makers with the requisite information facilitating their situational understanding so that they may successfully

---

[10] Sitaram Asur and Bernardo A. Huberman, "Predicting the Future with Social Media," in *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* 1 (August–September 2010): 492–499, http://ieeexplore.ieee.org/document/5616710/; Brendan O'Connor et al., "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," in *Proceedings of the International AAAI Conference on Weblogs and Social Media* 11, no. 122–129 (Washington, DC: 2010), 1–2; Spyros E. Polykalas, George N. Prezerakos, and Agisilaos Konidaris, "An Algorithm Based on Google Trends' Data for Future Prediction. Case study: German Elections," in *2013 IEEE International Symposium on Signal Processing and Information Technology* (December 2013): 69–73, http://ieeexplore.ieee.org/document/6781856/.

[11] Tunnicliffe and Tatham, *Social Media—The Vital Ground,* 8.

[12] Bourret, Wines, and Mendes, "Assessing Sentiment in Conflict Zones through Social Media."

[13] Ibid.

[14] Ibid.

[15] Erik Cambria et al., "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intelligent Systems* 28, no. 2 (March/April 2013): 15–21, http://222.252.30.203:8888/bitstream/123456789/13029/1/Intelligent%20systems%20and%20their%20applications.Vol.28.Iss.2.A.6.pdf.

accomplish their missions in full spectrum operations."[16] Social Network Analysis is one of the primary methods used by the military and correlates the number and strength of links between persons of interest and threat groups.[17] Although this method can produce results at the operational and tactical level regarding targets of interest, the benefit of open source sentiment gathering through social media is often overlooked which could assist commanders in understanding their operational environment.[18]

When it comes to the U.S. military, there is always an urgent and ever-present need to understand the operational environment, but some commanders and military leaders are skeptical of social media analysis.[19] In a 2011 White Paper titled "Strategic Landpower: Winning the Clash of Wills," written by the commanders of the U.S. Army, Marine Corps, and Special Operations Command, the commanders recognized that the United States has engaged in conflicts without understanding the physical, cultural, and social factors of the operational environment.[20] Their proposed solution to better understand these factors is through strategically deployed forces that can interact with foreign governments, militaries, and populations because they claim the only reliable means of assessing how people will act is through human to human contact.[21] However, in areas denied to U.S. forces and government agencies, the U.S. generally relies on human intelligence and other technical means to provide understanding of the "human domain."

While academia continues to build the link between social media and local sentiment, mostly for marketing and opinion polling, these methods have only begun to be extended to developing countries, much less today's conflict-ridden operational

[16] Department of the Army, *Intelligence,* FM 2-0 (Washington, DC: Department of the Army, May 2004), 1–1.

[17] John Scott, *Social Network Analysis* (Thousand Oaks, CA: Sage, 2012).

[18] Department of the Army, *Intelligence,* 1-1.

[19] Tunnicliffe and Tatham, *Social Media—The Vital Ground,* viii.

[20] James F. Amos, William H. McRaven, and Raymond Odierno, "Strategic Landpower: Winning the Clash of Wills" (white paper, United States Army, United States Marine Corps and the United States Special Operations Command, 2013), 8.

[21] Ibid.

environments.[22] Although the majority of military intelligence tasks focus on current or potential threats, methods of understanding the leanings of an entire population within an operational environment could alert commanders to potential events or activities that would change the nature of operations and the initiation of prompt and appropriate action.[23] The knowledge available through social media metadata analysis directly supports the U.S. Army Intelligence Mission Essential Task of Support to Strategic Responsiveness, which "centers on providing information and intelligence to the commander, which facilitates his achieving understanding of the enemy and the environment."[24] While multiple sources of information gathering are required to paint the most accurate picture, failing to analyze the leanings of a population at large limits the understanding of threat groups who often share the same cultural and religious experiences as the "disaffected portions of the populations where partnerships are most critical."[25] In addition to understanding a particular threat group, social media analysis could provide better insights into the demographic dimension of Military Intelligence which includes the "cultural, religious, and ethnic makeup of a given region, nation, or

---

[22] Apoorv Agarwal et al., "Sentiment Analysis of Twitter Data," *in Proceedings of the Workshop on Languages in Social Media*, eds. Meena Nagarajan and Michael Gamon (Association for Computational Linguistics, 2011), 30–38; Younggue Bae and Hongchul Lee, "Sentiment Analysis of Twitter Audiences: Measuring the Positive or Negative Influence of Popular Twitterers," *Journal of the American Society for Information Science and Technology* 63, no. 12 (2012): 2521–2535; Johan Bollen, Alberto Pepe, and Huina Mao, "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, eds. N. Nicolov and J. G. Shanahan (Barcelona, Spain: AAAI Press, July 2011), 1–10; M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter Brand Sentiment Analysis: A Hybrid System Using N-Gram Analysis and Dynamic Artificial Neural Network," *Expert Systems with Applications* 40, no. 16 (2013): 6266–6282; Long Jiang et al., "Target-Dependent Twitter Sentiment Classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* 1 (Stroudsburg, PA: Association for Computational Linguistics, 2011), 151–160; Lewis Mitchell et al., "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place," *PLOS ONE* 8, no. 5 (2013): e64417, https://doi.org/10.1371/journal.pone.0064417; O'Connor et al., "From Tweets to Polls"; Bourret, Wines, and Mendes, "Assessing Sentiment in Conflict Zones through Social Media"; Warren, "Mapping the Rhetoric of Violence"; Meena Nagarajan, Amit Sheth, and Selvam Velmurugan, "Citizen Sensor Data Mining, Social Media Analytics and Development Centric Web Applications," in *Proceedings of the 20th International Conference Companion on World Wide Web* (ACM, 2011); A. Feller et al., "Divided They Tweet: The Network Structure of Political Microbloggers and Discussion Topics," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media,* ed. N. Nicolov and J. G. Shanahan (Menlo Park, CA: AAAI Press, 2011).

[23] Department of the Army, *Intelligence,* 1-6.

[24] Ibid., 1-3.

[25] Bourret, Wines, and Mendes, "Assessing Sentiment in Conflict Zones through Social Media," 2.

non-state actor."[26] Population trends expressed through social media could not only illuminate threat group propensities but also provide commanders with information on individuals or groups that are sympathetic to U.S. objectives. Where coalitions are critical, understanding who might be a potential ally and who might be persuadable increases the potential for organic solutions to a particular problem set.

B. CONFLICT PREDICTION

Just as social media analysis continues to evolve, conflict prediction remains a highly researched topic. Resent research has sought to combine the behavioral and computer sciences to more effectively predict human behavior.[27] Applied behavior analysis (ABA) has been used to suppress malicious behavior by reinforcing appropriate behavior. This method establishes a baseline during clinical observations of antecedent response and seeks to predict and change those responses through consequences at the individual level.[28] In 2012, the ABA method was expanded to a global method of behavior prediction known as automated behavior analysis, which used computer technology to establish a baseline of worldwide events and how they affected the behavior of larger groups and even the collective behavior of a country.[29] To use this method effectively in expanded environments requires a six step process where past events which demonstrate the behavior you need to predict and the subdivision of antecedent and consequence sections are inputted into a database with numerical values.[30] The values for both antecedent and consequence columns then determine what behavior might occur based on previous events in a particular location. This method along with many other predictive statistical modeling methods relies heavily on choosing the correct antecedents to predict a particular response. Although the inclusion of automated assists aided prediction in a wide variety of threat domains, the process is still

---

[26] Department of the Army, *Intelligence,* 1-25.

[27] Gary Jackson, *Predicting Malicious Behavior: Tools and Techniques for Ensuring Global Security,* 1st ed. (Indianapolis, IN: John Wiley and Sons, 2012), xxix.

[28] Ibid.

[29] Ibid., xxxi.

[30] Ibid., 480.

time-consuming and fails to use the sentiment shared through open source networks such as social media.[31]

Spatial-temporal models have also proved effective in the statistical prediction of conflict. By utilizing in-sample conflict data to establish trends over time and space, researchers observed spatial units until a certain time and used that information to predict subsequent conflict.[32] During a 1992–1995 case study on Bosnia using spatial-temporal modeling, research showed that conflict displays a strong spatial and temporal dependence which improved predictive performances across examined scenarios.[33] This research method introduced mathematical simulations in addition to observational data.[34] By using historical geo-referenced conflict data and the inclusion of information about neighboring communities, the prediction and understanding of conflict was improved.[35]

While utilizing mathematical modeling for conflict prediction, studies have also shown that statistical significance as a primary goal can be counterproductive. In a 2010 article on predicting civil conflicts, the authors argued, "too much attention has been paid to finding statistically significant relationships, while too little attention has been paid to finding variables that improve our ability to predict civil war."[36] By conducting a side-by-side comparison of two models of civil war, this article demonstrated that statistical significance as the only metric for variable utilization can be misleading by ignoring the predictive attributes of the underlying models.[37] This realization will force variation in model specification to ensure that the findings highlighted in this research are not simply based on statistical significance, but on whether or not they improve predictive accuracy.

---

[31] Ibid., 481.

[32] Nils B. Weidmann and Michael D. Ward, "Predicting Conflict in Space and Time," *Journal of Conflict Resolution* 54, no. 6 (2010): 883–901, http://www.jstor.org/stable/25780759.

[33] Ibid., 899.

[34] Ibid., 884.

[35] Ibid., 885.

[36] Michael D. Ward, Brian D. Greenhill, and Kristin M. Bakke, "The Perils of Policy by P-Value: Predicting Civil Conflicts," *Journal of Peace Research* 47, no. 4 (2010), http://journals.sagepub.com/doi/pdf/10.1177/0022343309356491.

[37] Ibid.

C.   METHODOLOGY SYNTHESIS

As social media analysis continues to provide applicable insight into social leanings and conflict prediction methods become more critical, the synthesis of these predictive research methods is critical to understanding the mechanisms underlying human social and violent interactions.[38] Insight into population sentiment and the horizontal links created by social media could provide a more comprehensive conflict prediction capability.[39] In 2015, similar conflict data and social media research discovered significant correlations between geographic regions with higher cellular penetration and rates of conflict.[40] This research showed that interaction across the cellular medium of communication increased the likelihood of higher rates of violence within African states. Findings presented by Pierskalla and Florian (2013) further showed that "cell phone coverage has a significant and substantive effect on the probability of conflict occurrence" throughout the continent of Africa.[41] These findings contradict a similar study conducted in Iraq from 2004–2009. Shapiro and Weidmann (2015) concluded that increased mobile communications reduced the violent acts committed by insurgent groups both at the district and local level within Iraq.[42] Pierskalla and Florian (2013) explain that the likely reason for this contradiction is due to the context of political violence in African states when compared to the conflict in Iraq during Operation Iraqi Freedom.[43] In Iraq, with the aid of the U.S. military, government forces had a superior technological capability than that of many African forces. Therefore, these articles show that the technological capability of actors in a given conflict can alter whether or not cell phone coverage expansion is advantageous.

---

[38] Warren, "Mapping the Rhetoric of Violence."

[39] T. Camber Warren, "Explosive Connections? Mass Media, Social Media, and the Geography of Collective Violence in African States," *Journal of Peace Research* 52, no. 3 (2015): 297–311.

[40] Ibid., 306.

[41] Jan H. Pierskalla and Florian M. Hollenbach, "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa," *American Political Science Review* 107, no. 2 (2013): 207–24, doi:10.1017/S0003055413000075.

[42] Jacob N. Shapiro and Nils B. Weidmann, "Is the Phone Mightier than the Sword? Cellphones and Insurgent Violence in Iraq," *International Organization* 69, no. 2 (2015): 247–274.

[43] Pierskalla and Hollenbach, "Technology and Collective Action."

These findings must be taken into consideration as this research focuses on a geographical space where cellular penetration alone might not be predictive of a particular threat group's level of violence. Due to this reality, the use of a lexicon to determine sentiment will provide an additional metric to understand violence over and above the effects of infrastructure and development. Along these lines, a 2016 Naval Postgraduate School research paper utilized specific nationalist and separatist lexicons within Nigeria to achieve additional predictive leverage over likely and unlikely locations of conflict.[44] This research found a statistically significant relationship between areas with higher violent discourse through social media and areas with higher rates of violence.[45] However, the metrics used in this paper were based exclusively on machine translation of English language lexicons. To expand on this research, this thesis seeks to determine the relationship between social media sentiment and conflict within Iraq by utilizing a locally-derived Arabic sentiment lexicon drawn from automated analysis of Arabic social media, without the need for explicit translation between English and Arabic.

---

[44] Warren, "Mapping the Rhetoric of Violence," 1–6.

[45] Ibid., 16–17.

III. BACKGROUND—IRAQ

Iraq was selected as the case study for analysis based on its strategic importance, sectarian violence, and use of Twitter to express sentiments. Iraq is strategically located between the region's leading Sunni power, Saudi Arabia, and the dominant Shia power, Iran. Thus, the country's population is largely divided between the two main sects of Islam, consisting of approximately 55–60 percent Shia and 40 percent Sunni.[46] The Sunni-Shia divide dates back to 632 AD, when the death of Prophet Muhammad led to disagreement on the process to determine the succession of leadership for the Muslim faith.[47] This disagreement continues to separate the Sunni and Shia followers of Islam. In Iraq, under Saddam Hussein's Baath party regime, the Sunni minority controlled a monopoly of the government positions.[48] Following the removal of Saddam Hussein and the Baath party in 2003, the ascension of Shia control of the government resulted in the marginalization of Iraqi Sunnis.[49] This transfer of control from the Sunni minority to the Shia majority created conditions for resentment by both segments of the population. Complicating matters further, northern Iraq's population is heavily dominated by ethnic Kurds. The Kurds have long sought an independent Kurdistan, but currently the region is governed by the Kurdistan Regional Government as part of the Iraqi government. Figure 1 illustrates the disposition of Iraq's population by ethnic and sectarian divisions in 2014.

---

[46] Central Intelligence Agency, "Middle East: Iraq," in *The World Fact Book 2017* (Washington, DC: Central Intelligence Agency, 2017), https://www.cia.gov/library/publications/resources/the-world-factbook/geos/print_iz.html.

[47] Christopher M. Blanchard, "Islam: Sunnis and Shiites" (Washington, DC: Congressional Research Service, 2009), 1.

[48] Larry Diamond, "What Went Wrong in Iraq," *Foreign Affairs* (September/October 2004): 34–56, 43.

[49] Fawaz A. Gerges, "ISIS and the Third Wave of Jihadism," *Current History* 113, no. 767 (2014): 339–343, 339.
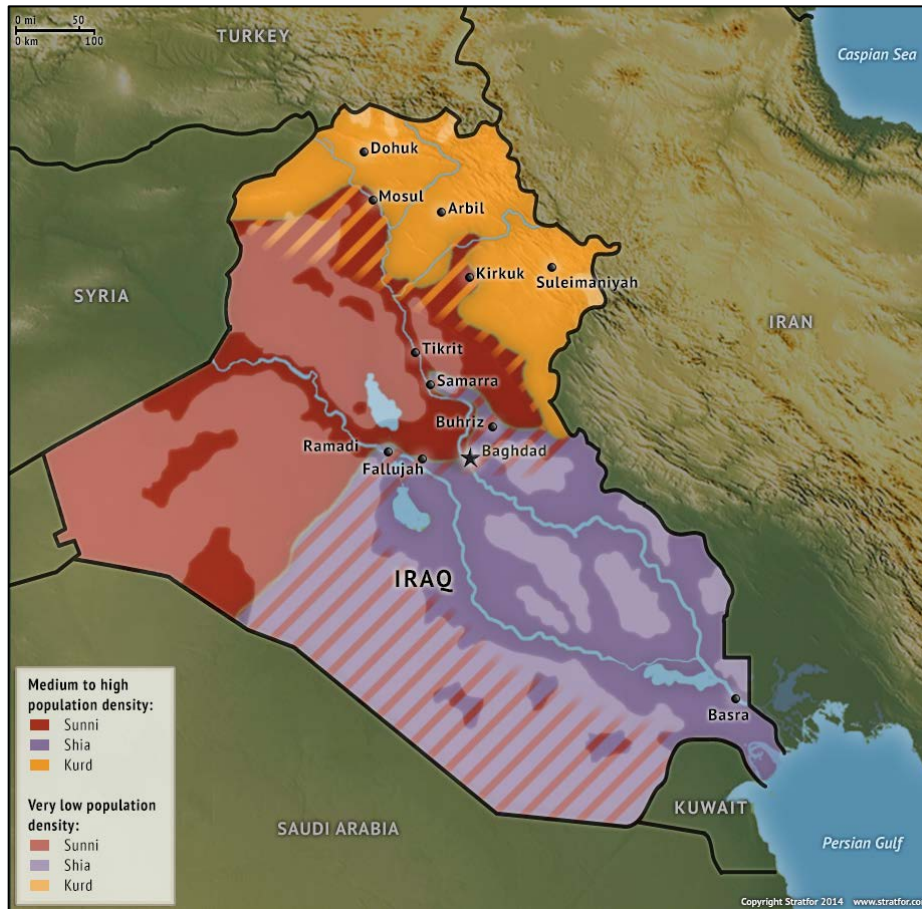
Figure 1. Iraq's Population Density by Ethnic and Sectarian Divisions[50]

In late 2012, a popular Sunni protest movement began to take hold within Iraq.[51] The protests were sparked by the arrest of Sunni aides to Rafia al-Isawi, the Iraqi Finance Minister.[52] These arrests brought to the forefront the issue of increasing harsh treatment of Sunnis under the Shia-led government of the Iraqi Prime Minister, Nouri al-Maliki. The Human Rights Watch 2013 Report cited, "The number of violent civilian deaths in Iraq increased in 2012, for the first time since 2009."[53] This protest movement spread

---

[50] Source: "Iraq's Population Density by Ethnic and Sectarian Divisions," Stratfor, August 10, 2014, accessed September 30, 2017, https://twitter.com/stratfor/status/498544417976770562?lang=en/.

[51] Kirk H. Sowell, "Maliki's Anbar Blunder," *Foreign Policy,* January 15, 2014, https://foreignpolicy.com/2014/01/15/malikis-anbar-blunder/.

[52] Ibid.

[53] "Iraq: A Broken Justice System," Human Rights Watch, January 31, 2013, https://www.hrw.org/news/2013/01/31/iraq-broken-justice-system.

across predominantly Sunni-populated cities in West and Northwest Iraq. These protests were largely divided between pro-government and anti-government demonstrations. The anti-government protests tended to occur in heavily Sunni populated areas, while the pro-government protests tended to occur in heavily Shia populated areas.[54] This divide contributed to the increased ethnic tensions and violence across Iraq.

Opposing political views among the population reflected both negative and positive sentiments across the country. The perceived government directed marginalization and discrimination of Sunnis by Prime Minister Maliki, led to increased negative sentiments and protest by Sunni political and tribal leaders.[55] These clashes escalated in Hawija on April 23, 2013, when Iraqi security forces violently suppressed a protest that resulted in the death of 44 protesters.[56] Furthering the Sunni-Shia divide, Shia militias began to mobilize and increased violent suppression of Sunni populations.[57] The Sunni-Shiite divide created the conditions for the Islamic State of Iraq and Syria (ISIS) to garner support and prosper in Sunni dominated areas. Further exploiting this divide, ISIS conducted vehicle borne improvised explosive device attacks targeting Iraqi security forces and critical infrastructure.[58] ISIS attacks against the government supported ISIS efforts to recruit supporters with negative and hostile sentiments towards the Iraqi government.[59]

The United Nations Assistance Mission in Iraq reported the year 2013 as the most violent year in Iraq since 2008, resulting in 7,818 civilians and 1,050 Iraqi security forces

---

[54] Sam Wyer, "Political Update: Mapping the Iraq Protests," Figure 2, Institute for the Study of War, January 11, 2013, http://www.understandingwar.org/backgrounder/political-update-mapping-iraq-protests.

[55] Zana K. Gulmohamad, "The Rise and Fall of the Islamic State of Iraq and Al-Sham (Levant) ISIS," *Global Security Studies* 5, no. 2 (Spring 2014): 4.

[56] Sowell, "Maliki's Anbar Blunder."

[57] Jessica D. Lewis, *Al-Qaeda in Iraq Resurgent: The Breaking the Walls Campaign, Part 1* (Washington, DC: Institute for the Study of War, September 2013), http://www.understandingwar.org/sites/default/files/AQI-Resurgent-10Sept_0.pdf, 8.

[58] Ibid., 9.

[59] Gerges, "ISIS and the Third Wave of Jihadism," 339.

killed.[60] Additionally, some 17,891 Iraqis were injured as a result of violence during this same period.[61] In December of 2013, ISIS overthrew Iraqi government control of Fallujah and parts of Ramadi. ISIS expansion continued throughout 2014, as the group captured the cities of Mosul and Tikrit in June.[62] Essential to the ISIS strategy was their use of social media to recruit, indoctrinate, and wage psychological warfare to overthrow the Iraqi government's control over vast areas.[63] Around 2011, many Jihadi groups began a shift from posting mostly on Internet forums to mainstream social media platforms like Twitter and Facebook.[64]

In 2014, there were an estimated 46,000-90,000 ISIS support accounts, making Twitter the most active social media platform for ISIS.[65] A study of ISIS Twitter users from September 2014 to December 2014 revealed the majority of ISIS support accounts were geo-referenced in Syria and Iraq with accounts that often posted messages in higher volumes than the average Twitter account.[66] Additionally, most ISIS support accounts had a higher number of followers than the average Twitter account.[67] A study analyzing ISIS networks on Twitter determined the followers of an ISIS sympathizer represented four Twitter user categories: international mass media, regional Arabic mass media, IS

[60] Luke Harding and Agencies in Baghdad, "Iraq Suffers Its Deadliest Year Since 2008," *Guardian*, January 1, 2014, https://www.theguardian.com/world/2014/jan/01/iraq-2013-deadliest-year-since-2008.

[61] Zachary Laub and Jonathan Masters, "The Islamic State in Iraq and Syria," Council on Foreign Relations, August 8, 2014, http://www.aina.org/news/20140808123646.pdf.

[62] Cameron Glenn, "Timeline: Rise and Spread of the Islamic State," Wilson Center, July 5, 2016, https://www.wilsoncenter.org/article/timeline-rise-and-spread-the-islamic-state.

[63] Jytte Klausen, "Tweeting the Jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq," *Studies in Conflict & Terrorism* 38, no. 1 (2015): 1–22, doi:10.1080/1057610X.2014.974948.

[64] Ibid., 3.

[65] Shaun Wright et al., "Resurgent Insurgents: Quantitative Research into Jihadists Who Get Suspended but Return on Twitter," *Journal of Terrorism* 7, no. 2 (May 2016): 1, http://doi.org/10.15664/jtr.1213.

[66] J. M. Berger and Jonathon Morga, *The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter* (analysis paper no. 20, Brookings Project on U.S. Relations with the Islamic World, March 2015), https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf, 3.

[67] Ibid.

fighters, and IS sympathizers.[68] In practice, ISIS social media networks appealed to both media outlets and individuals, which enabled ISIS to rapidly project their narratives.

For example, an ISIS sympathizer used Twitter messages referring to current events for propaganda, radicalization, and recruitment (Figure 2). In the figure, some of the messages include links to videos or "#" symbol to categorize the message for others to users to follow. The themes of the propaganda messages express counter-narratives designed to show the anti-ISIS coalition forces as liars who are incompetent and deliberately target civilians. In contrast, the radicalization messages seek to highlight recent acts of violence conducted by ISIS as a display of strength. The recruitment messages promote narratives that seek to inspire others to take action based on religious reasons. ISIS sympathizer Twitter messages like these are designed to appeal to like-minded individuals and share sentiments regarding current events. Additionally, the Twitter messages include a mixture of sarcasm, inference, and words associated with extreme sentiments. Words such as, "awesome" or "amazing" are often associated with extremely positive sentiments. While words like, "massacred" are often associated with extremely negative sentiments. The mixture of current events, religious ideology, and extreme sentiments make these ISIS sympathizer's Twitter messages more appealing to large audiences.

---

[68] Akemi Takeoka Chatfield, Christopher G. Reddick, and Uuf Brajawidagda, "Tweeting Propaganda, Radicalization and Recruitment: Islamic State Supporters Multi-Sided Twitter Networks," in *Proceedings of the 16th Annual International Conference on Digital Government Research* (ACM, 2015), 247.

Timestamp	@shamiwitness Tweet
	Propaganda
6/10/2014 10:38	Coalition planes massacred these children in airstrikes on #Hit ,#Anbar http://t.co/yCsEgkwDY6#Iraq
6/10/2014 17:48	these PKK fellas are exceptional liars.after the city was almost fully abandoned by civilians, they now claim 55 thousand civilians there.
7/10/2014 16:08	This is so awesome. US airstrikes also by mistake hit a Shia militia convoy near Tuz http://t.co/zUwWcEzBzk 2nd one after the Jabour strikes
8/10/2014 14:30	Ask the Americans how they liked fighting JTJ/AQI/MSC/ISI in street wars in Ramadi,Fallujah etc. (3)
10/10/2014 4:13	Amazing thing about Kobane:not a single genuine claim of massacres of civilians. Not one.Meanwhile, Assad keeps killing kids in Aleppo
	Radicalization
6/10/2014 14:45	@Himalaya888 you do realize IS wants to destroy every single nation-state,Arab or Kurd or communist doesnt matter, that they come across?
7/10/2014 18:59	Real warriors, I mean those who are not fags, dont need air support.
8/10/2014 12:18	#IS blows up police station in downtown Kobane where YPG were holed up. #Syria
12/10/2014 11:18	Notice this recent uptick of IS activity in Qara Teppe? multiple VBIEDs [*vehicle-borne improvised explosive device*] hitting PUK ,Pesh,Asayesh ?Guess where else that happened earlier?

	Recruitment
7/10/2014 18:21	This is the time for Muslim Kurds in Turkey to show whether they can ever counter PKK.
7/10/2014 20:05	@EbuRuana so if it is not right to make dua [*the Islamic act of calling out to Allah*] for a kafir [*a disbeliever, someone who rejects Allah and who does not believe in Muhammad as the final messenger of Allah*], what is the right thing to do ?
11/10/2014 16:29	@banihillal blend in like you are the most anti-jihadist guy one can think of. @isinlibya @ummussamah11
12/10/2014 20:14	@DaniyalYasin93 The question is to you:can you,as a Muslim,counter-argue any of the points made in that segment quoted from Dabiq magazine [the IS' propaganda magazine]?

Figure 2. Example of ISIS Sympathizer Twitter Messages.[69]

[69] Ibid., 246.

IV. RESEARCH METHODS

A. HYPOTHESIS

Building on previous research, we intend to further increase the field's understanding of the relationship between violence and social media.[70] We propose to do this by answering the question, *How can social media analysis be used as a tool for predicting violent events and understanding social sentiment during violent events?* We hypothesize that negative social media sentiment tends to occur before violent acts take place. If a predictive relationship is found between social media metadata and historical databases of violent acts in Iraq, the study could provide political and military leaders a new method of conflict prediction and therefore increase situational awareness and response time at the strategic, operational, and tactical levels. Finally, as a secondary objective, we propose that correlations between social media and sentiment data within Yemen found in previous research can be further confirmed by conducting similar research within a different geographical space.[71]

B. DATA AND METHODS

1. Social Media

The research was initiated by conducting Twitter key-word analysis to determine trends and sentiment within Iraq. NPS has 40 terabytes of licensed Twitter data archived from August 1, 2013, through July 31, 2014, that contains a randomized sample of 10% of the Twitter messages that appeared on Twitter during this period. In order to isolate the origin of the Twitter messages to Iraq, the coordinates used in our analysis are metadata, provided by Twitter as part of the licensed archive, which represent the coordinates of the self-reported hometown from each user profile. These coordinates are less reliable than the GPS coordinates associated with specific messages. However, these GPS coordinates are only available for approximately 1% of the archived data, because users must opt-in

---

[70] Bourret, Wines, and Mendes, "Assessing Sentiment in Conflict Zones through Social Media"; Warren, "Mapping the Rhetoric of Violence."

[71] Bourret, Wines, and Mendes, "Assessing Sentiment in Conflict Zones through Social Media."

to allow them to be collected. In contrast, the profile-based geo-coordinates are available for approximately 30 percent of the records in the NPS archive. Using this method, from August 1, 2013, through July 31, 2014, the we find that the NPS archive has geo-referenced within Iraq an average of nearly 2,300 tweets per day. We believe this is an acceptable rate as a recently completed NPS thesis focused on Yemen using similar volumes of data.[72]

2.   Sentiment Dictionary

Over the past few years, researchers have developed many English sentiment lexicons for web content analysis. However, the development of Arabic-specific sentiment lexicons remains limited. The difficulty of developing Arabic specific sentiment lexicons is caused in part by the fundamental differences in regional dialects (Egyptian, Levantine, Gulf, Maghrebi, Iraqi, Yemenite, etc.) spoken and used by individuals, and Modern Standard Arabic (MSA), the only formally taught written language of the Arab world.[73] Compounding the difficulty is the mixture of dialects to create a new unique and highly-regionalized dialect. For instance, Iraqi Arabic uses both a combination of Gulf and Levantine dialects. Furthermore, while using social media, people often use slang words, words from their regional dialect, or social media specific characters to express their sentiments instead of formal writing standards.[74] Due to these nuanced issues, we selected the Large-Scale Twitter-Specific Arabic Sentiment Lexicon (AraSenTi) for our analysis, because the AraSenti lexicon specifically focuses on the use of informal Arabic language found on Twitter.[75]

The AraSenTi Lexicon is a robust lexicon generated through a hybrid approach, combining Arabic tweet-specific words and a Pointwise Mutual Information (PMI) corpus that measures the strength of word associations with positive, neutral, and

---

[72] Ibid.

[73] Gehad S. Kaseb and Mona F. Ahmed, "Arabic Sentiment Analysis Approaches: An Analytical Survey," *International Journal of Scientific & Engineering Research* 7, no. 10 (October 2016): 712.

[74] Nora Al-Twairesh, Hend Al-Khalifa, and AbdulMalik Al-Salman, "AraSenTi: Large-Scale Twitter-Specific Arabic Sentiment Lexicons," *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Berlin, Germany: August 7–12, 2016), 697.

[75] Ibid.

negative datasets of tweets.[76] The lexicon consists of a total of 225,276 Arabic words with assigned values for individual words varying from -8 to 8. These values are based on whether the word was associated with a positive, neutral, or negative co-occurring text.[77] Words with values closer to -8 are deemed associated with extremely negative sentiments, while words with values closer to 8 are deemed associated with extremely positive sentiments. Words with values closer to zero are deemed associated with neutral or mixed sentiments. Due to the challenges associated with generating Arabic specific lexicons, we believe AraSenTi-PMI was the appropriate lexicon for our research. The Arabic tweet-specific hybrid approach to lexicon generation specifically addresses the nuances of data mining Iraqi tweets, while providing an automated means of quantifying positive and negative sentiments.

Another challenge of social media analysis in general is that the demographics associated with social media use tend to be younger, wealthier and urban.[78] With this acknowledged, our research attempts to show that violence done by anyone, tweeter or not, will be reflected in Twitter sentiment. A link established between Twitter and violence will be a link nonetheless, regardless of demographics, and provide potential for future analysis and prediction. To further reduce the potential for a spurious connection, we have also chosen several control variables associated with demographic and infrastructure development datasets, as will be discussed further below.

3.   Analysis Framework

For this research, our analysis is based on a spatial-temporal frame. To start, all our variables are spatially isolated to the borders of Iraq as derived from the "CShapes" dataset that provides historical maps of state boundaries and capitals in the post-World

---

[76] Ibid., 701.

[77] Ibid.

[78] Stephen Ansolabehere and Eitan Hersh, "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate," *Political Analysis* 20, no. 4 (Autumn 2012): 437–459, http://www.jstor.org/stable/23359641; Pablo Barberá and Gonzalo Rivero, "Understanding the Political Representativeness of Twitter Users," *Social Science Computer Review* 33, no. 6 (December 1, 2015): 712–729; Alan Mislove et al., "Understanding the Demographics of Twitter Users," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (AAAI Press, 2011), 554–557.

War II period.[79] We further restrict our variables by timeframe to match the timespan of our available Twitter data, from August 1st, 2013 to July 31st, 2014. To allow for proper geospatial analysis, we created a by calendar month grid covering Iraq at an interval of two-kilometer squares and then overlaid our variables to allow for a uniform unit of analysis. This method will allow for equivalent aggregation and comparison for both our dependent and independent variables, as well as our control variables.

4. Dependent Variable

As a starting point, the Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP GED) has a collection of geo-referenced data on individual acts of armed violence in conflicts since 1989. This data provides a means to analyze correlation between social media trends and violent events, by recording the date and location of deadly events occurring in the context of armed conflicts.[80] The unit of analysis in the raw dataset is an event, defined as "an incident where armed force was [used] by an organized actor against another organized actor, or against civilians, resulting in at least 1 direct death at a specific location and a specific date."[81] The raw UCDP GED dataset contains 128,264 events collected globally (excluding Syria) from January 1st, 1989 to December 31st, 2015, with the best spatial resolution of a fully geocoded individual village and the best temporal resolution of a day.[82]

For this research, this dataset was first isolated to the borders of Iraq, then stripped of all entries that did not have a location confidence greater than or equal to second level administrative division (for Iraq, the 102 "kaza" districts).[83] These events were then binned into monthly event counts starting with the first month of our Twitter

[79] Nils B. Weidmann, Doreen Kuse, and Kristian Skrede Gleditsch, "The Geography of the International System: The CShapes Dataset," *International Interactions* 36, no. 1 (2010): 86–106.

[80] Ralph Sundberg and Erik Melander, "Introducing the UCDP Georeferenced Event Dataset," *Journal of Peace Research* 50, no.4 (2013): 523–532.

[81] Mihai Croicu and Ralph Sundberg, "UCDP GED Codebook Version 5.0" (Department of Peace and Conflict Research, Uppsala University, 2016).

[82] Sundberg and Melander, "Introducing the UCDP Georeferenced Event Dataset"; Croicu and Sundberg, "UCDP GED Codebook Version 5.0."

[83] Ibid.; "Iraq Districts," Humanitarian Information Centre for Iraq, accessed on October 25, 2017, https://www.globalsecurity.org/military/world/iraq/images/013_A4_Iraq_Districts.jpg.

data, August 2013 through July 2014, and then overlaid on the established two-kilometer grid. This scoping identified 926 geo-referenced violent events inside our spatial-temporal grid, which will be used as our dependent variable, *Violent Events*. Our research hypothesizes a positive relationship between negative sentiments expressed in social media messages and subsequent rates of violent events. While this association appears intuitive, careful quantitative analysis will be required to determine whether it can be confirmed as a base for further research.

Figures 3 presents a visual representation of the UCDP GED geo-referenced observations (A) and their density (B). The raw observations are shown on the left, with each mark indicating a violent event within our analysis timeframe, totaling 926 events over 12 months. The heat map shows the density of these observations with red showing heavy density tapering to yellow showing less dense observations. The highest density of *Violent Events* occurs in and around Baghdad, the nation's capital, located roughly in the center of Iraq. We can also see that the pattern of violence in Iraq appear to correspond to high population centers and major roads.
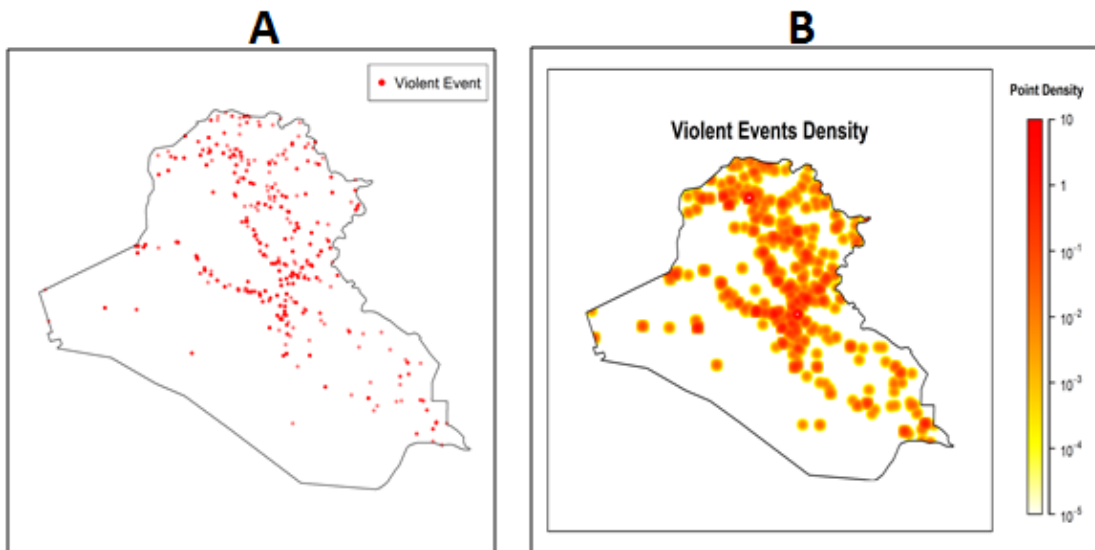

Figure 3. *Violent Events* Geo-referenced (A) and Density of *Violent Events* (B) in Iraq, August 2013–July 2014

21

5. Independent Variable

Our independent variables were derived from the comparison of two datasets: the previously described AraSenTi lexicon, and a historical archive of Twitter messages. The Twitter dataset is comprised of a NPS-licensed, random 10 percent cross-section of global Twitter posts from August 1st, 2013 through July 31st, 2014. The raw Twitter dataset contains over 40 terabytes of public messages and associated metadata. To facilitate the computation of such a large data set, we utilize a highly-parallelized, in-memory database application developed at NPS to speed computation.[84]

The comparison of these two datasets is used to derive our base sentiment data and is used to create the independent variables of *Total Twitter Messages, Negative Twitter Messages* and *Extremely Negative Twitter Messages.* Our research breaks each of these variables into calendar-month counts that are then geo-referenced to our already established two-kilometer grid for model analysis.

Total Twitter Messages is the count of all Twitter messages geo-referenced within Iraq's spatial-temporal area, yielding a raw corpus of Twitter messages totaling 829,537 observations. This variable serves as our baseline independent variable for contrast to the next two more restrictive independent variables. Figure 4 is a visual representation of *Total Twitter Messages* and their corresponding density. The heat map illustrates the density of these observations with red showing heavier density tapering to yellow showing less dense observations. The density and location of *Total Twitter Messages* across Iraq are similar to the *Violent Events* in Figure 3, with the densest locations occurring near high population centers and near roads.

---

[84] Warren, "Mapping the Rhetoric of Violence."
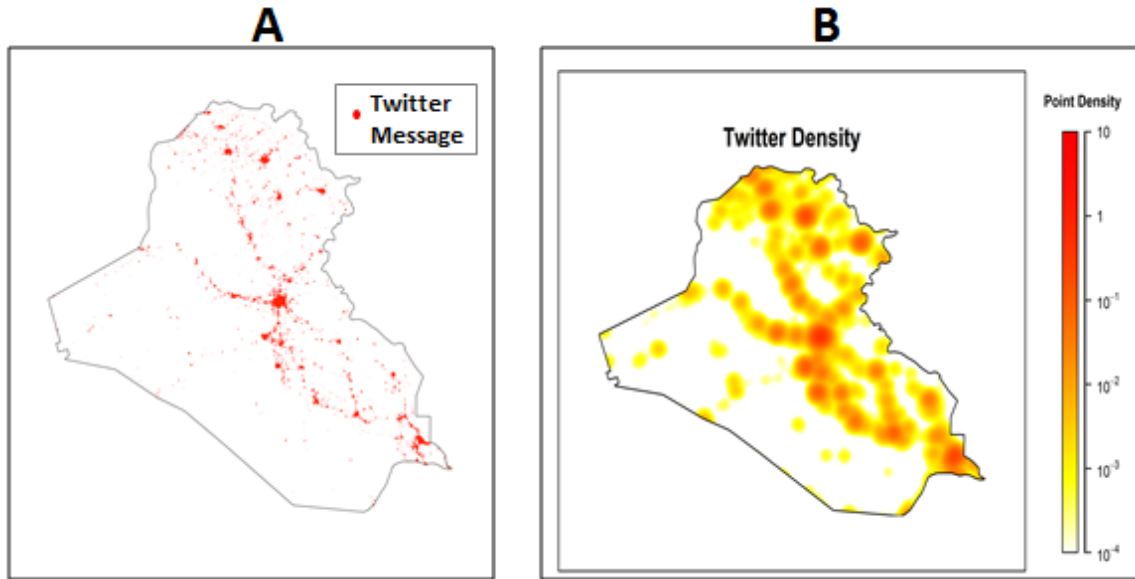
Figure 4. *Total Twitter Messages* Geo-referenced (A) and Density of *Total Twitter Messages* (B) in Iraq, August 2013–July 2014

Negative Twitter Messages isolates the independent variable to counting those messages containing negative sentiment terms, as identified by the AraSenTi lexicon, representing a total of 117,018 observations. *Extremely Negative Messages* further narrows the independent variable to counting messages containing terms in the top 0.1 percent (an AraSenTi score of >5.833 or <-6.153) of the distribution of negative sentiment terms identified in the AraSenTi lexicon, reflecting messages at the highest level of negative sentiment. While yielding a smaller dataset of only 1740 total messages, these tweets contain the most extreme terms identified by the lexicon, and therefore may provide greater levels of confidence in the assessment of spatial patterns of negative sentiments.

6.   Control Variables

Our control variables center around demographic and infrastructure information. These variables were included to ensure that the relationships observed between the dependent and independent variables were not simply a result of economic development, population density or existing infrastructure in a geographical space.

23

a. *Road Distance*

The control variable *Road Distance* is derived from the Global Roads Open Access Data Set (gROADSv1) from the Socioeconomic Data and Applications Center (SEDAC). This provides a well-documented global dataset of roads between population centers which is geospatially referenced. We incorporated the data by measuring the distance from a gridded cell to the nearest recorded road segment.[85] The dataset has been continually updated from the 1980s to 2010 and spatial accuracy varies. The Iraq subset used in our research contains 160,000 kilometers of georeferenced roads, primarily taken from the National Imagery and Map Agency (NIMA) in 2003 at 100k resolution.[86] Figure 5 shows the dataset when laid onto our 2-km grid with green representing the road networks of Iraq for the year 2010, the most current year available for our area.

Figure 5.   Road Density in Iraq, 2010
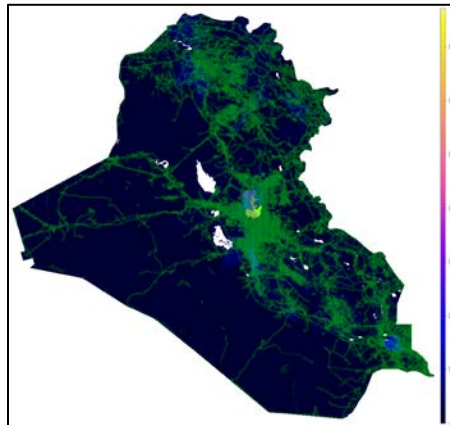
---

[85] Center for International Earth Science Information Network, Columbia University and Information Technology Outreach Services, and University of Georgia, "Global Roads Open Access Data Set, Version 1 (gROADSv1)" (Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2013), http://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1.

[86] Ibid.

b. *Night Light*

The control variable *Night Light* measures nighttime light emissions by utilizing the Version 4 Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS) Nighttime Lights Time Series dataset. These datasets are collected in 30 arc second grids via four sun-synchronous satellites to display global light emissions for the years 1992–2013. The data is collected on the individual light emission basis and then filtered for percent frequency of detection to normalize the variation in cloud-free persistent light observations. Our specific dataset was collected from satellite F18 in 2013, the most recent data available for our region.[87] Light emissions are a hallmark indicator of urbanization and as such will allow for our research to further control for urbanization as a factor for any link found between violence and Twitter sentiment.[88] Figure 6 shows the 2013 level of light emissions with dark to light scaling, yellow/white showing the highest levels of night light emissions down to the dark blue/black area of little to no light emissions.
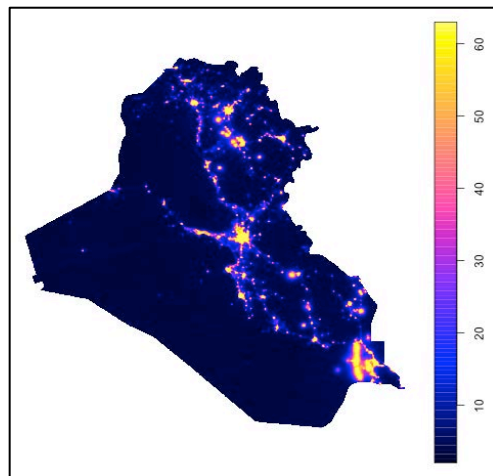
Figure 6.   Light Emissions in Iraq, 2013

---

[87] NOAA National Centers for Environmental Information, "DMSP-OLS Nighttime Lights Time Series," National Geophysical Data Center, Boulder, CO, last modified June 7, 2017, http://noaa.maps.arcgis.com/home/item.html?id=78147366ebac4f28862053b89169dfc9.

[88] Marc L. Imhoff et al., "A Technique for Using Composite DMSP/OLS 'City Lights' Satellite Data to Map Urban Area," *Remote Sensing of Environment* 61, no. 3 (1997): 361–370; Ting Ma et al., "Quantitative Estimation of Urbanization Dynamics Using Time Series of DMSP/OLS Nighttime Light Data: A Comparative Case Study from China's Cities," *Remote Sensing of Environment* 124 (2012): 99–107.

c.   *Population Density*

This control variable models the distribution of human population throughout the globe and was derived from Columbia University's Center of International Earth Science Information Network's (CIESIN) dataset titled Gridded Population of the World (GPWv4). This dataset is collected and updated yearly starting in 1995 with accuracy at approximately 1 km at the equator.[89] Our data from the GPWv4 dataset was pulled from the most current year available for Iraq, 2010. Historically, levels of population density and urban development have been linked to a positive influence on violence.[90] This control variable will allow our research to see if this is a contributing factor to any link that is seen between violence and Twitter sentiment.

d.   *Previous Violent Events and Deaths*

Also derived from the UCDP GED dataset are four control variables designed to account for the recent history of violent events in a particular spatial region. The first two variables, *Violent Events (1 Month Lag)* and *Violent Events (2 Month Lag)* record lagged counts of the total number of violent events in a grid-cell at temporal distances of 1 and 2 months respectively. The second two variables, *Deaths (1 Month Lag)* and *Deaths (2 Month Lag)*, record the sum of the death totals associated with each of the violent events in *Violent Events*. These variables are derived from the best (most likely) estimate of total deaths involved for the event, whether combatant or civilian, again at lags of 1 and 2 months respectively, and are included to control for the severity of recent violence at a given location. The mean death value for each violent event is 1.61 deaths, ranging from a minimum of 1 to a maximum of 906.

---

[89] Center for International Earth Science Information Network, Columbia University, "Documentation for the Gridded Population of the World, Version 4 (GPWv4)" (Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2016), http://dx.doi.org/10.7927/H4D50JX4.

[90] Robert Muggah, *Researching the Urban Dilemma: Urbanization, Poverty and Violence* (International Development Research Centre, 2012); John D. McCarthy, Omer R. Galle, and William Zimmern, "Population Density, Social Structure, and Interpersonal Violence: An Intermetropolitan Test of Competing Models," *American Behavioral Scientist* 18, no. 6 (1975): 771–789.

C.   REGRESSION ANALYSIS

Negative Binomial regression models were used due to our dependent variable being a count of violent events, with a unit of analysis given by 2km grid cell-months. All independent variables were logged prior to regression analysis to adjust for heavy-tailed distributions.[91] All time-varying independent variables were lagged by one month, and each variable was then aggregated into the 2km grid cells as described above. The results from these models are presented in Chapter V (see Table 1).

---

[91] William H. Greene, "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models" (working paper, New York University Stern School of Business, 1994); Joseph M. Hilbe, *Negative Binomial Regression* (Cambridge University Press, 2011).

THIS PAGE INTENTIONALLY LEFT BLANK

V. RESULTS

Four regression models (Table 1) were used to examine the predictive relationship between *Violent Events* in Iraq and messages sent through the Twitter network. Additionally, seven independent control variables were selected to ensure against spurious relationships. The first model demonstrates the relationship between the seven control variables and *Violent Events*. The second model builds on the first model and includes a measure of the total number of observed Twitter messages to establish whether including message volume in the model results in a better fit for predictive analysis. The third model builds on the second model and includes *Negative Twitter Messages*, the total volume of the previous month's negative messages. The fourth model replaces the independent variable, *Negative Twitter Messages*, used in the third model, and instead uses the independent variable, *Extremely Negative Twitter Messages*, which is drawn only from the most extreme terms in the lexicon. This enabled us to determine if Twitter messages associated with extreme sentiment produced more statistically significant results.

The results of the models produced three noteworthy findings. Finding one shows that aggregate rates of Twitter messaging activity matter for predicting violence. In regions of Iraq with higher rates of Twitter messages, lower rates of violence tended to occur in subsequent months. When the total volume of Twitter messages is included as a variable in the second model, the AIC score improved, which indicates a better fit for predictive analysis. The second finding indicates that measurements of negative sentiment can improve the accuracy of violence prediction, beyond the predictions generated by aggregate message volume. More specifically, we find a negative relationship between Twitter messages containing negative sentiments and the subsequent amount of observed violence (Model 3 and Model 4). This relationship is the opposite of what was hypothesized and is counter-intuitive to the results that were expected. Finding three reveals that extreme sentiment matters even more. *Extremely Negative Twitter Messages* (Model 4) proved to have a higher statistical significance than *Negative Twitter Messages* (Model 3) and generates the best fit for predictive analysis

(lowest AIC score). Interestingly, this fourth model also indicates that the count of *Extremely Negative Twitter Messages* has a statistically significant negative coefficient, contrary to our hypothesis. The improved predictive successes of Model 4 may also indicate that the use of higher veracity words for lexical analysis may produce more accurate predictive models.

Table 1. Regression Models

	Dependent Variable: Violent Events			
	(1)	(2)	(3)	(4)
Negative Twitter Messages (Full Lexicon) (1 Month Lag)			-0.218*	
			(0.119)	
Extremely Negative Twitter Messages (Top 0.1% of Lexicon) (1 Month Lag)				-0.385***
				(0.136)
Total Twitter Messages (1 Month Lag)		-0.071*	0.089	-0.006
		(0.040)	(0.098)	(0.048)
Road Distance	-0.281*	-0.276*	-0.277*	-0.273*
	(0.149)	(0.150)	(0.149)	(0.149)
Night Light	1.277***	1.328***	1.254***	1.209***
	(0.151)	(0.154)	(0.160)	(0.159)
Population Density	-0.148***	-0.130**	-0.127**	-0.083
	(0.051)	(0.054)	(0.054)	(0.056)
Deaths (1 Month Lag)	0.519***	0.474***	0.464***	0.463***
	(0.156)	(0.156)	(0.157)	(0.154)
Deaths (2 Month Lag)	0.603***	0.575***	0.547***	0.546***
	(0.125)	(0.126)	(0.128)	(0.127)
Violent Events (1 Month Lag)	-0.503	-0.400	-0.327	-0.300
	(0.379)	(0.380)	(0.383)	(0.376)
Violent Events (2 Month Lag)	0.315	0.437*	0.484**	0.479**
	(0.218)	(0.227)	(0.229)	(0.226)
Constant	-8.281***	-8.454***	-8.301***	-8.308***
	(0.524)	(0.541)	(0.545)	(0.541)
Observations	13,836	13,836	13,836	13,836
Log Likelihood	-975.350	-973.924	-972.469	-970.264
Akaike Inf. Crit.	1,966.700	1,965.848	1,964.937	1,960.529
Note:				*p< 0.1; **p<0.05; ***p<0.01

Logarithmic transformation applied to all independent variables.

A.   FINDING ONE—TWITTER MATTERS

Model 1 demonstrates the relationship between the seven independent variables and the dependent variable, *Violent Events*. These seven independent variables are used as control variables. In the model, most control variables acted as expected. The variable for *Road Distance* resulted in a negative coefficient, indicating the greater distance from a road results in less violence. The variable for *Night Light* resulted in a significant positive coefficient, signifying that violent events occur more frequently in developed areas corresponding to high nighttime light emissions. The variables *Deaths (1 Month Lag)* and *Deaths (2 Month Lag)* both produced significant positive coefficients, indicating that violence tends to follow in the wake of violence. The only exception to these expectations was the variable for *Population Density*, which produced a significant negative coefficient, meaning that in more populated areas, less violence occurred. However, this may be generated by collinearity between population density and our other measures of infrastructure. In addition, the independent variable for *Violent Events (1 Month Lag)* resulted in a negative coefficient, while the independent variable for *Violent Events (2 Month Lag)* produced a positive coefficient, indicating the length of a lag period makes a difference in predictive analysis regarding *Violent Events* in a month-long temporal periods. Model 1 serves as a baseline for the predictive analysis of the number of *Violent Events* across Iraq.

Model 2 expands on Model 1 by including *Total Twitter Messages* as an independent variable. This variable assesses the volume of Twitter messages as a means for predictive analysis rather than the type of sentiment observed in the Twitter message. The coefficient for *Total Twitter Messages* is negative and statistically significant. This relationship implies that regions of Iraq that produce higher rates of Twitter messages tended to produce lower rates of violent events. Including *Total Tweets* into the model resulted in a lower AIC score between Model 1 and Model 2 by the difference of 0.852. This indicates that including *Total Twitter Messages* as a variable in the equation generates a better fit for predictive analysis.

B.   FINDING TWO—NEGATIVE SENTIMENT IMPROVES PREDICTIONS

Model 3 compares the rate of *Negative Twitter Messages (Full Lexicon) (1 Month Lag)* to *Violent Events* while also including the seven control variables and *Total Twitter Messages*. In the model, *Negative Twitter Messages* resulted in a moderately significant negative coefficient (p < 0.1). This indicates that the spatial regions in Iraq with *Negative Twitter Messages* in a particular month tend to be the same spatial regions that are less likely to experience violence in the following month. Additionally, the model produced a better fit than previous models. The difference between AIC scores in Model 2 and Model 3 resulted in a lower AIC score by 0.911. This indicates a better fit for predictive analysis when *Negative Twitter Messages* is used as the independent variable.

C.   FINDING THREE—EXTREME SENTIMENT MATTERS MORE

Model 4 produced the best fit for predictive analysis compared to the other models. The negative and highly significant coefficient for *Extremely Negative Twitter Messages (Top 0.1% Lexicon) (1 Month Lag)* resulted in a p-value of $p < 0.01$. This indicates a highly significant factor in the results for predicting *Violent Events*. The AIC value also resulted in the lowest score of all the models. The AIC score for Model 4 is considerably lower than Model 3 by a difference of 4.408 points. Model 4 thus outperforms Model 2 and Model 3, both in terms of statistical significance and in terms of predictive success in analyzing the relationship between Twitter messages and violence. This indicates that using the most extreme subset of negative sentiment terms for lexical analysis produced statistically more accurate results than using the full AraSenTi Lexicon set of negative sentiment terms.

A visualization of the negative binomial regression results for Model 4 (Figure 9. A-F) demonstrates how each statistically significant independent variable impacts the expected value of the dependent variable (*Violent Events*), when all other variables in Model 4 remain fixed at their means. This method allows us to isolate the independent variables and visually understand their effects in the model. The Y-axis on each figure is the expected outcome in terms of number of *Violent Events* in a grid cell-month. The X-axis shows a shift in the independent variable from its minimum to maximum value. The

red line in the visualizations is the prediction line, which demonstrates how the expected number of *Violent Events* changes as a function of the independent variables. The line's slope indicates either a negative or a positive relationship of the independent variable in relation to *Violent Events*. The grey portions around the prediction line represent the 95 percent confidence interval of the prediction.

For example, *Road Distance* ($p < 0.1$) (Figure 9A) demonstrates a negative relationship to *Violent Events*. Essentially, the further away from a road, the less amount of violence is expected to occur. This is intuitive, as the rural desert terrain of Iraq would likely cause less people to interact the further away people get from a road. The independent variable *Night Light* ($p < 0.01$) (Figure 9B) illustrates a highly significant positive relationship to *Violent Events*. At zero levels of *Night Light*, we see an expected count of *Violent Events* close to zero, but once the *Night Light* variable is increased to its maximum value the expected number of *Violent Events* increases towards 0.03. This relationship is also intuitive. In rural areas with low emissions of *Night Light*, people are likely to interact less and *Violent Events* are less likely to occur. *Violent Events (2 Month Lag)* ($p < 0.05$) (Figure 9C), also has a positive effect on the expected count of *Violent Events*. Both *Deaths (1 Month Lag)* ($p < 0.01$) (Figure 9D) and *Deaths (2 Month Lag)* ($p < 0.01$) (Figure 9E) also show highly significant positive effects.

Most importantly, the visualization for *Extremely Negative Twitter Messages (Top 0.1% Lexicon) (1 Month Lag)* ($p < 0.01$) (Figure 9F) illustrates the negative and highly significant effect on the expected count of *Violent Events*. Moving from the zero level of *Extremely Negative Twitter Messages* to the maximum value, the expected level of *Violent Events* significantly decreases by nearly 80 percent. This counter-intuitive relationship between *Extremely Negative Twitter Messages* and *Violent Events* is the opposite of our hypothesis. This represents a noteworthy finding for understanding violence through social media, because if our intuitive understanding of the relationship between sentiments and violence in a conflict zone is fundamentally wrong, then further research in this field is required.
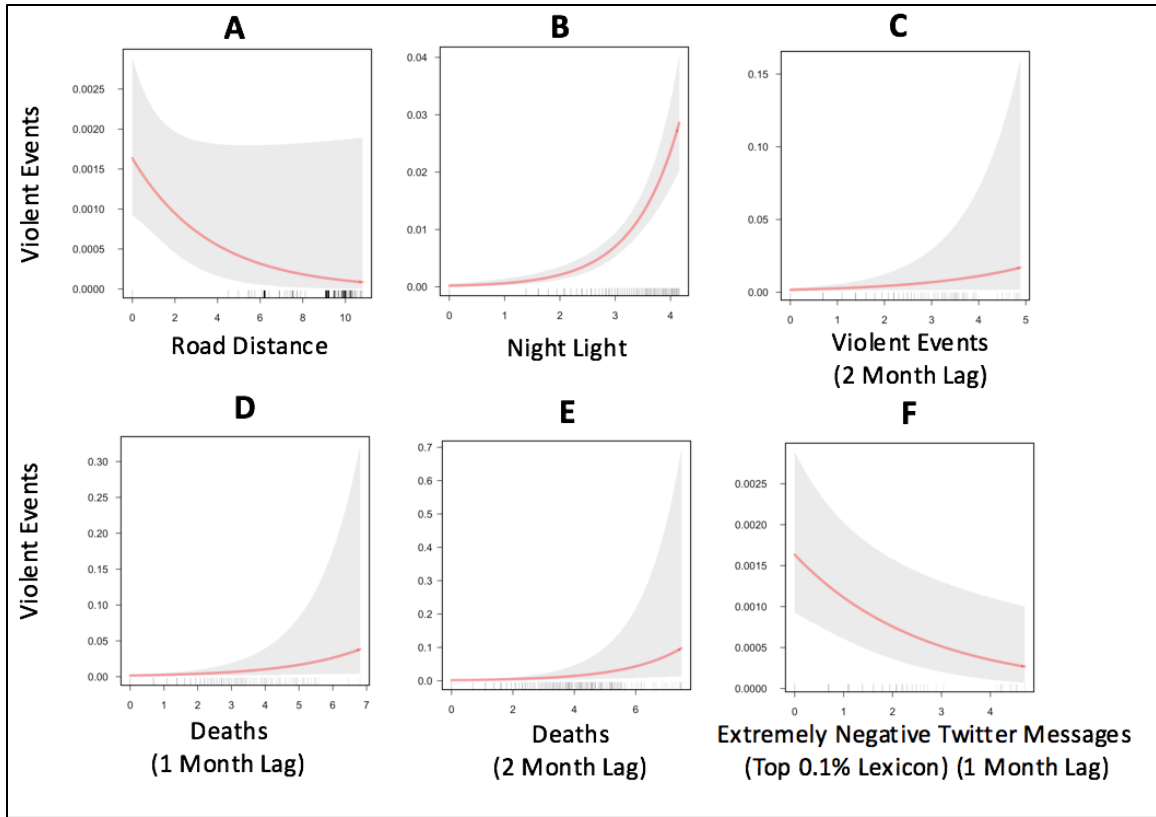
Figure 7. Predicted Effects of Statistically Significant Regression Variables
(Model 4)

THIS PAGE INTENTIONALLY LEFT BLANK

VI.   ADDITIONAL RESEARCH

Our efforts have also identified four areas of potential future research. These research areas involve expanding on the lexicon and threats discussed in this thesis in addition to replicating the research methods used in other operational areas of interest.

A.   SENTIMENT FOLLOWING VIOLENT EVENTS

Additional research is required to determine sentiment following violent events. Understanding the leanings of various population groups following a specific violent event could illuminate who a population tends to blame for events and whether that coincides with the facts on the ground. Understanding those trends and what types of violent acts generate the most negative sentiment by a population could aid the U.S. in choosing which tools to use to prevent or mitigate the catastrophic loss of local support. We recommend using the same conflict data and adjusting the lexicon to build word lists that would determine positive and negative sentiments towards the government and non-state actors. Our research only addressed general sentiments. It did not address whether there was more popular support for ISIS or the Iraqi government. ISIS occupied vast areas of Iraq and seized the city of Mosul with little resistance. An additional lexical analysis that specifically includes terms for ISIS and the Iraqi government could differentiate popular support during the conflict. Understanding the relationship between popular support and violent events could provide insight on how popular support affects insurgencies.

B.   DOES THE NATURE OF THE CONFLICT MATTER?

Our research found that there is a negative relationship between tweets with negative sentiments and the occurrence of violent events in the following month. This contradicts the expectations drawn from previous sentiment analysis research. We believe the nature of the ISIS conflict could have been responsible for this occurrence due to the increased usage of social media to recruit and spread pro-ISIS propaganda. The usage of social media by ISIS might have influenced increased social media usage by the general population to stay abreast of threat activities. Additional research in a different conflict

zone where the predominant threat group had limited or no social media utilization could be conducted to confirm or deny this speculation.

C.   ARABIC LEXICONS AND TWITTER

This research utilized a large-scale, Twitter-specific Arabic sentiment lexicon. Arabic sentiment lexicons continue to be insufficient at addressing multiple Arabic dialects, as well as the nuances of sarcasm. We did not conduct human analysis to translate the Twitter messages to determine the accuracy of the sentiments expressed in the Iraqi geo-referenced tweets. Additionally, we did not conduct lexical analysis of other languages, to include English. Since Twitter allows users to connect across the globe, tweets geo-referenced in Iraq, but written in foreign languages are likely targeted towards international audiences and less likely towards local audiences. Furthermore, we did not attempt to identify Twitter bot accounts. Bot accounts are computer-generated user accounts designed to amplify specific messages by constantly tweeting and retweeting messages. It is extremely difficult to identify these accounts, but many bot accounts are often accounts with short life spans. Eliminating tweets from accounts with short life spans could serve to eliminate some of these bots. Arabic Twitter analysis continues to be an area requiring further research and methods to increase the accuracy of sentiment results.

VII. CONCLUSION

This thesis produced three findings: (1) Measuring spatial and temporal variation in the volume of Twitter messages improved predictions of the timing and location of violent events in Iraq, (2) predictions were further enhanced by measurement of terms reflecting negative sentiments in local dialects, and (3) extremely negative terminology seems to be more useful than moderately negative terminology in generating accurate predictions of violent events. The thesis began by asking the research question, "*How can social media analysis be used as a tool for predicting violent events and understanding social sentiment during violent events?*" This research was pursued because there is a gap in understanding how social media sentiments are related to violent events. The research specifically focused on analyzing social media sentiments through Twitter messages geo-referenced in Iraq from August 1, 2013, through July 31, 2014. Iraq was selected as the case study due to the high level of violence and conflict observed during this time period, which corresponded with ISIS violently occupying large areas of Iraq. Additionally, understanding how social sentiments in Iraq relate to acts of violence is important as Iraq continues to be strategically significant to U.S. interests and regional stability.

To analyze Iraqi sentiment, as expressed through social media, a sentiment dictionary was required. Instead of translating an English lexicon to Arabic, the AraSenti lexicon was selected for its specific strength of including the use of informal Arabic language associated with Twitter messages. This lexicon enabled us to categorize Iraqi tweets into positive or negative sentiment categories. These tweets also provided geo-referenced metadata, allowing us to spatially analyze the use of Arabic terminology across Iraq.

This research uncovered an inverse relationship between negative sentiment and levels of violence, which contradicted initial assumptions. When negative words were used, subsequent violent events occurred at lower rates. Understanding why this occurred could change how we think about operations and the sentiment of a population in conflict zones. It may be that when violence is lower, one is be more willing to express negative sentiments without fear of reprisal. Another possibility lies in the difference between

sadness and anger-related words. It is possible that a further division of negative words by more specific emotional content could produce a different result. Additionally, the ability to express negative feelings through social media might decrease the need to express discontent through violence. In this instance, social media might have provided an emotional release of discontent that actually prevented the expression of discontent in more dangerous ways.

One implication of this finding is that military commands should consider promoting the use and implementation of mediums where a local population can express discontent as an outlet to relieve pressure. Providing individuals access to social media, town hall meetings, and other public forums where freedom of speech is allowed could lower the expression of discontent through violence. Even in cultures where expression is restricted in public forums, social media provides an outlet that is less restrictive. However, because our data does not allow us to differentiate between these mechanisms, future research will be required to confirm or disconfirm these hypotheses.

LIST OF REFERENCES

Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums." *ACM Transactions on Information Systems (TOIS)* 26, no. 3 (2008).

Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. "Sentiment Analysis of Twitter Data." In *Proceedings of the Workshop on Languages in Social Media*, edited by Meena Nagarajan and Michael Gamon, 30–38. Association for Computational Linguistics, 2011.

Alexa Internet, Inc. "The Top 500 Sites on the Web." Accessed February 14, 2017. http://www.alexa.com/topsites.

Al-Twairesh, Nora, Hend Al-Khalifa, and AbdulMalik Al-Salman. "AraSenTi: Large-Scale Twitter-Specific Arabic Sentiment Lexicons." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,* 697–705. Berlin, Germany: August 7–12, 2016.

Amos, James F., McRaven, William H., and Odierno, Raymond. "Strategic Landpower: Winning the Clash of Wills." White paper, United States Army, United States Marine Corps and the United States Special Operations Command, 2013.

Ansolabehere, Stephen, and Eitan Hersh. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." *Political Analysis* 20, no. 4 (Autumn 2012): 437–459, http://www.jstor.org/stable/23359641.

Asur, Sitaram, and Bernardo A. Huberman. "Predicting the Future with Social Media." In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* 1 (August–September 2010): 492–499. http://ieeexplore.ieee.org/document/5616710/.

Barberá, Pablo, and Gonzalo Rivero. "Understanding the Political Representativeness of Twitter Users." *Social Science Computer Review* 33, no. 6 (December 1, 2015): 712–729.

Berger, J. M., and Jonathon Morga. *The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter.* Analysis paper no. 20, Brookings Project on U.S. Relations with the Islamic World, March 2015. https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf.

Bifet, Albert, and Frank Eibe. "Sentiment Knowledge Discovery in Twitter Streaming Data." In *DS'10 Proceedings of the 13th International Conference on Discovery Science*, edited by Bernhard Pfahringer, Geoff Holmes, and Achim Hoffmann, 1–15. Berlin, Germany: Springer-Verlag, 2010.

Blanchard, Christopher M. *Islam: Sunnis and Shiites*. Washington, DC: Congressional Research Service, 2009.

Bollen, Johan, Alberto Pepe, and Huina Mao. "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena." In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media,* edited by N. Nicolov and J. G. Shanahan, 1–10. Barcelona, Spain: AAAI Press, July 2011.

Bourret, Andrew K., Joshua D Wines, and Jason M Mendes. "Assessing Sentiment in Conflict Zones through Social Media." Master's thesis, Naval Postgraduate School, 2016. http://calhoun.nps.edu/handle/10945/51650.

Cambria, Erik, Björn Schuller, Yunqing Xia, and Catherine Havasi. "New Avenues in Opinion Mining and Sentiment Analysis." *IEEE Intelligent Systems* 28, no. 2 (March/April 2013): 15–21. http://222.252.30.203:8888/bitstream/123456789/13029/1/Intelligent%20systems%20and%20their%20applications.Vol.28.Iss.2.A.6.pdf.

Center for International Earth Science Information Network, Columbia University. "Documentation for the Gridded Population of the World, Version 4 (GPWv4)." Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2016. http://dx.doi.org/10.7927/H4D50JX4.

Center for International Earth Science Information Network, Columbia University and Information Technology Outreach Services, and University of Georgia. "Global Roads Open Access Data Set, Version 1 (gROADSv1)." Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2013. http://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1.

Central Intelligence Agency. "Middle East: Iraq." In *The World Fact Book 2017*. Washington, DC: Central Intelligence Agency, 2017. https://www.cia.gov/library/publications/resources/the-world-factbook/geos/print_iz.html.

Chatfield, Akemi Takeoka, Christopher G. Reddick, and Uuf Brajawidagda. "Tweeting Propaganda, Radicalization and Recruitment: Islamic State Supporters Multi-Sided Twitter Networks." In *Proceedings of the 16th Annual International Conference on Digital Government Research* (ACM, 2015), 239–249.

Croicu, Mihai, and Ralph Sundberg. "UCDP GED Codebook Version 5.0." Department of Peace and Conflict Research, Uppsala University, 2016.

Department of the Army. *Intelligence.* FM 2-0. Washington, DC: Department of the Army, 2004.

Diamond, Larry. "What Went Wrong in Iraq." *Foreign Affairs* (September/October 2004): 34–56.

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter." *PloS ONE* 6, no. 12 (2011): e26752.

Fan, Rui, Jichang Zhao, Yan Chen, and Ke Xu. "Anger Is More Influential than Joy: Sentiment Correlation in Weibo." *PloS ONE* 9, no. 10 (2014): e110184.

Gerges, Fawaz A. "ISIS and the Third Wave of Jihadism." *Current History* 113, no. 767 (2014): 339–343.

Ghiassi, M., J. Skinner, and D. Zimbra. "Twitter Brand Sentiment Analysis: A Hybrid System Using N-Gram Analysis and Dynamic Artificial Neural Network." *Expert Systems with Applications* 40, no. 16 (2013): 6266–6282.

Glenn, Cameron. "Timeline: Rise and Spread of the Islamic State." Wilson Center. July 5, 2016. https://www.wilsoncenter.org/article/timeline-rise-and-spread-the-islamic-state.

Golder, Scott A. and Michael W. Macy. "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength across Diverse Cultures." *Science* 333, no. 6051 (2011): 1878–1881.

Greene, William H. "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models." Working paper, New York University Stern School of Business, 1994.

Gulmohamad, Zana K. "The Rise and Fall of the Islamic State of Iraq and Al-Sham (Levant) ISIS." *Global Security Studies* 5, no. 2 (Spring 2014): 4.

Harding, Luke, and Agencies in Baghdad. "Iraq Suffers Its Deadliest Year Since 2008." *Guardian*. January 1, 2014. https://www.theguardian.com/world/2014/jan/01/iraq-2013-deadliest-year-since-2008.

Hilbe, Joseph M. *Negative Binomial Regression.* Cambridge University Press, 2011.

Huang, Shu, Wei Peng, Jingxuan Li, and Dongwon Lee. "Sentiment and Topic Analysis on Social Media: A Multi-Task Multi-Label Classification Approach." In *Proceedings of the 5th Annual ACM Web Science Conference.* New York: ACM, 2013. doi:10.1145/2464464.2464512.

Human Rights Watch. "Iraq: A Broken Justice System." January 31, 2013.
https://www.hrw.org/news/2013/01/31/iraq-broken-justice-system.

Humanitarian Information Centre for Iraq. "Iraq Districts." Accessed on October 25,
2017.
https://www.globalsecurity.org/military/world/iraq/images/013_A4_Iraq_Districts
.jpg.

Imhoff, Marc L., William T. Lawrence, David C. Stutzer, and Christopher D. Elvidge. "A
Technique for Using Composite DMSP/OLS 'City Lights' Satellite Data to Map
Urban Area." *Remote Sensing of Environment* 61, no. 3 (1997): 361–370.

Jackson, Gary. *Predicting Malicious Behavior: Tools and Techniques for Ensuring
Global Security.* 1st ed. Indianapolis, IN: John Wiley and Sons, 2012.

Jiang, Long, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. "Target-Dependent
Twitter Sentiment Classification." In *Proceedings of the 49th Annual Meeting of
the Association for Computational Linguistics: Human Language Technologies* 1.
Stroudsburg, PA: Association for Computational Linguistics, 2011.

Kaseb, Gehad S., and Mona F. Ahmed. "Arabic Sentiment Analysis Approaches: An
Analytical Survey." *International Journal of Scientific & Engineering Research*
7, no. 10 (October 2016): 712–723.

Klausen, Jytte. "Tweeting the Jihad: Social Media Networks of Western Foreign Fighters
in Syria and Iraq." *Studies in Conflict & Terrorism* 38, no. 1 (2015): 1–22.
doi:10.1080/1057610X.2014.974948.

Laub, Zachary, and Jonathan Masters. "The Islamic State in Iraq and Syria." Council on
Foreign Relations. August 8, 2014.
http://www.aina.org/news/20140808123646.pdf.

Lewis, Jessica D. *Al-Qaeda in Iraq Resurgent: The Breaking the Walls Campaign, Part
1.* Washington, DC: Institute for the Study of War, September 2013.
http://www.understandingwar.org/sites/default/files/AQI-Resurgent-
10Sept_0.pdf.

Ma, Ting, Chenghu Zhou, Tao Pei, Susan Haynie, and Junfu Fan. "Quantitative
Estimation of Urbanization Dynamics Using Time Series of DMSP/OLS
Nighttime Light Data: A Comparative Case Study from China's Cities." *Remote
Sensing of Environment* 124 (2012): 99–107.

McCarthy, John D., Omer R. Galle, and William Zimmern. "Population Density, Social
Structure, and Interpersonal Violence: An Intermetropolitan Test of Competing
Models." *American Behavioral Scientist* 18, no. 6 (1975): 771–789.

Mislove, Alan, Sune Lehmann Jørgensen, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist. "Understanding the Demographics of Twitter Users." In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.* AAAI Press, 2011, 554–557.

Mitchell, Lewis, Morgan R. Frank, Kameron Decker Harris, Peter Sheridan Dodds, nd aChristopher M. Danforth. "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place." *PLOS ONE* 8, no. 5 (2013): e64417. https://doi.org/10.1371/journal.pone.0064417.

Muggah, Robert. *Researching the Urban Dilemma: Urbanization, Poverty and Violence.* International Development Research Centre, 2012.

Nagarajan, Meena, Amit Sheth, and Selvam Velmurugan. "Citizen Sensor Data Mining, Social Media Analytics and Development Centric Web Applications." In *Proceedings of the 20th International Conference Companion on World Wide Web*. ACM, 2011.

National Intelligence Council. *Global Trends 2030: Alternative Worlds.* Washington, DC: Office of the Director of National Intelligence, December 2012.

Feller, A., M. Kuhnert, T. O. Sprenger, and I. M. Welpe. "Divided They Tweet: The Network Structure of Political Microbloggers and Discussion Topics." In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, edited by N. Nicolov and J. G. Shanahan. Menlo Park, CA: AAAI Press, 2011.

Nielsen, Finn Årup. "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs." In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big Things Come in Small Packages,* 93–98. 2011. arXiv preprint arXiv: 1103.2903.

NOAA National Centers for Environmental Information. "DMSP-OLS Nighttime Lights Time Series." National Geophysical Data Center, Boulder, CO. Last modified June 7, 2017. http://noaa.maps.arcgis.com/home/item.html?id=78147366ebac4f28862053b89169dfc9.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." In *Proceedings of the International AAAI Conference on Weblogs and Social Media* 11, no. 122–129, 1–2. Washington, DC: 2010.

Pierskalla, Jan H., and Florian M. Hollenbach. "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa." *American Political Science Review* 107, no. 2 (2013): 207–24. doi:10.1017/S0003055413000075.

Polykalas, Spyros E., George N. Prezerakos, and Agisilaos Konidaris. "An Algorithm Based on Google Trends' Data for Future Prediction. Case Study: German Elections." In *2013 IEEE International Symposium on Signal Processing and Information Technology* (December 2013): 69–73. http://ieeexplore.ieee.org/document/6781856.

Scott, John. *Social Network Analysis*. Thousand Oaks, CA: SAGE, 2012.

Shapiro, Jacob N., and Nils B. Weidmann. "Is the Phone Mightier than the Sword? Cellphones and Insurgent Violence in Iraq." *International Organization* 69, no. 2 (2015): 247–274.

Sowell, Kirk H. "Maliki's Anbar Blunder." *Foreign Policy,* January 15, 2014, https://foreignpolicy.com/2014/01/15/malikis-anbar-blunder/.

Statista. "Acting Social Network Penetration in Selected Countries as of January 2017." 2017. https://www.statista.com/statistics/269615/social-network-penetration-by-region/.

Stratfor. "Iraq's Population Density by Ethnic and Sectarian Divisions." August 10, 2014. Accessed September 30, 2017, https://twitter.com/stratfor/status/498544417976770562?lang=en/.

Sundberg, Ralph, and Erik Melander. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50, no.4 (2013): 523–532.

Tunnicliffe, Ian, and Steve Tatham. *Social Media—The Vital Ground: Can We Hold It?* Carlisle, PA: Strategic Studies Institute and U.S. Army War College Press, 2017.

Twitter. "#numbers." Accessed February 14, 2017. http://about.twitter.com.

Uppsala Conflict Data Program. "UCDP Conflict Encyclopedia." Uppsala University. Accessed February 17, 2017. www.ucdp.uu.se.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47, no. 4 (2010). http://journals.sagepub.com/doi/pdf/10.1177/0022343309356491.

Warren, T. Camber. "Explosive Connections? Mass Media, Social Media, and the Geography of Collective Violence in African States." *Journal of Peace Research* 52, no. 3 (2015): 297–311.

———. "Mapping the Rhetoric of Violence: Political Conflict Discourse and the Emergence of Identity Radicalization in Nigerian Social Media." Presented at the Annual Meeting of the American Political Science Association, San Francisco, CA, September 3–6, 2015.

Weidmann, Nils B., and Michael D. Ward. "Predicting Conflict in Space and Time." *Journal of Conflict Resolution* 54, no. 6 (2010): 883–901. http://www.jstor.org/stable/25780759.

Weidmann, Nils B., Doreen Kuse, and Kristian Skrede Gleditsch. "The Geography of the International System: The CShapes Dataset." *International Interactions* 36, no. 1 (2010): 86–106.

Wright, Shaun, David Denney, Alasdair Pinkerton, Vincent A. A. Jansen, and John Bryden. "Resurgent Insurgents: Quantitative Research into Jihadists Who Get Suspended but Return on Twitter." *Journal of Terrorism* 7, no. 2 (May 2016): 1. http://doi.org/10.15664/jtr.1213.

Wyer, Sam. "Political Update: Mapping the Iraq Protests." Figure 2. Institute for the Study of War. January 11, 2013, http://www.understandingwar.org/backgrounder/political-update-mapping-iraq-protests.

Younggue, Bae, and Lee Hongchul. "Sentiment Analysis of Twitter Audiences: Measuring the Positive or Negative Influence of Popular Twitterers." *Journal of the American Society for Information Science and Technology* 63, no. 12 (2012).

Yu, Sheng, and Subhash Kak. *A Survey of Prediction Using Social Media*. Stillwater, OK: Department of Computer Science, Oklahoma State University, 2012.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California