AN ANALYSIS OF DISTURBANCES IN CRITICAL ENERGY INFRASTRUCTURE THROUGH SOCIAL MEDIA

by

Katelyn Thomas
A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Geoinformatics and Geospatial Intelligence

Committee:

____________________	Dr. Arie Croitoru, Thesis Director
____________________	Dr. Anthony Stefanidis, Committee Member
____________________	Dr. Andrew Crooks, Committee Member
____________________	Dr. Dieter Pfoser, Department Chairperson
____________________	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
____________________	Dr. Peggy Agouris, Dean, College of Science
Date: ________________	Spring Semester 2018 George Mason University Fairfax, VA

An Analysis of Disturbances in Critical Energy Infrastructure Through Social Media

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Katelyn Thomas
Bachelor of Science
George Mason University, 2016

Director: Arie Croitoru, Associate Professor
Department of Geography and Geoinformation Science

Spring Semester 2018
George Mason University
Fairfax, VA

ACKNOWLEDGEMENTS

I would first like to thank my loving family and friends for all of their continuous support and encouragement throughout my years of study, especially through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

I would like to acknowledge and thank Dr. Arie Croitoru from George Mason University's Department of Geography and Geoinformation Science, who has supported me throughout my thesis with his patience, knowledge, and invaluable help.

I would also like to acknowledge members of my committee, Dr. Anthony Stefanidis and Dr. Andrew Crooks, for their time and interest in my thesis.

Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Alberta Electric System Operator	AESO
Ambient Geographic Information	AGI
American Community Survey	ACS
American Electric Power	AEP
California ISO	CAISO
Comma-Separated Values	CSV
Department of Energy	DOE
District of Columbia	D.C.
Electric Reliability Council of Texas	ERCOT
Electrical Reliability Organization	ERO
Energy Information Administration	EIA
Event Detection with Clustering of Wavelet-based Signals	EDCoW
Federal Emergency Management Agency	FEMA
Federal Energy Regulatory Commission	FERC
feet	ft
Florida Reliability Coordinating Council	FRCC
Geostationary Operational Environmental Satellite	GEOS
inches	in
Independent Electricity System Operator	IESO
Independent System Operators	ISO
ISO New England	ISONE
Midcontinent Independent System Operator	MISO
Midwest Reliability Organization	MRO
National Hurricane Center	NHC
nautical mile	nmi
Network Interface Device	NID
New York ISO	NYISO
North American Electric Reliability Corporation	NERC
Northeast Power Coordinating Council	NPCC
Outage Management System	OMS
Portable Document Format	PDF
Regional Transmission Organizations	RTO
ReliabilityFirst Corporation	RFC
Southeastern Electric Reliability Council	SERC
Southwest Power Pool	SPP
Texas Regional Entity	TRE

Volunteered Geographic Information .... VGI
Western Energy Coordinating Council .... WECC
United States .... U.S.

ABSTRACT

AN ANALYSIS OF DISTURBANCES IN CRITICAL ENERGY INFRASTRUCTURE THROUGH SOCIAL MEDIA

Katelyn Thomas, M.S.

George Mason University, 2018

Thesis Director: Dr. Arie Croitoru

The onset of power grid failures and outages due to severe weather happen instantly. During a cyclonic weather event, critical infrastructure sustains damages and can be destroyed from tropical-force winds, storm surge, flooding, and tornadoes. Electricity is the most vulnerable infrastructure to severe weather impacts. Often, damages to electrical equipment and their impact can be hard to locate. Also weather conditions may not permit for a safe dispatch to locate, assess, and repair utility equipment. Therefore, this study is tasked to examine how social media users report blackouts and damages to electrical equipment of utility providers. Specifically, it aims to explore the possible relationship between the volume and the spatial footprint of weather and power-related tweets to the spatial extent of power outages reported by utility companies. Social media platforms, such as Twitter, now play a pivotal role in crisis management during severe weather events. However, limited research has been

completed on the concept of using social media to predict future disruptions in the energy grid. In order to determine whether or not this concept is feasible, a geosocial, spatiotemporal, and geospatial analysis are carried out, examining the relationships between social media usage and power outages in two case studies: Hurricane Harvey (2017) and Superstorm Sandy (2012). The results of these case studies suggest that, at least in some cases, social media can serve as a possible information source about the occurrences and the spatial extent of power outages due to extreme weather events.

1 INTRODUCTION

Hurricane Harvey was a record-breaking hurricane that struck the Texas coastline. Hurricane Harvey became the first major hurricane (category 3 or higher) to hit the Texas coast since Hurricane Brett in 1999, the first hurricane to strike the Houston area since Hurricane Ike in 2008 [1], and the first major hurricane to make a United States landfall since Hurricane Wilma in 2005. On August 25th, Hurricane Harvey made landfall between Port Aransas and Rockport, Texas as a destructive category 4 hurricane. After making landfall, Harvey continued moving in an easterly direction before its center of circulation stalled over southeast Texas on August 26th. Over the next four days, the slow forward momentum caused devastating flooding and catastrophic damage. The Harris County Flood Control District reported a total of 1 trillion gallons of water fell across Harris County over the four-day period [2].

Superstorm Sandy was also a once in a lifetime storm, whose track took aim at the United States' east coast, placing nearly 50 million people in harm's way. At the time of its landfall, Sandy was the second costliest storm to ever hit the U.S. [3]. Also, Sandy holds the record for the largest diameter of a tropical cyclone [3]. The track of powerful Sandy proved to be a worst-case scenario for the mid-Atlantic states. As Sandy tracked north, it merged with an Arctic air mass and begun to lose some of its tropical characteristics; no longer a hurricane, Sandy earned the title of *superstorm*. Late on October 29th, Sandy made its landfall near Brigantine, New Jersey as a post-tropical cyclone [3]. Landfall occurred simultaneously with high tide, which resulted in record tide levels [4]. Superstorm Sandy delivered strong winds, record flooding, extensive rainfall, and even a blizzard to the east coast— and is often referred to as a *perfect storm*. Overall, 24 states were impacted by Superstorm Sandy and years later are still in recovery [5].

As a result of Hurricane Harvey and Superstorm Sandy, several states experienced severe damage to critical infrastructures. For this study, critical infrastructure is defined as sectors whose assets, systems, and networks, whether physical or virtual, are considered so vital to the United States that their incapacitation or destruction would have a debilitating effect on security, national economic security, and/or national public health or safety [6]. Since the energy sector sustained damages from Harvey and Sandy, which negatively impacted security, economic security, and public health, it will serve as the focal point of this research study.

During Hurricane Harvey, Texas' energy sector suffered major damages which led to widespread power outages. The power outages stemmed from utility poles and transmission structures that were knocked down, destroyed or flooded. The severe flooding prevented utility crews from reaching damaged equipment and extended the duration of outages. Also, all flooded electrical equipment and circuit breakers had to be inspected by utility crews before they could reactivate and restore electricity. Four major power and utility pools that were affected were American Electric Power Texas (AEP Texas), Electric Reliability Council of Texas (ERCOT), CenterPoint Energy, and Entergy Corp [7]. Overall, power outages were felt by 1.7 million customers across fourteen different power and utility pools [7] [8]. The longest outage lasted more than 12 and a half days [7]. The number of Texas outages (within a 24-hour period) peaked on August 29th with 312,698 customers without electricity. The peak of Louisiana's power outages (within a 24-hour period) occurred on August 30th with 11,857 customers without power [8].

In the aftermath of Superstorm Sandy, 21 states' energy sectors were affected. Similar to Hurricane Harvey, power outages originated from downed utility poles and flooded transmission structures. In the following days, electrical structures remained underwater making it impossible to restore power until the water receded. Power outages affected 8.5 million customers [3] across 22 electric and utility companies [9]. The lapse in electricity was felt as northern as Maine, southern as North Carolina, and western as Illinois [10]. The longest outage lasted over 14 days [9]. The total number of customers without electricity peaked on October 30th with 8.2 million customers [10]. Also, on this day the percentage of customers without power in the states of New Jersey and New York peaked at 65% and 23%, respectively [10].

One of the most common ways utility companies are notified of a power outage is by its customers calling to report an outage in their neighborhood. During a catastrophic weather event, such as Hurricane Harvey and Superstorm Sandy, cell towers become clogged due to the high volumes of incoming and outgoing calls. Ideally, phone lines should be used only in cases of emergencies. Therefore, social media platforms, such as Twitter, can potentially offer the best avenue to deliver reports of down power lines, damages to electric equipment, and power outages. Tweets with geographic information or that are geotagged could allow utility technicians to quickly locate outages and possibly damaged electrical gear. Geotagged tweets with pictures enable utility technicians to visually assess environmental conditions and damaged equipment before dispatching a team for repairs. Twitter data can also aid utility providers in prioritizing critical areas by the hardest hit regions.

The integration of social media into every aspect of our lives, now offers assistance in a plethora of diverse disciplines. In extreme weather events, especially hurricanes, dangerous weather conditions can develop suddenly. These dangerous conditions have the potential to abruptly cause widespread blackouts. Additional resources, such as social media, could assist in pinpointing impacted areas experiencing power outages faster than traditional reporting. As a result, an increase the volume of Twitter activity could potentially identify areas with active power outages. Possibly, utility and electric companies could utilize the concept of event detection in social media platforms to serve as an extra form of verification. Current research explores whether

Twitter can play a pivotal role in crisis management before, during, and after severe weather events.

This introduction provides brief glimpses into selected topics, motivations, and possible uses for this research. Related research that is complementary to themes explored throughout this introduction and research study will be introduced in chapter 2. Topics covered include volunteer geographic information and event detection in social media and background information about Hurricane Harvey, Superstorm Sandy, and the U.S. energy grid. The following chapter, chapter 3, will present the objectives and research questions. Then, the data sources and their structures and characteristics are discussed in chapter 4. Chapter 5 outlines the methodology used to analyze and answer each research question defined in chapter 3. The results from each analysis are displayed in chapter 6 and are further discussed and analyzed in chapter 7. Lastly, chapter 8 will state this study’s findings, discuss the challenges and limitations, and possible areas for future research.

2 LITERATURE REVIEW

The literature review provides a glimpse into related research and scientific studies. The supplied literature provides brief introductions to the meteorological histories of Hurricane Harvey and Superstorm Sandy, social media's role in volunteered geographic information, event detection in social media, the U.S. energy grid, and advancements in power outage detections. This literature review aids in understanding the purpose of selected research questions, data sources and analysis techniques harnessed by this research study. The given literature review arms the reader with contextual knowledge of event detection in social media and how it can be utilized to detect disturbances in critical infrastructure due to the aftermath of Hurricane Harvey and Superstorm Sandy.

2.1 Meteorological History of Hurricane Harvey

On August 13th, the National Hurricane Center (NHC) began to monitor an area of low pressure located southwest of the Cape Verde Islands [11]. Like many powerful hurricanes, Hurricane Harvey originated as a tropical wave off the west coast of Africa. Six hours later, the area of low pressure was upgraded to Tropical Storm Harvey [12]. As Harvey transverse westerly across the Atlantic Ocean, the system was weakened by upper-level wind shear and degenerated to an open wave [13]. However, in just 56 hours over the Gulf of Mexico, Harvey underwent a period of rapid intensification and grew from a regenerated tropical depression into a category 4 hurricane [14]. On August 25th, Hurricane Harvey made landfall between Port Aransas and Rockport, Texas. Harvey broke the previous record of 54 hours by remaining a named storm for 117 hours after making a Texas landfall [15]. Within 24 hours of making landfall, the eye of Harvey begun to stall over south Texas before it gradually drifted back into the Gulf of Mexico [14]. Hurricane Harvey made a second landfall near Cameron, Louisiana on August 30th.

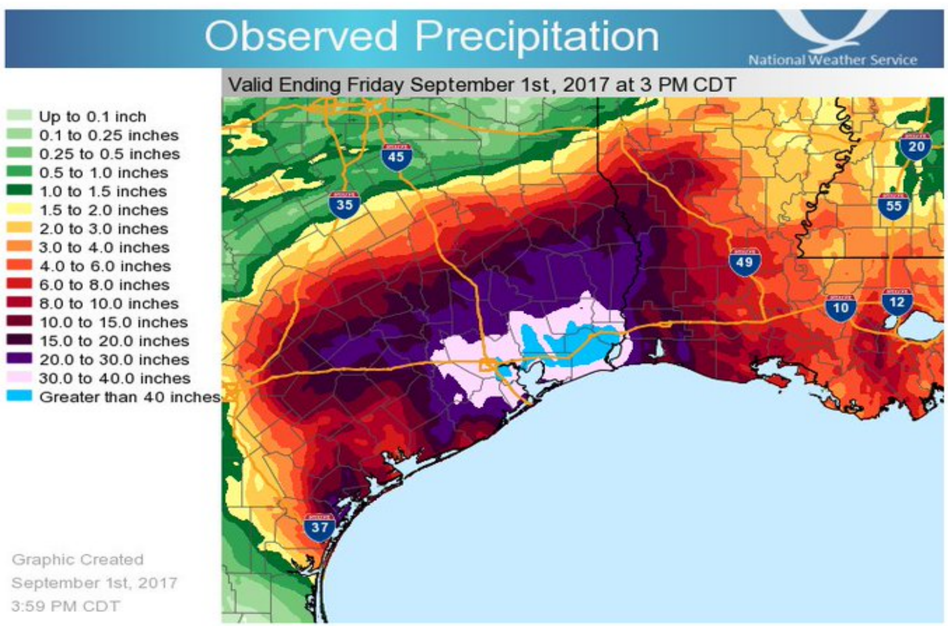

Figure 1: Observed 7-Day precipitation over southern Texas. Source: [14]

As Hurricane Harvey's 130 mph winds pushed ashore, homes and infrastructures were destroyed by its powerful wind bands and storm surge. Although, Harvey spawned off numerous tornadoes reaching from Texas to Tennessee [14], the most widespread damages were attributed to its monumental rainfall totals. The Federal Emergency Management Agency (FEMA) estimates over 19 trillion gallons of rainwater fell over parts of Texas [16]. Within seven days, over 6.7 million people within a 29,000-square mile area had received at least 20 inches of rain [17]. Historically, Harris County typically receives 50 inches of rainfall a year. However, as a result of Harvey stalling over southeast Texas, over two dozen rainfall gages registered seven-day readings topping 40 inches (Figure 1) [2]. The largest rainfall total ever recorded in the continental United States from a single storm was felt by Houston (51.88 inches) as a result of Hurricane Harvey [16]. Houston received a year worth of rain within seven days, resulting in approximately 780,000 Texans forced to evacuate their homes.

After the storm passed, roughly 42,000 Texans were left in temporary housing across 692 shelters [16] and a total of 176,219 homes were impacted by Harvey [18]. Harris County's 22 watersheds and approximately 120,000 structures and infrastructure were flooded [2]. Twenty-four hospitals were evacuated, 61 communities lost drinking water capability, 23 ports were closed, and roughly 780 roads were impassable [16]. Over the time period of Harvey, there were 122,331 people and 5,234 pets rescued by local, state and federal first responders. Despite emergency services working around the clock, Texas officials report approximately 80 people died as a result of Hurricane Harvey [19]. Texas Governor Abbott estimates the total damage from Hurricane Harvey is between $150 billion to $180 billion [20].

Federally Declared Areas of Major Disaster

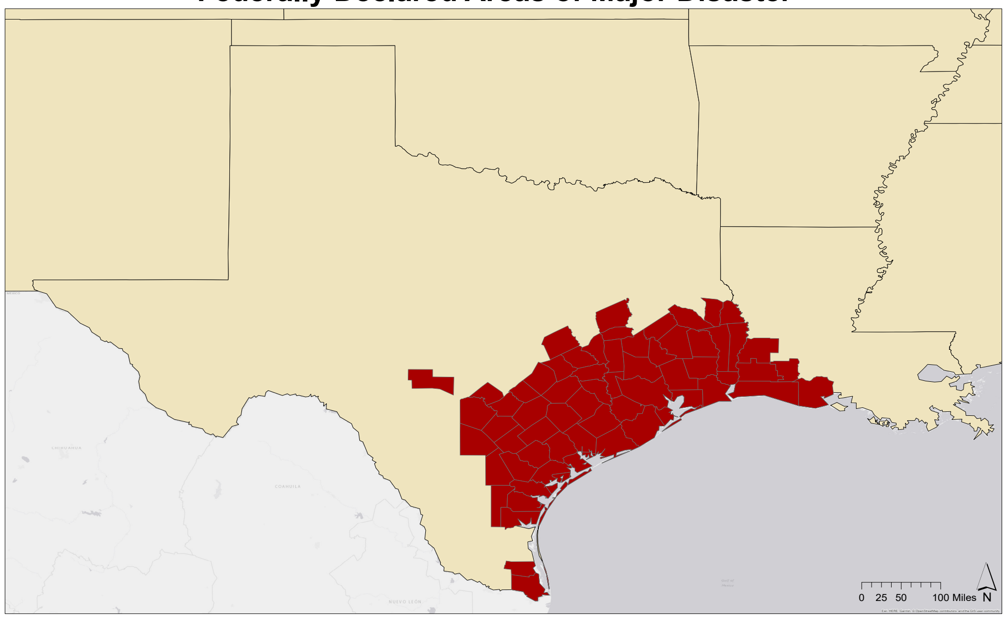

Figure 2: Highlighted counties (in red) were declared disaster areas in the wake of Hurricane Harvey.

Hurricane Harvey caused catastrophic damage throughout Texas and Louisiana. The recovery efforts continued months after the storm. Texas Governor Abbott renewed a State of Disaster proclamation on November 20th for 60 Texas counties. The counties that were declared major disaster areas are Angelina, Aransas, Atascosa, Austin, Bastrop, Bee, Bexar, Brazoria, Brazos, Burleson, Caldwell, Calhoun, Cameron, Chambers, Colorado, Comal, DeWitt, Fayette, Fort Bend, Galveston, Goliad, Gonzales, Grimes, Guadalupe, Hardin, Harris, Jackson, Jasper, Jefferson, Jim Wells, Karnes, Kerr, Kleberg, Lavaca, Lee, Leon, Liberty, Live Oak, Madison, Matagorda, Milam, Montgomery, Newton, Nueces, Orange, Polk, Refugio, Sabine, San Augustine, San Jacinto, San Patricio, Trinity, Tyler, Victoria, Walker, Waller, Washington, Wharton, Willacy and Wilson counties (Figure 2) [21]. An Emergency Declaration was declared for Louisiana on August 28th. Louisiana Governor Edwards designated Beauregard, Calcasieu, Cameron, Jefferson Davis, and Vermillion parishes as federal disaster areas (Figure 2) [22].

2.2 Meteorological History of Superstorm Sandy

Superstorm Sandy originated as a tropical wave off the west coast of Africa on October 11th, 2012 [3]. Over the next several days, the tropical disturbance encountered atmospheric conditions which inhibited further development. However, on October 18th the wave entered the eastern Caribbean Sea where environmental conditions became favorable for hurricane development [3]. By October 22nd and approximately 305nmi south-southwest of Kingston, Jamaica, the tropical low began to organize itself and was upgraded to a tropical depression [3]. Six hours later, an Air Force Reserve Hurricane Hunter aircraft reported that the depression strengthened to a tropical storm [3]. Tropical storm Sandy continued to intensify as a middle to upper-level trough forced Sandy to accelerate north-northeastward. By October 24th, aircraft reconnaissance observed that Sandy had intensified into a category 1 hurricane [3]. Later that day, Sandy came ashore between Kingston and South Haven, Jamaica [3]. As Hurricane Sandy moved over the deep, warm waters of the Cayman Trench, it underwent a period of rapid intensification. Then as a major hurricane, Sandy made a second landfall over Cuba on October 25th [3].

After making landfall in Cuba, Sandy began to weaken as it turned to the northeast. This steering pattern drove Sandy through the Bahamas, which weakened the system to below hurricane strength. However, the system's radii had nearly doubled since it's landfall over Cuba; reconnaissance aircraft data determined the radius of maximum winds was over 100nmi [3]. This enabled the storm to benefit from the warm advection aloft. Therefore, by October 27th Sandy was able to regain hurricane intensity. Nevertheless, shortly after Sandy encountered a blocking pattern stationed over the North Atlantic, which prevented the storm from moving out to sea [3]. After spending days in favorable conditions and roughly 220nmi southeast of Atlantic City, Sandy re-intensified into a category 2 hurricane [3]. Late on October 29th, Sandy moved over cooler water and became extratropical by 2100 UTC [3]. Two hours later, the heart of post-tropical cyclone Sandy, made landfall near Brigantine, New Jersey [3]. After landfall, Superstorm Sandy moved through southern New Jersey, northern Delaware, and southern Pennsylvania. By October 31st, Sandy's structure started degenerating over northeastern Ohio. The remnants continued on to Ontario, Canada before integrating with an area of low pressure [3].

Superstorm Sandy came ashore as a large, extratropical cyclone, whose storm surge and winds caused destruction along the mid-Atlantic coastline. Although Superstorm Sandy did not make landfall as a hurricane, hurricane-force winds were felt along the coasts of New Jersey and Long Island, New York [3]. In the end, seven different states reported feeling hurricane strength wind gusts [3]. In addition to powerful winds, Sandy's pressure at landfall (945.5mb) set a record for the lowest sea-level

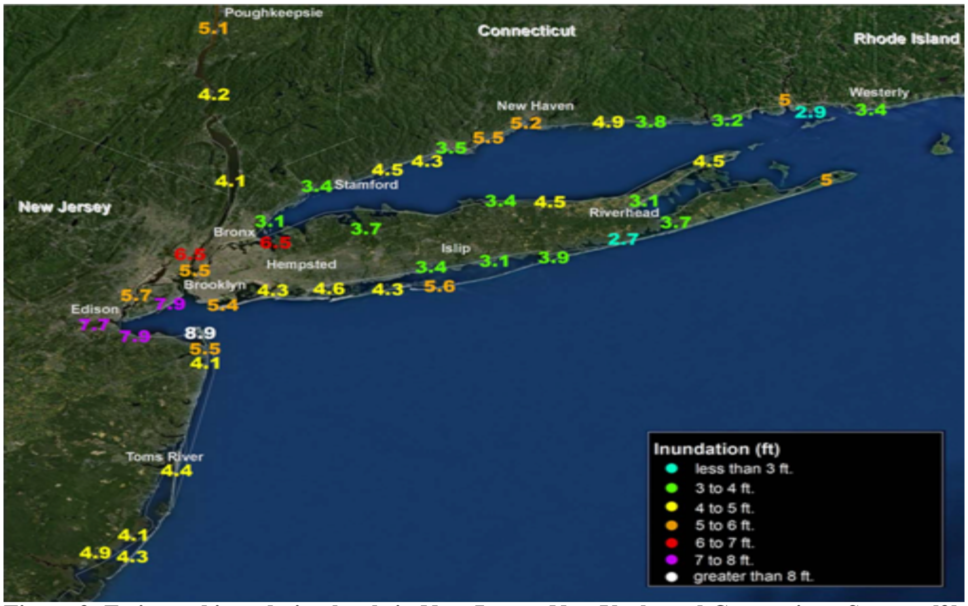

Figure 3: Estimated inundation levels in New Jersey, New York, and Connecticut. Source: [3]

pressure ever recorded north of North Carolina [3]. Nevertheless, these factors all contributed to Sandy's most destructive force—storm surge. The entire east coast from Florida to Maine saw a rise in water levels. Ultimately, New Jersey (8.57 ft), New York (12.65 ft), and Connecticut (9.83 ft) reported the highest storm surges [3]. The low-lying coastlines of New Jersey and New York, especially New York City, exposed essential infrastructures to the raw power of Sandy's surge and damaging waves. Record storm tides brought catastrophic flooding. The highest inundation felt in New York occurred in Staten Island and Manhattan with floodwaters 4-9ft above ground level (Figure 3) [3]. Similarly, in New Jersey the highest inundation was 4-9ft felt by Monmouth and Middlesex counties (Figure 3) [3]. More than 80% of Atlantic City was underwater [23]. Although rainfall did contribute to the extensive inundation across New Jersey and New York, the highest rainfall total (12.83 in) was recorded in Bellevue, Maryland [3]. Although very seldomly seen with tropical cyclones, Superstorm Sandy remarkably caused a widespread blizzard. The snowfall fell along the Appalachian Mountains from North Carolina to Pennsylvania. Wolf Laurel, North Carolina and Richwood, West Virginia recorded the highest snowfall totals of 36 inches [3].

At the time of its landfall, Superstorm Sandy was the second costliest ($71.4 billion) tropical cyclone to ever hit the U.S. [24]. Five states issued evacuations in preparation for Sandy. New York City mayor Michael Bloomberg ordered the mandatory evacuation of 375,000 residents [25]. In the immediate aftermath 23,000 people took refuge in temporary shelters [26]. At least 650,000 homes were either damaged or destroyed by the storm's surge [23]. New Jersey estimated a total of $8.3 billion in small business losses [3]. Approximately 8.5 million customers were without power for weeks. Lower Manhattan lost all power, leaving important infrastructures like hospitals and transportation hubs completely in the dark and closed [23]. Due to the extended power outages and cold weather about 50 people died from hypothermia, falls in the dark (by senior citizens), or carbon monoxide poisoning. Overall, Sandy was responsible for 147 deaths [3]. Seventy-two direct deaths were reported in the United States; forty-one of those deaths were attributed to storm surge [3].

Federally Declared Areas of Major Disaster

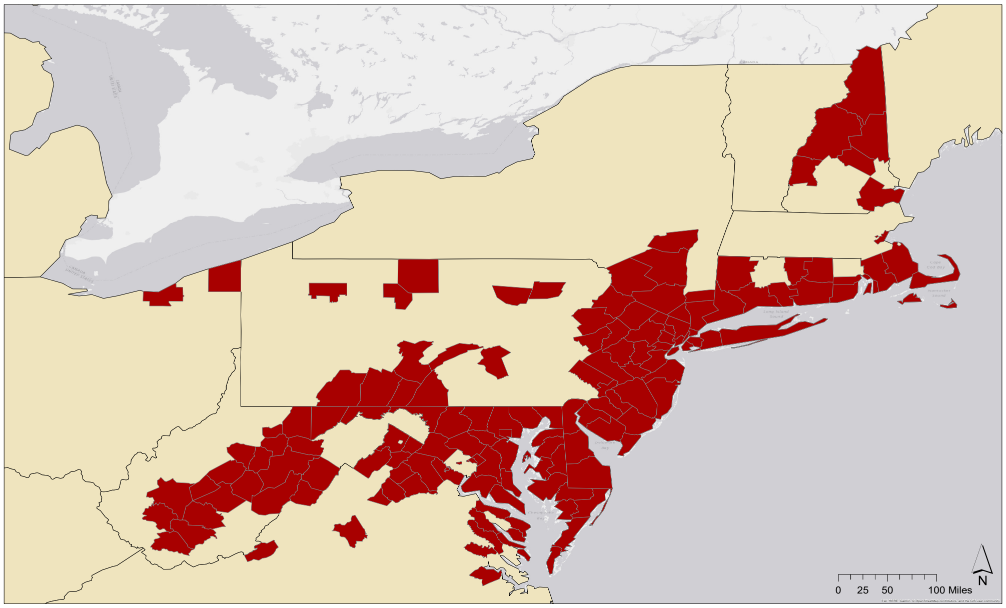

Figure 4: Highlighted counties (in red) were declared disaster areas in the wake of Superstorm Sandy

Superstorm Sandy caused damage along the U.S. eastern seaboard, however catastrophic destruction was seen throughout New York and New Jersey. The recovery efforts continue years after the storm. On October 28th, Disaster Declarations were declared for the states of Connecticut, District of Columbia (D.C.), Massachusetts, Maryland, New York, and New Jersey; by October 29th and 30th additional Disaster Declarations were declared for the states of Delaware, New Hampshire, Pennsylvania, Rhode Island, Virginia, and West Virginia. Additionally, over 12 states had counties which were declared areas of Major Disaster—Connecticut (7 counties), D.C., Delaware (3 counties), Massachusetts (6 counties), Maryland (24 counties), New Hampshire (6 counties), New Jersey (21 counties), New York (14 counties), Ohio (2 counties), Pennsylvania (18 counties), Rhode Island (4 counties), Virginia (28 counties), and West Virginia (18 counties) (Figure 4) [27].

2.3 Social Media and Geoinformatics

The growth of Web 2.0 has sparked an expansion of the use of social media and its wide-range of applications. Popular social media platforms, such as Facebook, Twitter, and Instagram, now allow users to easily add geographic information into their posts and pictures. Furthermore, users with location services turned on have their geographic coordinates embedded into everything they post on social media without having to manually add or tag locations. As a result, the growth and advancements of social media has widened the potential and capabilities of volunteered geographic information (VGI) [28]. Volunteer geographic information is a term used to describe the collection of widespread geographic information, which is provided voluntarily by private citizens. Recently, a focus has been placed on the VGI concept of sensor networks. One type of sensor network consists of humans and our ability to compile and interpret our surroundings while freely roaming the Earth [28]. In times of a natural disaster, volunteer geographic information with humans acting as sensors, can report conditions through electronic devices faster than satellites [28]. In the future, volunteer geographic information could possibly be the solution to overcoming common problems, such as inadequate imagery and dangerous weather conditions.

Volunteer geographic information provides researchers with a new source for geographic information. This new information source calls for innovative data mining techniques to enable researchers to explore data [29]. Popular VGI crowdsourcing platforms are OpenStreetMap and Wikimapia. The purpose of these platforms is for users to provide accurate geographic information that will be used for mapping purposes. Volunteered geographic information is sometimes the only available data source because access to geographic information in specific areas is a risk to national security [28]. However, volunteered geographic information provided by social media is published differently than other crowdsourcing platforms. Social media often contains geographic information without it being knowingly provided by an application. Geographic content is often embedded in the author's message, reply, picture, or links. Therefore, embedded geographic information in social media must be harvested and analyzed before it can be used [30]. This extracted information is an extension of the VGI concept called Ambient Geographic Information (AGI) [30].

A prominent example of an AGI data source is the social media platform, Twitter. Twitter is a popular micro-blogging web service, in which users can share "tweets" with other users [31]. In 2016, Twitter reported 313 million monthly active Twitter users [32]. Twitter users are provided with many different avenues to communicate with one another. Users interact by responding to each other in one of two ways: *mentions* or *replies* and *retweeting*. A tweet is considered a mention when it is written to address a specific user [31]. There are many different reasons cited by Twitter users for reasons why they retweet; one of the top reasons being to spread information to new audiences [33]. Therefore, breaking news tends to be often retweeted in the form of links to articles from media sources [33]; nearly 92% of retweets contain a URL or link [34]. The focus is on content for retweets, therefore the number of retweets can represent the content value of one's tweets [34].

2.4 Early Event Detection in Social Media

Current literature investigates social media's event detection ability. There exist two categories for event detection algorithms—feature-pivot and document-pivot methods. A feature-pivot method, which shows promising results in detecting events is the Event Detection with Clustering of Wavelet-based Signals (EDCoW) [35]. The EDCoW algorithm analyzes tweets published on Twitter to detect events. This approach requires each event of interest to have at least two words to describe the event [35]. This is practical because large events tend to be described by several different words by many people. The EDCoW algorithm applies clustering techniques to detect events [35]. Another approach to using a feature-pivot algorithm is to model event topics as "burst of activities", where certain features rise suddenly in frequency as the event is ongoing [36]. To detect events the algorithm uses an infinite-state automaton, in which events are modeled as state transitions [36]. Additional research has been done to model sets of words used during bursts of activity to apply to future events [37]. These studies advise that algorithms should model an individual word's appearance as a binomial distribution and then identify the burst of each word with a threshold-based heuristic [37]. A research study used the "burst of activity" feature-pivot algorithm to detect events from a major English newspaper from Hong Kong over a two-year period. The results showed that feature-pivot clustering approach was highly successful in detecting events from bursts of features (words) [37].

Event detection algorithms, such as the ones outlined above, have been used to detect real world events from Twitter activity. The main advantages of Twitter-based detection systems over sensor-based systems is that they are inexpensive, have a quick detection speed [38], and Twitter is widely used. The growing use of Twitter during disasters offers a new information source that could provide authorities with an enhanced emergency situation awareness [39]. Also, Twitter's ability to reflect a variety of events in short messages proves to be well suited as a source of real-time event content [40]. In order to identify events, recent research experiments have successfully used clustering techniques to group together similar tweets based on common features [40]. These clusters are then furthered studied to determine identifying features that can be used to train classifiers to distinguish future events [40]. In one study, a train classifier was overall successful in differentiating between real world event and non-event clusters [40].

Social media location services can supplement traditional geographic information sensors, such as seismic sensors and remote sensing, to produce a fuller, coherent situational report [38]. Yet, researchers have experimented with relying solely on Twitter as a sensor to detect real world events, such as earthquakes and hurricanes. One study, focusing on detecting earthquakes and typhoons in Japan, was able to produce a probabilistic spatiotemporal model. This early earthquake reporting system can detect an earthquake within the first minute of the initial shaking [41]. Therefore, the early detection system is able to issue alerts within the first minutes of an earthquake; this is roughly six times faster than traditional broadcasted alerts [41]. Using Twitter as an early event detector resulted in successfully detecting 96% of major Japanese earthquakes [41]. Real-time monitoring of social media is essential because early detection of seismic waves can be used to warn people further along the coastline, allowing them to get to safety in the case of a tsunami [38]. The spatiotemporal model also uses particle filtering to estimate the location of the earthquake's epicenter [41]. The same study also had success using particle filtering to estimate the trajectory of a typhoon [41].

Another option researchers are utilizing to examine natural disasters through a social media lens is automated social media analytics and mapping platforms. Social media crisis-mapping platforms map geographic data for areas at risk of natural disaster through geo-parsing real-time Twitter data streams [38]. Then they use statistical analysis to generate real-time crisis maps [38]. Geographic information collected from gazetteers, street maps, and VGI are compiled to make a crisis map [38]. An example of a social media analytic platform, which uses crisis mapping is Floodtags. Floodtags, filters, maps, and analyzes Twitter data. One research study examined Twitter activity during the 2015 Philippine floods and an additional 80 smaller floods in Pakistan [42]. By using the longitude and latitude coordinates stored in each tweet's metadata, Twitter acted as a crowdsourcing virtual sensor network [38]. Ultimately, the study found that during a local flooding event Twitter was able to detect flooding two days earlier than any official reports [42]. The research also discovered that while satellites were more appropriate for monitoring widespread flooding, Twitter is better skilled to monitor floods of any magnitude [42].

As a result, social media analytic platforms can offer quick, additional assistance to emergency services in the wake of a natural disaster. Utilizing the public's collective intelligence is especially useful during emergency incidents, in which people within blackout areas experience limited communication ability [39]. Utility companies could utilize social media analytic platforms to aid in mapping outage areas and possibly locating compromised electrical equipment by identifying the extent of inundation. After all, the purpose of a social-media assisted platform is to produce a coherent situation assessment picture, that could be presented to emergency and relief responders to help coordinate response efforts and improve overall situational awareness [38].

2.5 The U.S. Energy Grid

2.5.1 Regulatory Structure

The DOE outlines the regulatory structure of the United States' electricity structure in the Department of Energy's publication, United States Electricity Industry Primer [43]. The United States' electric grid is overseen by a major regulatory body, regional organizations, and utilities' ownership structures. Overall, the electricity sector is managed by the Federal Energy Regulatory Commission (FERC). The FERC is the independent agency within the U.S. Department of Energy, which regulates the interstate transmission of electricity within the United States [43]. Within the electricity sector, some of FERC’s responsibilities include regulating the transmission and wholesale of electricity in interstate commerce, reviewing electricity companies’ mergers and acquisitions, reviewing siting applications for electric transmission projects, protecting the reliability of the high voltage interstate transmission system, and overseeing financial reporting regulations and conduct of regulated companies [43]. In recent years, the Energy Policy Act of 2005, expanded FERC's power to implement regulations regarding the availability of reliable energy resources [43]. At the consumer level, the FERC is responsible for attaining reliable, efficient, and sustainable energy services at a reasonable cost [43].

In 2006, FERC designated the North American Electric Reliability Corporation (NERC) as the government’s electrical reliability organization (ERO) [43]. The NERC is an international regulatory authority, which ensures the reliability of the bulk power system in North America. As an ERO, NERC has jurisdiction over electric users, owners, and operators of the bulk power system [43]. The NERC is also authorized to enforce and develop reliability standards, monitor the bulk power system, assess seasonal and long-term reliability, and train and certify personnel [43]. North America’s bulk power system consists of four distinct power grids called interconnections. The NERC’s area of responsibility spans across the continental United States, Canada, and the northern portion of Baja California, Mexico. The four interconnections are the Western

Interconnection, ERCOT Interconnection, Eastern Interconnection, and Quebec Interconnection. Interconnections are zones, which utilities are electrically tied together during usual system conditions (Figure 5) [43]. Each interconnection operates independently of one another and strive to operate at a synchronized average frequency.

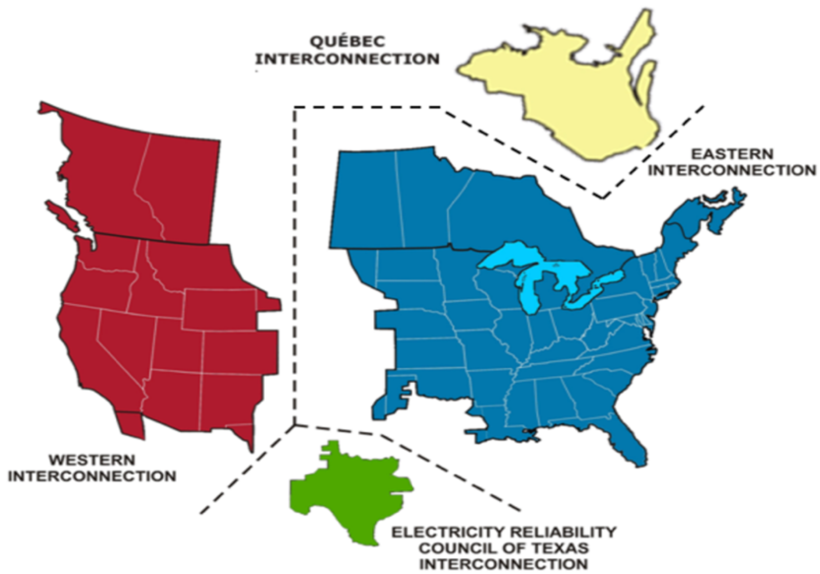

Figure 5: The North American Power Grid Interconnections. Source: [43]

The FERC further defines areas within NERC regions as regional transmission organizations (RTO) and independent system operators (ISO) [43]. The formation of RTOs and ISOs comes at the recommendation of the FERC. There are seven ISOs and four RTOs in North America (Figure 6). The seven ISOs are California ISO (CAISO), New York ISO (NYISO), Electric Reliability Council of Texas (ERCOT), Midcontinent Independent System Operator (MISO), ISO New England (ISONE), Alberta Electric System Operator (AESO), and Independent Electricity System Operator (IESO) [43]. The

Electric Reliability Council of Texas is an ISO and NERC region. The four RTOs are PJM Interconnection (PJM), Midcontinent Independent System Operator (MISO), Southwest Power Pool (SPP), and ISO New England (ISONE) [43]. Midcontinent Independent System Operator and ISO New England are also ISOs.

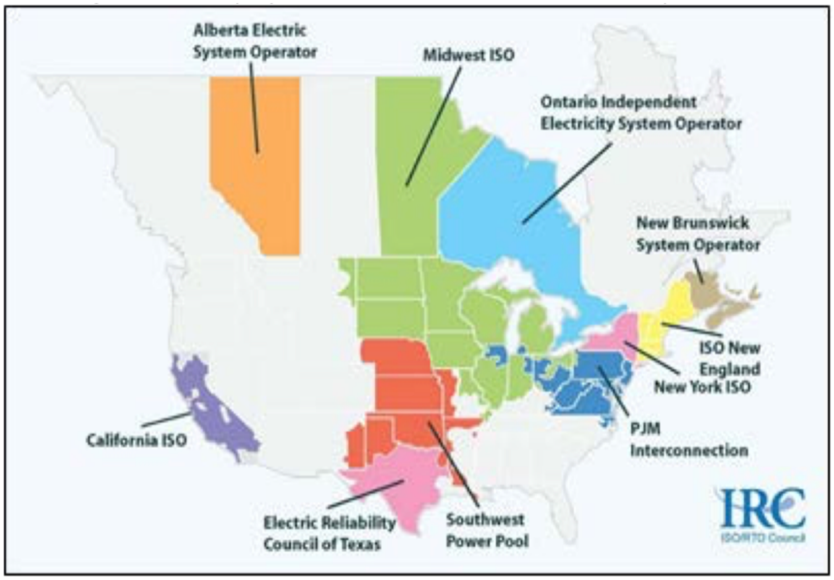

Figure 6: The North American Transmission Operators. Source: [43]

According to the Department of Energy, ISOs operate the area's electricity grid, administers the region's wholesale electricity markets, and provides reliability planning for the territory's bulk electricity system. RTOs are similar to ISOs; however, they have greater responsibility for the transmission network established by the FERC [43]. In addition, RTOs coordinate and control the functions of the electric system within their region and monitor their transmission network by providing fair transmission access [43]. Both, ISOs and RTOs, are involved in regional planning to ensure the needs of the electric system are met with the appropriate infrastructure. Before RTOs and ISOs were created, power and utility companies were responsible for developing and coordinating transmission plans [43]. This still holds true for regions without ISO and RTO coverage.

2.5.2 Delivery

According to the Department of Energy, the structure of electricity delivery can be categorized into three functions: generation, transmission, and distribution. Each function is linked by key assets known as substations (Figure 7). The U.S. energy generation begins in a power plant. A power plant often has one or more generators, which sometimes can be used across different types of fuel [43]. The United States' energy supply is generated by a diverse mixture of fuels. In 2014, the Department of Energy reported there were 19,023 commercial generators at 6,997 operational power plants in the United States [43]. Electricity is generated when mechanical energy is transformed into electrical power. Once electricity is generated "step-up substations" are used to increase power generation voltage to the transmission system level [43].

Figure 7: Flow chart of the delivery of electricity. Source: [43]

At the transmission system level, electric power travels more than 360,000 miles of transmission lines connecting to approximately 7,000 power plants [43]. Of those

360,000 miles of transmission lines, there are 180,000 miles of high-voltage transmission lines [43]. Transmission line networks are designed to carry energy over long distances with minimal power loss. This is possible by boosting voltages at certain points along the electricity supply chain [43]. RTOs and ISOs routinely regulate transmission systems. Once power has reached a load center, a "step-down substation" uses transformers to decrease the voltage to a medium range for major distribution networks [43]. This is done because it is cost efficient to transmit on a sub-transmission network at a voltage level between standard transmission and distribution voltages [43]. Transformers are crucial substation equipment for delivering electricity to customers. However, many transformers are located in remote regions and are vulnerable to weather events, acts of terrorism, and sabotage [43]. The loss of transformers at substations raises a momentous concern for energy security in the electricity supply chain due to shortages in inventory and manufacturing materials, increased global demand, and limited domestic manufacturing capabilities [43].

The low voltage can now be carried over distribution power lines to commercial and residential customers. At the distribution system level, power generation voltage is further reduced by a distribution substation [43]. Distribution transformers are the cylindrical devices, which can be seen mounted on local power lines or on concrete pads in neighborhoods; distribution transformers can also be located underground [43]. After the electricity reaches a distribution substation it is finally deliverable to the customer.

2.5.3 Storm Preparedness and Restoration

Even though power infrastructure is highly redundant and resilient, customer outages do occur as a result of system disruptions. The most common disruptions are due to weather-related events [43]. The energy sector actively prepares, plans, and anticipates severe weather events. One of the ways the energy sector is regularly preparing its infrastructure is called "hardening" [43]. Hardening is the process of physically changing a utility's infrastructure to mitigate the effects of storm damage [43]. The goal of hardening is to increase the durability and dependability of transmission and distribution assets. One hardening method is undergrounding, which is the act of burying transmission and distribution lines underground [43]. Although undergrounding protects electrical lines from severe incidents above ground, the trade-offs are that it is expensive, difficult to repair, and more susceptible to flooding [43].

Newer hardening methods, such as the smart grid and microgrid, are attributed to technological advancements [43]. The Department of Energy says smart grids are slowly becoming implemented by utilities, which allows for utilities to quickly identify outage areas. Smart Grid receives information from electricity distribution systems, enabling it to be able to detect problems and potentially re-route power while alerting system operators to the location of the issue [44]. Smart Grids allow for utilities to use their resources and personnel more efficiently by eliminating the need to send out crews just to identify a problem area. A less common hardening technique, yet effective, is the microgrid. A microgrid is described as an isolated "island" of electricity generation, transmission, and distribution [43]. In the event of a service interruption, microgrids are able to disconnect from the main grid and operate independently [43].

Once a storm is forecasted to strike an area, utility companies pre-position mobilized crews and restoration resources [43]. Mobile command centers are set up to serve as a hub for communication and coordination for restoration efforts. Mutual assistance—agreement between neighboring utilities to assist one another when outages occur—is enacted [43]. Once the storm passes, utility companies can begin their restoration process. The restoration process begins with utilities dispatching crews to access the damage to power lines and substations [43]. However, locations where Smart Grids are installed, eliminate this step of the restoration process. Also, at this point utility customers can report outages to their local electric companies [43]. This helps utility restoration teams to know where they need to direct their crews and resources.

Next, restoration crews need to alleviate hazardous conditions. The Department of Energy states that crews prioritize repairs to equipment that poses a threat to the public. At this stage, down and damaged lines and substations are turned off. Starting at the base of the electricity delivery process, crews first begin repairing damaged power plants, then move on to repairs to the high-voltage transmission lines and finally the substations [43]. At this point, power is attempted to be restored to critical public safety and health infrastructure, such as emergency responders [43]. Subsequently, repairs continue on distribution lines and substations to restore power to businesses and residential neighborhoods [43]. Delays in this process are often attributed to equipment that has endured significant damage or lingering storm impacts, such as flooding.

2.3.4 Impacts

In addition to periods of blackouts, electricity outages have propagating effects because the energy infrastructure is interdependent upon its entities [43]. This mutual dependence can make power restoration difficult. For example, the creation and delivery of oil and gas heavily relies on the supply of electricity. Nonetheless, the generation of electricity requires the continuous supply of resources, such as natural gas, coal, and oil [43]. Furthermore, critical infrastructures, such as water treatment facilities, pumping stations, gas stations, communications systems, and natural and petroleum pipelines, depend on a consistent supply of electricity [43].

2.6 Advancements in Power Outage Detections

The United States' energy infrastructure which serves the power grid is aging. As older systems and components are retired, they are replaced with newer components often linked to communications or automated systems [44]. As previously mentioned, one improvement modernizing the electrical grid is the Smart Grid [44]. Smart Grid is an evolving electric power network that has the ability to aid in avoiding extensive power outages. Another modernization for detecting power outages is patented by Verizon. A network device receives a loss of power alarm from a network interface device (NID) associated with a customer premise [45]. The loss of power alarm includes a particular

NID identifier that is associated with a specific address and identifies a power outage in a particular region associated with the address [45]. The systems then determine if other loss of power alarms have been received from the same area [45]. However, the privacy of customer information has been raised as an issue with smart meters [44].

The electricity industry is beginning to look to social media to assist in detecting and locating areas experiencing power outages. Social media users can provide nearly real-time observations in disaster zones [39]. Tweets can use embedded links or photos to offer more information. This is particularly useful during emergency situations, when people within blackout areas experience limited communications methods [39]. One company called DataCapable is already exploring this idea. Recently, DataCapable has created a new outage reporting tool in the form of an app. DataCapable’s *UtiliSocial* portal gathers posts from social media and the DataCapable app and maps it out. The map is then ready to be used by a utility's Outage Management System (OMS) tools to direct their repair teams to outage areas [46]. This technology could prove to be essential during natural disasters.

3 OBJECTIVES AND RESEARCH QUESTIONS

The incorporation of social media into our daily lives, now offers assistance in a variety of diverse disciplines. In extreme weather events, such as hurricanes, dangerous conditions can develop suddenly. Strong tropical force wind bands can knock down powerlines and damage utility boxes. Flash floods can develop in a matter of minutes, which makes repairing down electrical lines an extremely dangerous task during the aftermath of a storm. Also, storm surge can shut down entire electrical plants resulting in complete power grids shutting off. Overwhelmed, clogged cell towers coupled with no power can make identifying problem areas and prioritizing repairs difficult. Thus, additional resources such as social media, could help locate the extent of areas without power. When disastrous weather strikes, Twitter users can act as “eyewitnesses” by updating other users of developing conditions. For example, Twitter users can report power outages in areas, some so remote that otherwise wouldn’t be officially reported. The VGI concept of using humans as geospatial sensors is described as a citizen science [28]. Citizen science is a term used to describe networks of people who serve as observers in some scientific field of study [28]. This research hinges on citizen science and its nontraditional approach to assist in detecting disturbances in the energy infrastructure.

Power outages, which were a consequence of Hurricane Harvey and Superstorm Sandy, will be the center focus of this research. Twitter activity during Harvey and Sandy, service interruption datasets and shapefiles of the energy grid from the U.S. Energy Information Administration (EIA), reported customer outages from the U.S. Department of Energy (DOE), census data from the U.S. Census Bureau and information about Harvey and Sandy's aftermath will be used to answer this study's research questions. This research will explore whether power outages due to a severe weather event can be identified using Twitter. This study will start by assessing the textual content of collected tweets. A geosocial analysis of the Twitter network will provide insight on the size and range of the social network. An investigation will explore the most commonly used words and hashtags to determine which events and topics widely engaged the Twitter network. The messages in tweets can also divulge information about the locations of life-threatening aftermath. This evaluation can reveal whether or not a linkage between social media activity and power outages exists and if it can be later used to identify locations without power. Therefore, a geosocial analysis of the population captured from social media will allow us to answer the first research question:
(1) Does the textual content of Twitter reveal an association between disturbances in critical infrastructures and social media usage?

Next, this research will examine when power outages were felt according to Twitter. Based on Twitter activity, a temporal analysis can be used to further study a timeline of when customers started to experience power grid blackouts. This timeline will be used to compare to a documented timelines of statewide customer power outages. A spatiotemporal analysis can tell whether or not social media activity could act as an indicator of a loss of electricity. Also, this analysis has the ability to reveal if a spike in social media activity could indicate a power outage prior to traditional reports by utility and power companies. Therefore, a spatiotemporal analysis will allow us to answer the second research question in this study:

(2) Does Twitter capture when customers lost power due to Hurricane Harvey and Superstorm Sandy?

Lastly, a geospatial analysis will be performed to determine whether increased Twitter activity can be used as an indicator of power outages. Also, the analysis can disclose which power and utility providers were heavily affected. The results from the geospatial analysis will provide a visual, which compares the Twitter activity and reported power outages within each utility service area relative to one another. This analysis can be used by utility companies to aid in future storm preparations. This geospatial analysis will allow us to answer the third and final set of research questions of this study:

(3) Can the volume of Twitter activity serve as an indicator for the location of an active power outage? Is there a correlation between the volume of Twitter activity and reported power outages?

4 DATA STRUCTURES AND CHARACTERISTICS

The following chapter provides a thorough overview of the datasets used in this study. The data structures and storage are the primary focuses in the 4.1.1, 4.1.2, and 4.1.3 subsections. Section 4.2 of this chapter discusses the characteristics of the datasets.

4.1 Data Structures

4.1.1 Twitter Datasets

The primary datasets used in the Harvey analysis are Twitter datasets collected worldwide between August 24th - September 2nd, 2017. The first dataset was collected using the keywords *flood*, *flashflood*, and *flash flood*. The second dataset was collected using the keywords *hurricane*, *tornado*, *storm*, *thunderstorm*, *cyclone*, *derecho*, *heat*, *heatwave*, *draught*, *typhoon*, *ice*, *snow*, *snowstorm*, *blizzard*, and *Harvey*. The datasets were collected and formatted into tab-separated files. The datasets were received from George Mason University's Center for Geospatial Intelligence after initial pre-processing. Additional pre-processing was necessary to preserve special characters, correctly encode, and to obtain only the geo-located tweets. This pre-processing step was completed by running the dataset through a python script.

The primary datasets used in the Sandy analysis are Twitter datasets collected worldwide between October 29th and October 30th, 2012 using the keywords *hurricane* and *Sandy*. The dataset was collected and formatted into a tab-separated file. The dataset was also received from George Mason University's Center for Geospatial Intelligence after initial pre-processing. Additional pre-processing was necessary to preserve special characters, correctly encode, and to obtain only the geo-located tweets. This pre-processing step was completed by running the dataset through a python script.

ArcGIS 10.6 software stored the datasets and organized each row and column by tweets and their nineteen attributes. Each row represents a single tweet and each column contains an attribute for that tweet. The nineteen attributes include ID, location, country, state, zip, longitude, latitude, publication time, author, coordinates from, retweeted ID, retweet author's user ID, retweet author's user name, quoting, response ID, response author, language, text, and extracted links. The ID attribute is a tweet identifier assigned by Twitter. The next six attributes offer information about geo-tagged tweets. The location, country, state, and zip are the location/country/state/zip from which the tweet is sent. The longitude (x) is the longitude of the geolocated tweet and the latitude (y) is the latitude of the geolocated tweet. Each tweet's timestamp, the date and time it was published, is stored in the *published_at* attribute. The author's twitter handle and name are collected and saved in the author column. The *coords_from* attribute offers information about the origin of the geospatial information of the tweet.

The next four attributes offer information about retweets. The retweeted ID field holds the unique Twitter assigned identifier of the retweeted tweet. The retweet author's user ID, is the numeric ID of the user who sent the original tweet. The *retweeted_uname* attribute is the username of the author of the original tweet. The quote field is a collection of 1's and 0's—a numeral 1 is present if the retweeted tweet quotes the original tweet and a numeral 0 if the retweet is not quoted. The next two attributes contain information about tweets that are responses to another tweet. The *response_id* attribute includes the ID of the tweet to which the current tweet responds and the *response_author* attribute is the author's Twitter handle of the tweet to which the current tweet responds. The last three fields indicate information about the tweet's message. The *lang* attribute holds information about the tweet's language declared by the author's profile. Lastly, the text field is the text of the tweet and the extracted links column contains links embedded within the tweet.

4.1.2 Energy Datasets

4.1.2.1 U.S. Energy Information Administration Data

The U.S. Energy Information Administration datasets were retrieved from the EIA's *Electric Power Monthly* reports. Within these reports there are tables published that capture all reports of disturbances in electric services during the year. The tables retrieved are titled, "Major Disturbances and Unusual Occurrences, Year-to-Date 2012" [9] and "Major Disturbances and Unusual Occurrences, Year-to-Date 2017" [7]. The datasets were downloaded and formatted into tab-separated files. Tableau stored these datasets and organized each row and column by each report and its eleven attributes. Each row represents a reported disturbance in service and each column contains an attribute for that event. The eleven attributes include year, month, event and time, restoration date and time, outage duration, utility/power pool, the NERC region, area affected, type of disturbance, megawatts lost, and number of customers affected.

The first five attributes offer information about the event's outage date, restoration date, and the duration of the service interruption. The next attribute identifies the utility company/power pools, which experienced the power failure event. The succeeding two attributes give geographic information about where the electricity outage transpired. The NERC column signifies the regional council that suffered the outage. The following are the eight different councils represented in NERC: Florida Reliability Coordinating Council (FRCC), Midwest Reliability Organization (MRO), Northeast Power Coordinating Council (NPCC), ReliabilityFirst Corporation (RFC), Southeastern Electric Reliability Council (SERC), Southwest Power Pool (SPP), Texas Regional Entity (TRE), and Western Energy Coordinating Council (WECC). The area affected column labels the state and/or the counties where the outages were felt. The succeeding column, type of disturbance, offers information about the cause for the loss of power. Lastly, the last two attributes offer numerical information about the loss of power (in megawatts) and the number of customers affected.

4.1.2.2 Department of Energy Data

During a natural disaster, the Department of Energy publishes situation reports documenting general overviews and impacts on the electricity, oil, and natural gas sectors. During Hurricane Harvey and Superstorm Sandy's immediate aftermaths, the report updates were published bi-daily. As restoration efforts continued for days and weeks later, situation reports were only published daily. The reports are published and formatted in a text portable document format (PDF). Each report contained important energy information such as, executive summaries, electricity sector impacts, oil and natural gas sector impacts, an incident overview, and electricity outage data. Overall, there were 21 situation reports documenting electricity outages for Hurricane Harvey [8] and 19 situation reports detailing electricity outages for Superstorm Sandy [10]. The electricity outage tables were exported and stored in separate excel sheets.

Electricity outages due to Hurricane Harvey were documented bi-daily starting from August 26th until September 3rd and daily reports from September 4th through September 6th [8]. Each row in the outages table represented a state and the columns held their attributes. The first column identifies the states that faced power outages due to Harvey. The next column stores the current number of confirmed customers experiencing outages. The third column contains the percentage of confirmed customers without power for each state. The last column has the peak number of customer outages within a 24-hour period. Within the text of the situation reports are details about each utility and power companies' outages and restoration efforts. Information regarding the date, time, utility company, and number of outages was used to create an additional dataset for Hurricane Harvey's geospatial analysis. The DOE's situation reports provide more information about the affected utility companies than the datasets provided by EIA.

Power outages due to Superstorm Sandy were also recorded and stored in a similar style. Energy situation reports were recorded bi-daily beginning October 29th through November 6th [10]. One additional report was published on November 7th [10]. Identical to Harvey's tables, rows in Sandy's customer outage tables also denoted a state and the columns held their attributes. The first column identified the impacted states. The following column held the current number of customers outages in each state. The third attribute stored the percentage of customers without power. The next column contains the peak number of reported outages and the last attribute reveals the number of customers whose power has been restored since the peak.

4.1.2.3 The United States Energy Grid Data

The United States' energy grid structure is represented by map layers in a shapefile format. The energy grid is recreated structurally starting from the local level and working up to the national level. Vector data of transmission lines, substations, utility service territories, and NERC regions are used in this analysis to reconstruct the electric grid. These shapefiles are retrieved from the EIA's maps layer information page [47] and are stored in ArcGIS 10.6 software.

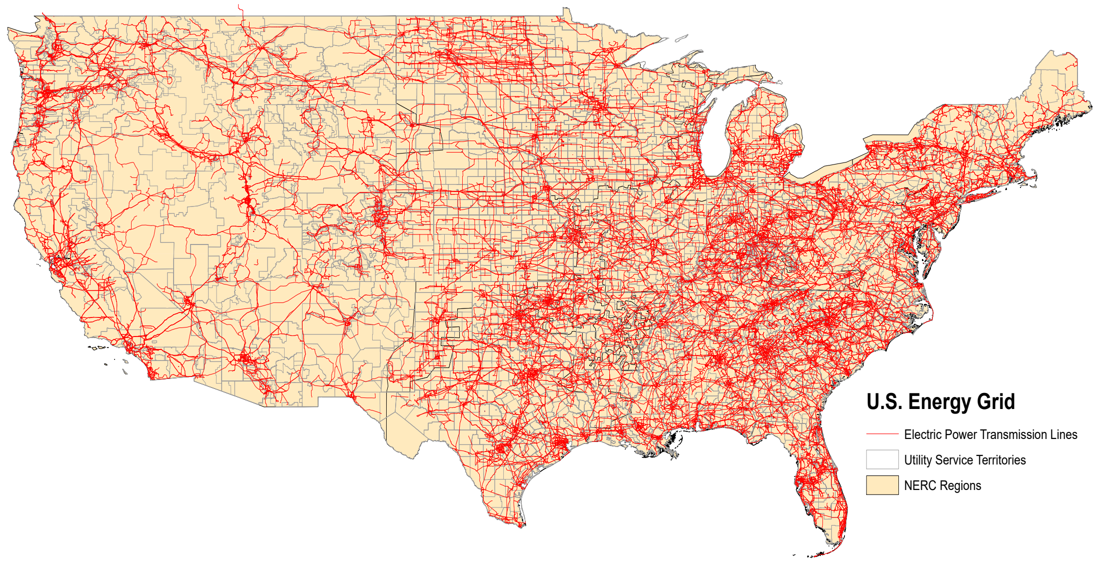

Figure 8: U.S. Energy Grid recreated in ArcGIS

The transmission lines feature class represents the system of structures, wires, insulators and associated hardware, which carry electric energy from one point to another. While the substations feature class includes all locations where power on a transmission line is tapped by another transmission line. This feature class contains transmission lines that are above and underground. The transmission lines and substations shapefiles' source are the Department of Homeland Security's Geospatial Platform. The transmission lines are polyline features, whereas the substations are point features (Figure 8). The transmission lines feature class contains 56,909 objects that have 13 attributes. Each transmission line has a unique identifier, *ID*. The rest of the attributes offer information about the transmission lines, such as the owner, operating status, the voltage (in kV), the voltage class, beginning and ending substations, the source, and the review date of the geographic placement. The substation feature class has 66,617 objects with 17 attributes. Each substation also has a unique identifier, *ID*. The other attributes include the substation's name, city, state, zip code, country, number of incoming and outgoing transmission lines, the maximum and minimum voltage, the source method, latitude, longitude, and the source.

The utility service territories are a polygon feature class, which represents all the electric power retail service territories in the United States. These electric power utilities service areas are responsible for the retail sale of electric power to local customers. The utility service territories feature class' source is also the Department of Homeland Security's Geospatial Platform. The utility service territories feature class contains 2,873 polygonal features with 33 attributes (Figure 8). Each utility service territory has a unique identifier, *ID*. Also recorded for each service territory is the name, phone number, address, city, state, zip code, country, website, owner type, regulated status, holding company, the RTO/ISO control area, electric power generation peaks and caps, net generation, number of customers, and the source.

The NERC regions feature class represents all the NERC regions in the United States. The polygon feature class source is the EIA's approximation based on a NERC map. The NERC feature class contains nine polygon objects with two attributes (Figure 8). The two attributes are the NERC region abbreviation and the full NERC names. The feature class contains an object for uncategorized NERC memberships, as well as the eight known NERC regions—FRCC, MRO, NPCC, RFC, SERC, SPP, TRE, and WECC.

4.1.3 U.S. Census Data

The United States census block data was retrieved from the United States' Census Bureau's Tiger/Line Download Center [48]. The Tiger/Line Download Center has available downloads for both the 2010 Census and the 2016 American Community Survey (ACS). The 2010 census block group housing characteristic table and census block group shapefile were retrieved for Superstorm Sandy. Whereas, the 2016 ACS census block group housing characteristic table and census block group shapefile were downloaded for Hurricane Harvey's case study. Within each housing characteristic table, the column of interest was the "Total Households" column, which contains the number of households within each census block group. Block groups were chosen because they are the smallest census geometry and therefore will provide more precision in calculations of the population of interest.

4.2 Data Characteristics

4.2.1 Hurricane Harvey Tweet Distribution

The Hurricane Harvey flood dataset contains 1,348,585 geographically located tweets out of a total of 2,366,531 tweets. From the original dataset, 56.9% of tweets are geo-located. The second dataset, Harvey's weather dataset contains 7,769,176 geotagged tweets, which is 54.1% of the original 14,352,975 tweets. The two datasets are combined together to give a more accurate overview of the impacted population. Once combined and all duplicate tweets are removed, the Hurricane Harvey Twitter dataset contains

8,588,000 geotagged tweets. Hurricane Harvey's geotagged tweets contain precise and imprecise geographic coordinates. Precisely geotagged tweets use coordinates obtained from the user's GPS or computed through cell tower triangulation. Imprecisely geotagged tweets estimate the user's geographic coordinates from either their location listed in their Twitter profile or it is inferred from the user's IP address. The geotagged dataset contains 0.69% precise geotagged tweets and 99.31% imprecise geotagged tweets.

As expected, the distribution of the geotagged tweets appears to mirror the United States' population distribution. The datasets are densely clustered around every metropolitan area. There are some compact clusters that extend alongside the Pacific coast of Washington, Oregon, and California and there are elongated clusters along the windward side of the Cascade and Sierra Nevada mountain ranges. Moving easterly, the distribution becomes sparse in the Rocky Mountain and Great Plains states. From east of the Great Plains to the east coast, the distribution of tweets appears densely packed. The majority of the geotagged tweets appear tightly clustered around the Texas cities of Houston and Dallas. The highest concentrations of tweets extend between the Texas cities of Dallas, Austin, San Antonio, Galveston, and Houston. There is also a compact cluster which extends along the Interstate-95 corridor from Boston, Massachusetts to Washington D.C. The datasets also have a world wide geographic spread. Internationally, the data is sparsely distributed with small clusters around capital cities.

Data collection begins one day before Hurricane Harvey makes landfall in southern Texas. Hurricane Harvey's Twitter activity steadily increases as the hurricane makes landfall. Over the next 24 hours Hurricane Harvey begins to gradually move towards southeastern Texas, where it becomes stalled over the city of Houston. During this time period there appears to be a sharp, rapid increase in Twitter activity. Between August 25^{th} and August 26^{th}, the data exhibits a 40% increase in Twitter activity. The dataset experiences its third largest peak (67,359 tweets) during this time period on August 26^{th}. Later during August 26^{th}, activity levels out around 45,000 tweets hourly. By morning on August 27^{th}, activity sharply increases to an absolute maxima of 67,823 tweets and is shortly followed by another peak of 67,741 tweets two hours later. From these peaks, activity slowly, cyclical decreases and reaches an absolute minima of 5,355 tweets on August 29^{th}. By August 30^{th}, Twitter activity seems to follow a daily diurnal cycle which gradually decreases daily for the rest of the collection period.

4.2.2 Superstorm Sandy Tweet Distribution

The Superstorm Sandy dataset is comprised of 2,268,692 tweets. There are 1,470,864 geotagged tweets, which is roughly 65% of the original dataset. Superstorm Sandy's geotagged tweets also contain precise and imprecise geographic coordinates. Precisely geotagged tweets use coordinates obtained from the user's GPS or computed through cell tower triangulation. While, imprecisely geotagged tweets estimate the user's coordinates from their location listed in their Twitter profile. The geotagged dataset contains 1.55% of precise geotagged tweets and 98.45% of imprecise geotagged tweets.

The Sandy dataset also follows the United States' population distribution and has highly concentrated clusters around every major U.S. city. There are looser, elongated clusters in California along the Pacific coast, the Sierra Nevada range, and southern California. Looking easterly, the Rocky Mountain and Great Plains states are scarcely distributed with tweets. The majority of the geotagged tweets are located in the eastern half of the continental United States. The cities of Boston, New York City, Philadelphia, Baltimore, Washington D.C., Pittsburg, Detroit, and Chicago are heavily clustered with tweets. Also, Long Island and the coastline of New Jersey are moderately dispersed with tweets. The highest concentration of tweets extends along the Interstate-95 corridor from Boston, Massachusetts to Washington D.C. Similar to the Hurricane Harvey Twitter distribution, the geographic spread is not just nationally, it is worldwide. Outside of the United States the dataset is moderately clustered around capital cities.

The data collection starts at 2300 UTC on October 29th which corresponds to Sandy's landfall. At the time of landfall, the volume of Twitter activity abruptly increases. The maxima in the volume of tweets occurs on October 30th with 94,273 tweets. The volume of activity remains constant with 90,000 tweets for the next four hours before the volume plunges. The volume continues to decrease until it reaches a minimum with 59,044 tweets. Then in the morning hours of October 30th, the Twitter activity rises for the next 12 hours. Four hours later, Twitter activity levels out before it sharply declines until the data collection ends.

5 METHODOLOGY

The following chapter provides a thorough overview of the methodological approaches used in this study (Figure 9). The processing methods and analytical techniques used to answer the research questions introduced in Chapter 3 are addressed in this chapter. The sections of 5.1, 5.2, and 5.3 of this chapter discusses the software, scripts, procedures, and methods used in each of the analyses of the datasets.

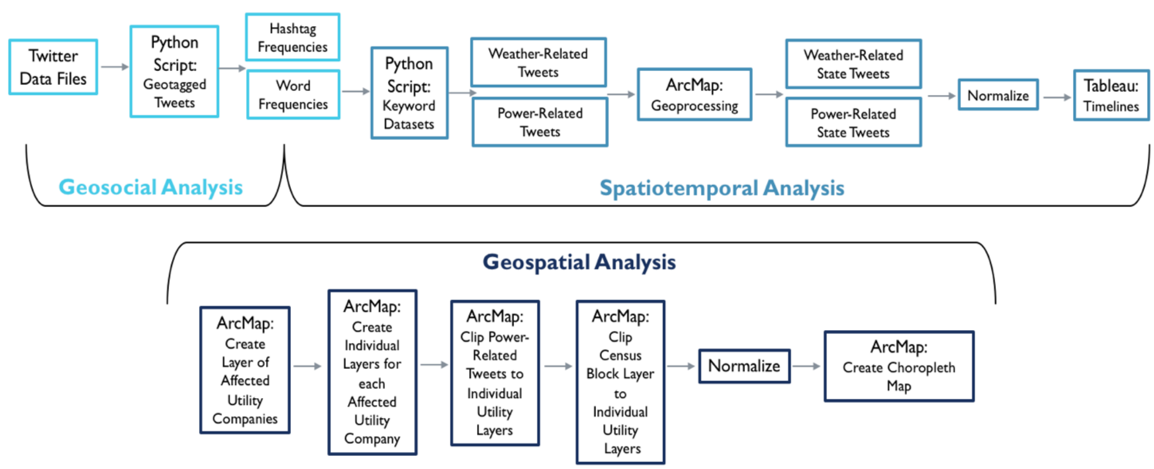

Figure 9: Flow diagram of the methodologies used to answer this study's research questions.

5.1 Geosocial Analysis

In order to examine the content of tweets and to better understand the network of the Twitter datasets, a geosocial analysis was conducted. The investigation of the Twitter network begun by finding the most commonly tweeted words and hashtags during the collection period. A python script was used to filter through tweets to get a frequency count for each tweeted word. The word frequency script received each Twitter text file as the input; the output is an excel spreadsheet. The spreadsheet contains the word counts for only tweets that are geotagged. Within the spreadsheet, each row holds a word and its frequency count. The Superstorm Sandy and Hurricane Harvey Twitter dataset were individually run through the word frequency script and have separate spreadsheet outputs. The output spreadsheets were sorted in descending order and further filtered for stop words. Stop words are common words that when taken out of context don't particularly convey a certain meaning. Of the remaining words, the twenty most frequently used words were selected and organized into a table.

A similar method was used to determine the most frequently used hashtags during the collection periods. A hashtag frequency script took each Twitter dataset as its input and the output was an excel spreadsheet which contained the hashtags from geotagged tweets in one column and their frequency counts in the corresponding column. The Hurricane Harvey and Superstorm Sandy Twitter datasets were individually run through the hashtag frequency script and each have separate spreadsheet outputs. The output spreadsheets sorted frequency counts in a descending order. Then the top twenty hashtags were selected and organized into a table. The frequency output spreadsheets were further searched for other interesting hashtag usage, which could reveal more information about the social network at the time of the storms.

5.2 Spatiotemporal Analysis

In order to temporally analyze the volume of Twitter activity during the reported outages, the initial task was to create a subset of tweets discussing power outages. Before a new dataset can be constructed, the word usage from each dataset is studied. This step makes sure the words chosen to depict power outages, accurately represents the language used on Twitter. The data processing used in this spatiotemporal analysis utilizes the word frequency outputs from the previous section. For each output dataset, the frequencies are sorted in decreasing order and stop words were filtered. Then each output spreadsheet is thoroughly searched to determine the most commonly used words that pertain to power outages. The keywords chosen from the Hurricane Harvey dataset are the following: *gas*, *power*, *plant*, *energy*, *infrastructure*, *dark*, *outages*, *lights*, and *electricity*. Additionally, the keywords selected for the Superstorm Sandy Twitter dataset are the following: *power*, *dark*, *lights*, *darkness*, *outages*, *blackout*, *outage*, and *electricity*. These words were considered possible indications of a loss of electricity.

Once the power outage keywords are identified, the new data subsets could be created. A python script aided in filtering each collected tweet's text field word-by-word to determine if it matched one of the power keywords. If the tweet's text contained a keyword and was geographically located, that tweet's ID, latitude, longitude, published date/time, and text fields are rewritten to a new comma-separated values (CSV) file. This process is completed for the Hurricane Harvey and Superstorm Sandy Twitter datasets.

This study also examines whether weather events can serve as possible secondary indicator of power outages. Therefore, a set of weather-related keywords captured by the Twitter network needs to be identified. Once again, the word frequency outputs from the geosocial analysis are used to search thoroughly for extreme weather incidents that could be responsible for damage to electrical equipment and ultimately cause outages. The same method to determine the power outage keywords is used to find the keywords for weather incidents. The keywords chosen from the Hurricane Harvey dataset are the following: *flood*, *flooding*, *floods*, *flooded*, *waters*, *water*, *floodwaters*, *surge*, *tornado*, and *underwater*. The chosen Superstorm Sandy keywords are: *water*, *flooding*, *flooded*, *Frankenstorm*, *snow*, *flood*, *floods*, and *surge*. After the weather incident keywords are selected, a python script is used again to filter through each tweet's text field to determine if it contains any of the weather keywords. If the tweet's text contained a keyword and is geotagged, that tweet's ID, latitude, longitude, published date/time, and text fields are rewritten to a new CSV file. This process is completed for the Superstorm Sandy and Hurricane Harvey Twitter datasets.

These newly created power outage and weather incident datasets are now ready for further refinement. The keyword datasets are loaded into ArcMap and individually clipped to each state that experienced an outage of power due to Hurricane Harvey or Superstorm Sandy. The states of Texas and Louisiana are used for the Harvey case study and the states of Connecticut, Delaware, District of Columbia, Illinois, Indiana, Kentucky, Maine, Maryland, Massachusetts, Michigan, New Hampshire, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Rhode Island, Tennessee, Vermont, Virginia, and West Virginia are used for the Sandy case study. After the clipping geoprocess is completed for each state, each clipped layer's attribute table is exported as a new text file. After this step, the number of power-related and weather-related tweets are revealed for each state. The statewide number of tweets are normalized with each state's total number of households. The normalization allows for an accurate account of tweet volume for each state. Hurricane Harvey's case study used the 2016 ACS to normalize data for every 1,000 households and Superstorm Sandy used the 2010 Census to normalize data for every 100,000 households.

Before the volume of Twitter activity and customer outages could be analyzed, each state's bi-daily reported customer outages have to be totaled and added to a new excel spreadsheet. The state's total customer outages are normalized with their respective number of households to offer a more precise interpretation of outage volume. This also allows for the data to form a fuller representation of the population. Next, the new keyword Twitter datasets and the DOE's customer outages are loaded into the Tableau software as data sources. Each Twitter keyword data source is plotted with their published date and time on the x-axis and the number of tweets on the y-axis. This shows the volume of Twitter activity, mentioning power outages or weather incidents, over the collection period. Similarly, the DOE's customer outages data are plotted with the date and time on the x-axis and the number of reported customers without power on the y-axis. Time series plots of power-related tweets versus reported outages and weather-related tweets versus reported outages are made for each state included in the dataset. This allows for a cleaner and clearer visual for the spatiotemporal analysis.

5.3 Geospatial Analysis

The final exploration of this research looks at the Twitter and power outage data spatially. ArcMap is used to better visualize and examine the relationship between Twitter activity and hurricane induced power outages. The first step to examine this relationship is to create new sublayers from the larger utility service territories shapefile. The individual service territory layers will be used in future steps to find the number of tweets and households in each service area. Using the EIA's "Major Disturbances and Unusual Occurrences" datasets, each power/utility company that experienced a disturbance in electricity during Superstorm Sandy is identified in the utility service territories shapefile. The affected utility territories and their attributes are selected and saved as a layer. Then each selected service area is individually saved as their own layer. The additional dataset created from the DOE's situation reports are used for Hurricane Harvey's geospatial analysis. The affected utility companies listed in the situation reports are selected from the utility service territories shapefile and saved as one layer. Then each service provider is individually saved as their own layer. The same procedure is repeated, but at the state level for both Harvey and Sandy case studies.

The following steps in this analysis are used to find the number of power-related tweets and households in each utility service area and state. In order to complete this, the power keyword dataset is loaded into ArcMap as a table. The latitude and longitude are used to spatially plot each tweet. The geoprocessing clip tool is then used to 'clip' the tweets layer to each service area. In this case, the input layer is the power-related tweets and the clipping layer is each individual utility company. The output of this geoprocess is a layer of tweets over the input area. The attribute table for each output reveals the number of tweets contained in that service territory or state. In a separate excel sheet, each service provider and their reported power outages and number of tweets are tracked. This process is repeated for all service areas and states for Hurricane Harvey and Superstorm Sandy.

A similar procedure is used to find the number of households per utility provider and state. The 2010 Census block groups shapefile and housing characteristics table is loaded into ArcMap for Superstorm Sandy; and the 2016 ACS block groups shapefile and housing characteristics table is loaded in ArcMap for Hurricane Harvey. The housing characteristics table is joined to the block group shapefile by their "GEOID" attributes. Then the geoprocessing tool is used to 'clip' the block groups to each service territory and state. The input layer is the block groups shapefile and the clipping layer is a service area or state. The output of this geoprocess is a layer of block groups which are contained within the input layer. In each output's attribute table, the column holding the number of housing units is totaled using the statistics tool. The total number of households for each service territory and state is recorded in the previously mentioned excel sheet with each service provider and their reported number of outages and tweets. This practice is repeated for all service areas and states for Hurricane Harvey and Superstorm Sandy.

Retrieving the number of households for every service territory and state allows the analyst to normalize the number of tweets and report outages. Hurricane Harvey’s data is normalized for every 1,000 households, while Superstorm Sandy’s data is normalized for every 100,000 households. The normalization difference is due to the difference in housing density. The Northeast and Mid-Atlantic states have a higher population density than Texas and Louisiana, and therefore have a higher housing density. It is imperative to normalize all data to allow for a more accurate and precise interpretation of Twitter usage and power outages. The new values of tweets per every 1,000 or 100,000 households and power outages per every 1,000 or 100,000 households, for Harvey and Sandy respectively, are stored in the excel sheet.

The last step of the geospatial analysis is to open the layer with all service areas, for each the Harvey and Sandy case studies, in ArcMap and begin an editing session. In this editing session, the information recorded and calculated in the excel sheet is transferred to the selected service areas’ attribute table. Now, the attribute table contains columns consisting of the raw tweet counts, raw number of power outages, the normalized tweet counts, and the normalized number of power outages. These new columns are used to make a choropleth map to visualize the relationship between Twitter usage and power outages.

6 RESULTS

The next chapter presents the results for each of the research questions described in Chapter 3. The results from the geosocial, spatiotemporal, and geospatial analyses are broken up into the separate sections of 6.1, 6.2, and 6.3. Subsections 6.1.1, 6.2.1, and 6.3.1 address the analysis for Hurricane Harvey, while subsections 6.1.2, 6.2.2, and 6.3.2 discuss the results for Superstorm Sandy. The results are further discussed in the next chapter, chapter 7.

6.1 Geosocial Analysis

The geosocial analysis found the most commonly tweeted words and hashtags from the Twitter network during Hurricane Harvey and Superstorm Sandy. The examination of the results allows for a better understanding of the social community during a natural disaster. The results below are of the geosocial analysis for all geotagged tweets from the Hurricane Harvey and Superstorm Sandy Twitter datasets.

6.1.1 Hurricane Harvey

The geosocial analysis revealed the main topics of discussion throughout the Twitter network. The Twitter network was made up of 8,588,000 geotagged tweets. The top two most commonly tweeted words within the Hurricane Harvey dataset identified the location that suffered from major flooding—*Texas* and *Houston* (Table 1). Texas appeared 783,296 times and Houston was used in 711,954 instances. However, although Rockport and Port Arthur were the cities where Hurricane Harvey made landfall, they were only mentioned 40,463 and 16,096 times respectively. As well as Louisiana, which was also affected by Harvey's aftermath, only appeared 42,369 times. Therefore, this analysis revealed that Houston and Texas are the locations in which the social community showed more concern.

Table 1: 20 Most Common Words Tweeted during Hurricane Harvey.

Word	Frequency	Word	Frequency
Texas	783,296	**Flooding**	152,792
Houston	711,954	**God**	145,078
Flood	648,507	**Category**	123,186
Storm	625,368	**Climate**	112,304
Help	403,740	**Tropical**	111,600
Victims	341,300	**Donate**	109,659
Relief	270,014	**Water**	102,044
Katrina	199,763	**History**	99,555
Please	163,252	**Warning**	99,381
Million	158,036	**Disaster**	99,044

Of the top 20 most commonly used words, over half of the words have meteorological connotations (Table 1). These words consisted of *flood, storm, Katrina, flooding, category, climate, tropical, water, history, warning,* and *disaster*. The most common words pertained to updates on Hurricane Harvey and the rising floodwaters surrounding Houston and its neighboring areas. Many people compared the images of the historical flooding to flooding experienced from Hurricane Katrina. Other popular words consisted of *help, victims, relief, please, million, God,* and *donate* (Table 1). Tweets with these words generally tweeted about the number of affected people, links for donating, and areas in need of emergency rescues and resources.

Although the dominating topics of concern were flooding and rescues, the topic of power did appear within the Twitter dialogue. There are many different words to describe a loss of electricity and power outages. Therefore, there was not an overwhelmingly popular word to describe a power outage. Within the social network the word *power* appeared 19,327 times, *energy* was used 8,684 times, *dark* appeared in 4,008 instances, *lights* appeared 2,647 times, *outage*(s) was used 2,366 times, *electricity* appeared in 1,760 instances, *darkness* appeared 594 times, *blackout*(s) was used 396 times, and *powerless* appeared 71 times. Also interesting was that the affected utility and power providers were only mentioned approximately 800 times within the Harvey dataset. These results are also reflected in the hashtag analysis.

The prevailing hashtag topics were Hurricane Harvey and it's sequential flooding in Houston, Texas. The top hashtag was #harvey, which was used 815,464 times and was used over five times more than the second most popular hashtag, #hurricaneharvey. Within the top 20 hashtags, Hurricane Harvey appeared four times as #harvey, #hurricaneharvey, #hurricane, and #harvey2017 (Table 2). As seen in the word frequency analysis, Houston and Texas also appeared in hashtag form. The #houston and #texas are mentioned 96,341 and 58,858 times respectively (Table 2). Similar to the word frequency analysis, the main focus of hashtags was on Houston. However, the landfall cities of Port Arthur and Rockport appeared more frequently as hashtags than they did in the word frequency analysis. The #rockport is the 39th most common hashtag with 5,387 mentions and #portarthur is the 45th most common hashtag with 4,678 mentions.

The most common topic found within the hashtag analysis was flooding. In the top 20 most common hashtags, there are five hashtags that mention flooding (Table 2). The popular hashtags of #houstonflood, #flood, #harveyflood, #flooding, and #houstonfloods talk about the historic flooding in Houston due to Hurricane Harvey. Another five hashtags were concerned with flood relief efforts and spreading situational awareness (Table 2). The hashtags #houstonstrong, #harveyrelief, #texasstrong, #prayfortexas, and #redcross were used as a national campaign to bring awareness to the natural disaster. Lastly, the hashtags #txwx, #houwx, #climatechange, and #goes16 have a meteorological denotation (Table 2). Texas weather (#txwx) and Houston weather (#houwx) were frequently used in tweets to distribute weather updates by meteorological Twitter accounts. The #goes16 hashtag refers to the National Oceanic and Atmospheric Administration's Geostationary Operational Environmental Satellite (GEOS-16), whose images circulated imagery of Hurricane Harvey on Twitter.

The main focus of hashtags was on the historic flooding and helping flood victims. In contrast, the topic of power and electricity outages was not a popular hashtag trend. Again, Twitter users used many different words to describe an outage of electricity. The top four were #power, #electricity, #poweroutage, and #poweroutages. The #power was the 2,076th common hashtag and was only used 86 times. The 3,426th most common hashtag was #electricity and was used 47 times. Lastly, #poweroutage and #poweroutages were the 3,497th and 4,798th most common hashtags and were used merely 46 and 31 times, respectively. Overall, the Twitter network did not show as much concern over power outages as they did for the historic flooding.

Table 2: 20 Most Common Hashtags Tweeted during Hurricane Harvey.

Hashtag	Frequency	Hashtag	Frequency
#harvey	815,464	**#harveyrelief**	15,231
#hurricaneharvey	145,241	**#houwx**	11,314
#houston	96,341	**#texasstrong**	9,598
#texas	58,858	**#climatechange**	9,521
#hurricane	46,148	**#harveyflood**	9,443
#harvey2017	39,948	**#flooding**	9,311
#houstonflood	28,591	**#houstonfloods**	7,192
#houstonstrong	25,816	**#goes16**	6,994
#txwx	20,028	**#prayfortexas**	6,780
#flood	16,519	**#redcross**	6,470

6.1.2 Superstorm Sandy

The geosocial analysis unveiled the Twitter network's main topics of discussion during Superstorm Sandy. The social media dataset is comprised of 1,470,864 geotagged tweets. Within those 1.47 million tweets, the two most commonly used words refers to states that suffered heavily from Superstorm Sandy. New York and New Jersey are represented by *york* and *jersey*, which were mentioned 103,218 and 76,747 times respectively (Table 3). The social network revealed that New York, New Jersey, New York City, the east coast, Manhattan, and metropolitan areas were among the most talked about regions during Superstorm Sandy.

Table 3: 20 Most Common Words Tweeted during Superstorm Sandy.

Word	Frequency	Word	Frequency
York	103,218	**Manhattan**	29,163
Jersey	76,747	**Thoughts**	25,579
Power	70,778	**Please**	25,469
NYC	69,047	**Without**	23,111
Safe	56,548	**Metro**	22,189
Storm	55,536	**Rain**	22,187
East	54,748	**Water**	22,168
Coast	49,693	**Million**	19,780
Superstorm	34,233	**Flooding**	17,358
Prayers	31,292	**Damage**	17,222

The third most commonly used word was *power*. Power appeared 70,778 times and was a major topic of discussion among the Twitter network (Table 3). Frequently, tweets conversing about power outages often mentioned the number of customers without power. Therefore, the words *without* and *million* indicate discussions of power outages and appear in the top 20 commonly used words (Table 3). Within the top 20 most tweeted words, *storm, superstorm, rain, water, flooding*, and *damage* were used to talk about Superstorm Sandy and its lingering aftermath (Table 3). The social community was also dominated by tweets of sympathy. The words *safe*, *prayers*, *thoughts*, and *please* are commonly used to offer signs of concern and well wishes (Table 3). Overall in the top 20 most commonly used words, there were seven words discussing locations of impacts, nine words to describe Superstorm Sandy and its aftermath, and four words denoting sympathy towards the natural disaster (Table 3).

During Superstorm Sandy, power was a large topic of discussion throughout Twitter. Power was the third most commonly tweeted word, but many other words were used to signify discussions of power outages. Some of these words were *lights*, *darkness, outage(s), blackout(s), electricity, energy, utility*, and *powerless*. *Lights* was used 5,195 times, *darkness* had 5,137 mentions, *outages* appeared 4,914 times, *blackout* was tweeted in 4,289 instances, *outage* appeared 4,112 times, *electricity* was used 3,537 times, *energy* had 967 mentions, *utility* was tweeted 432 times, and *powerless* appeared 424 times within the social network. Also, there was over 6,700 occurrences in which names of major utility and power companies were mentioned throughout the Twitter community. Therefore, power outages were a prevailing topic of discussion on Twitter during Superstorm Sandy.

The 20 most common hashtags resemble the dominant topics of discussion from the word frequency analysis. The most commonly used hashtag to talk about Superstorm Sandy was #sandy. The #sandy was tweeted 848,927 times (Table 4). There were five other hashtags within the top 20 hashtags that were used to directly discuss Superstorm Sandy—#hurricane, #hurricanesandy, #frankenstorm, #superstorm, and #storm (Table 4).

Once again, hashtags containing the locations of impacted areas were found among the top 20 hashtags. These hashtags included #nyc, #newyork, #ny, #manhattan, #newjersey, and #nj (Table 4). The landfall city of Brigantine was the 1,105th most popular hashtag. A more popular hashtag was Atlantic City, which was the next closest major city to Sandy's landfall. The #atlanticcity was the 189th most common hashtag and was tweeted 363 times. However, the social media community appeared to be most concerned about New York City.

Table 4: 20 Most Common Hashtags Tweeted during Superstorm Sandy.

Hashtag	Frequency	Hashtag	Frequency
#sandy	848,927	**#manhattan**	4,227
#nyc	39,160	**#newjersey**	3,742
#hurricane	34,291	**#cnn**	3,646
#hurricanesandy	22,515	**#superstorm**	3,560
#frankenstorm	12,803	**#911buff**	2,906
#prayforusa	10,953	**#nj**	2,293
#newyork	9,524	**#storm**	2,234
#staysafe	6,370	**#fema**	2,064
#ny	5,604	**#blackout**	1,643
#news	5,093	**#redcross**	1,470

The social media community suggests that there was a large presence of emergency relief agencies during Sandy because they appeared as trending hashtags. The Federal Emergency Management Agency and the American Red Cross had their own popular hashtags, #fema and #redcross; #fema was tweeted 2,064 times and #redcross was mentioned 1,470 times (Table 4). The community also used hashtags to send well wishes to those affected by the catastrophic event. The #prayforusa, #staysafe, #prayers, #prayfornewyork, and #prayforamerica were popular hashtags across the Twitter network. Unlike the word frequency analysis, hashtags pertaining to breaking news, news networks and news accounts seemed to be commonly tweeted. The #news, #cnn, and #911buff (@911buff is a popular breaking news Twitter account) fall within the top 20 tweeted hashtags (Table 4). Lastly, the term blackout is the 19th most commonly used hashtag. The #blackout is mentioned 1,643 times (Table 4). Other popular hashtags regarding power outages include #nopower, #power, #poweroutage, and #powerout. These hashtags were used roughly a combined 1,400 times. There were also 13 other hashtags with smaller frequencies, which mention power outages. The social network suggests that power outages were a main concern of many Twitter users.

6.2 Spatiotemporal Analysis

The spatiotemporal analysis developed timelines for power and weather-related tweets. These timelines were used to further examine how their timing coincided to reported power outages. The reported power outages were documented twice daily. In order to produce accurate timelines, the Twitter datasets are binned according to the outages reported time periods. The results below illustrate this analysis at the state level for Hurricane Harvey and Superstorm Sandy.

6.2.1 Hurricane Harvey

The spatiotemporal analysis developed a timeline for power outages and compared them to social media discussions of floods and power in Texas and Louisiana. The Hurricane Harvey power keyword dataset contained 108,949 tweets and the Hurricane Harvey weather keyword dataset contained 1,281,152 tweets. The weather keyword dataset was roughly 12 times larger than the power keyword dataset. This significant difference of volume can be seen mirrored from the geosocial analysis from the previous section above. The geosocial analysis revealed that the Twitter network was more concerned with the historic flooding event in Houston than power outages.

The state of Texas begun experiencing a spike in power outages on August 25th at 1000 UTC (Figure 10). During this time, the volume of both keyword datasets were slowly increasing and following a diurnal pattern. From August 24th to August 30th, the power keyword dataset remained constantly centered around 1-2 tweets per every 1,000 households; while the volume of the weather dataset begun to quickly increase on August 26th. The weather-related tweets arrived at a small peak on August 27th at 0300 UTC. This sharp increase can be seen as a gradual increase in power outages. An absolute maximum in the weather-related tweets followed on August 28th at 0300 UTC. This peak corresponded to a peak in power outages on August 28th. Simultaneously, there were roughly 30.6 tweets per every 1,000 households from the weather dataset and approximately 27.8 power outages per every 1,000 households.

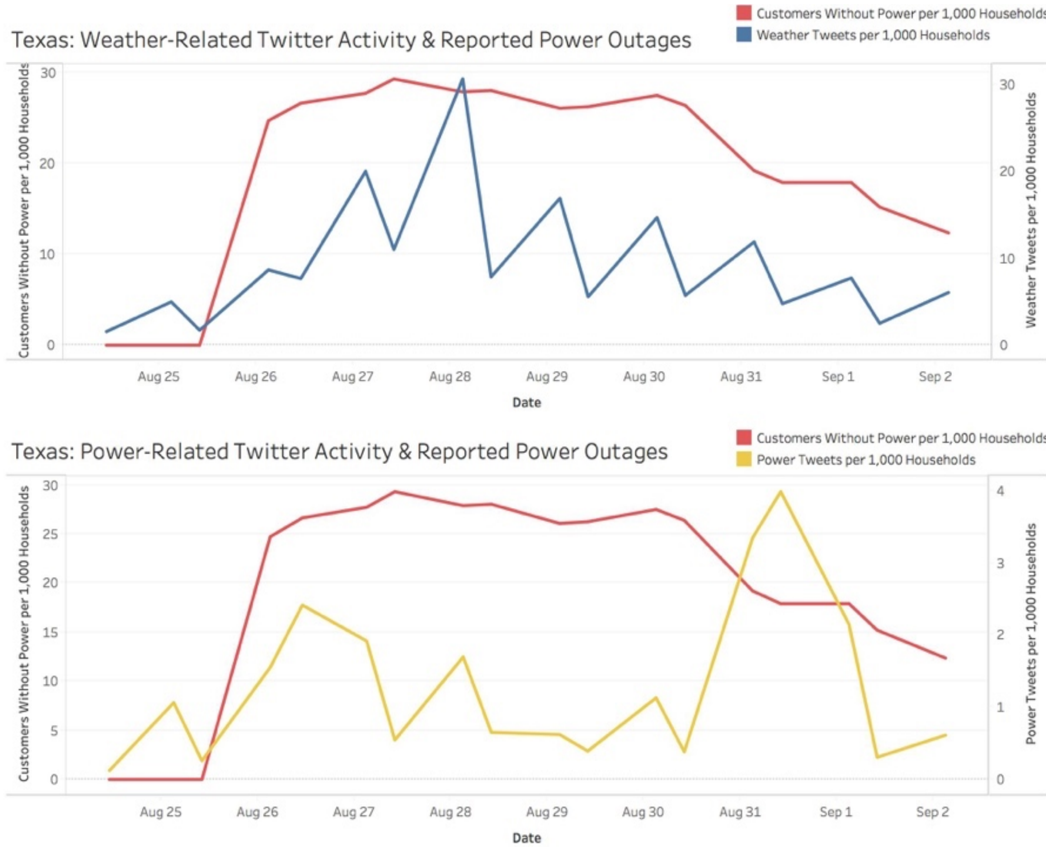

Figure 10: Timelines (in UTC) plotting the volume of weather-related Twitter activity (top) and power-related Twitter activity (bottom) against the volume of customers experiencing power outages in Texas.

Later on August 28th, there is a noticeably downward trend in all three datasets. Over the next two days, the Twitter datasets experience a dip in volume on August 29th at 1000 UTC. This dip in volume follows a slight dip in the power outage dataset at 0300 UTC. After this dip, the volume of power-related tweets begun to gradually increase over the next day and a half. The next peak in Twitter activity comes from the power-related Twitter dataset throughout August 31st. The dataset peaked on August 31st at 1000 UTC reporting 3.98 tweets for every 1,000 households. During this time, the volume of power outages had continued to plunge, but shortly leveled out near 18 power outages for every

1,000 households. Shortly after, all three datasets showed increasingly downward trends until the end of the data collection period.

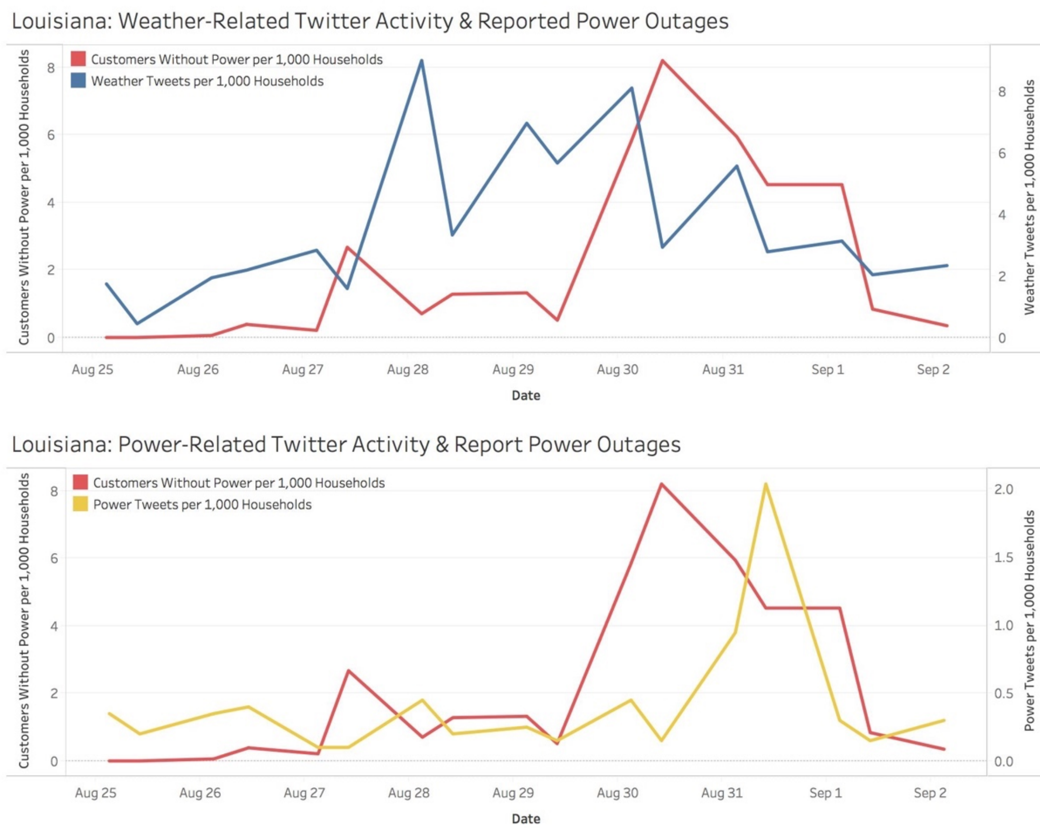

Figure 11: Timelines (in UTC) plotting the volume of weather-related Twitter activity (top) and power-related Twitter activity (bottom) against the volume of customers experiencing power outages in Louisiana.

Opposite trends were seen for the state of for Louisiana (Figure 11). The power keyword dataset closer reflected the power outage dataset. The power-related tweets appeared to have a five-hour offset after the reported power outages. Both Twitter datasets appeared to gradually increase between August 24th and August 26th, while the reported power outages remained constant at zero. By August 27th, all three datasets showed a sharp increase in volume. Around 1000 UTC, the reported power outages seemed to have a small peak of 2.6 power outages for every 1,000 households. Four hours later, the power-related tweets followed with a peak of 0.45 tweets per every 1,000 households and the weather-related tweets peaked nine hours later with 8.99 tweets per every 1,000 households. From this peak, the weather dataset progressively declined until the end of the data collection. While power-related tweets continued to mimic dips in the reported power outages throughout August 28th - 29th. By August 29th, the volume of power-related tweets and the reported power outages begun to exhibit an upward trend. As Twitter activity decreased, reported power outages peaked with 8.2 power outages per every 1,000 households on August 30th. The power-related dataset didn't peak again until 24 hours later with roughly 2 tweets for every 1,000 households. After their final volume summits, both datasets gradually dropped until the end of the collection period.

In Texas, weather-related tweets appeared to have a closer relationship to reported power outages than the power-related tweets. Both the weather keyword dataset and the reported outages dataset showed an overall increasing trend between August 25th and August 28th. They then started to gradual descend over the next day. A decrease in weather-related tweets occurred on August 29th at 0300 UTC and it is associated with a slight drop in reported outages. After this small peak, both datasets progressively decreased until the end of collection on September 2nd. Therefore, the monitoring of weather-related tweets revealed insight into the timing of Texas power outages. In Louisiana, the power-related tweets seem to have a closer relationship to reported power outages. Reports of power outages begun to come in early on August 27th. Hours later, power-related tweets started to rise. Days later, a large spike in power outages are reported. Then twenty-four hours later the power-related tweets responded with a spike in volume.

6.2.2 Superstorm Sandy

Superstorm Sandy's spatiotemporal analysis developed timelines for power outages and compared them to social media discussions of floods and power across twenty states and the District of Columbia. The Superstorm Sandy power keyword dataset contained 106,483 geotagged tweets and the Superstorm Sandy weather keyword dataset contained 96,554 geotagged tweets. The power keyword dataset contained more tweets than the weather keyword dataset. This difference of volume is reflected from the geosocial analysis, which revealed that the widespread power outages were a dominate topic of discussion throughout the Twitter network. The spatiotemporal analysis discovered that trends in Twitter activity and reported power outages were more similar across states that were in closer proximity to each other. This analysis will closely examine the results of each state by the geographic regions of New England, Mid-Atlantic, South Atlantic, and the Mid-West/Central.

The New England states include Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont. The New England states saw a rapid increase in Twitter activity over the first hours of collection. The power-related Twitter dataset came to a peak on October 30th between 0000 and 0100 UTC. Likewise, the weather-related Twitter dataset experienced a spike in activity at 0100 UTC. The reported power outages also saw a steep increase roughly 4 hours later. Simultaneously, all New England states' (except for Vermont) weather-related datasets drastically declined and remained oscillating between 2 – 3 tweets per every 100,000 households until the end of collection. After undergoing a rapid increase, the reported power outages plateau and remained at a steady volume from 0500 to 0900 UTC. Over the next hour, Rhode Island and

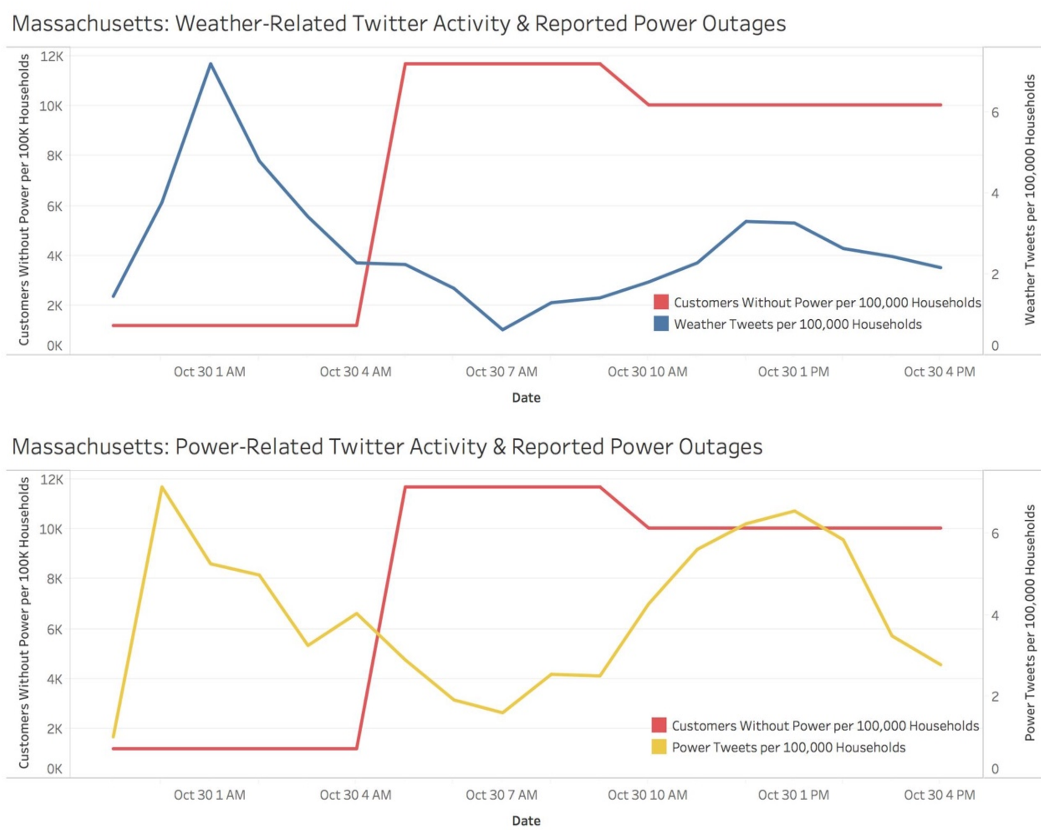

Figure 12: Timelines (in UTC) plotting the volume of weather-related Twitter activity (top) and power-related Twitter activity (bottom) against the volume of customers experiencing power outages in Massachusetts.

Connecticut displayed a slight increase in power outages. While Maine, Massachusetts, New Hampshire, and Vermont saw a drop in customer outages. This drop coincided with an earlier dip in the power and weather-related tweets across all New England states. This dip can be seen between 0600 and 0800 UTC above in Figure 12 for Massachusetts. After this minimum, all states see another peak in power-related tweets between 1100 and 1300 UTC. While the power-related discussions increase, the reported outages remained steady until the end of the data collection. By 1400 UTC, the power-related Twitter dataset begun to decline until the end of collection. Overall, peaks in the volume of power-related and weather-related tweets seemed to be offset prior to the surge of reported power outages across the New England states.

The Mid-Atlantic states saw nearly identical trends between the reported outages and Twitter activity as the New England states. However, the volume of reported power outages for New York and Pennsylvania is nearly doubled that of the New England states. Additionally, the reported power outages in New Jersey was over six times greater than the average number of documented outages for the New England states. The Mid-Atlantic states also saw a rapid increase in Twitter activity over the first hours of collection. In New Jersey and Pennsylvania, the power-related dataset came to a peak on October 30th at 0000 UTC. New York started to see a steep incline of power-related tweets at 0000 UTC, however the dataset did not peak until two hours later. The weather keyword datasets seemed to be nearly identical across all Mid-Atlantic states. All states have a spike in weather-related Twitter activity at 0100 UTC. These peaks in Twitter

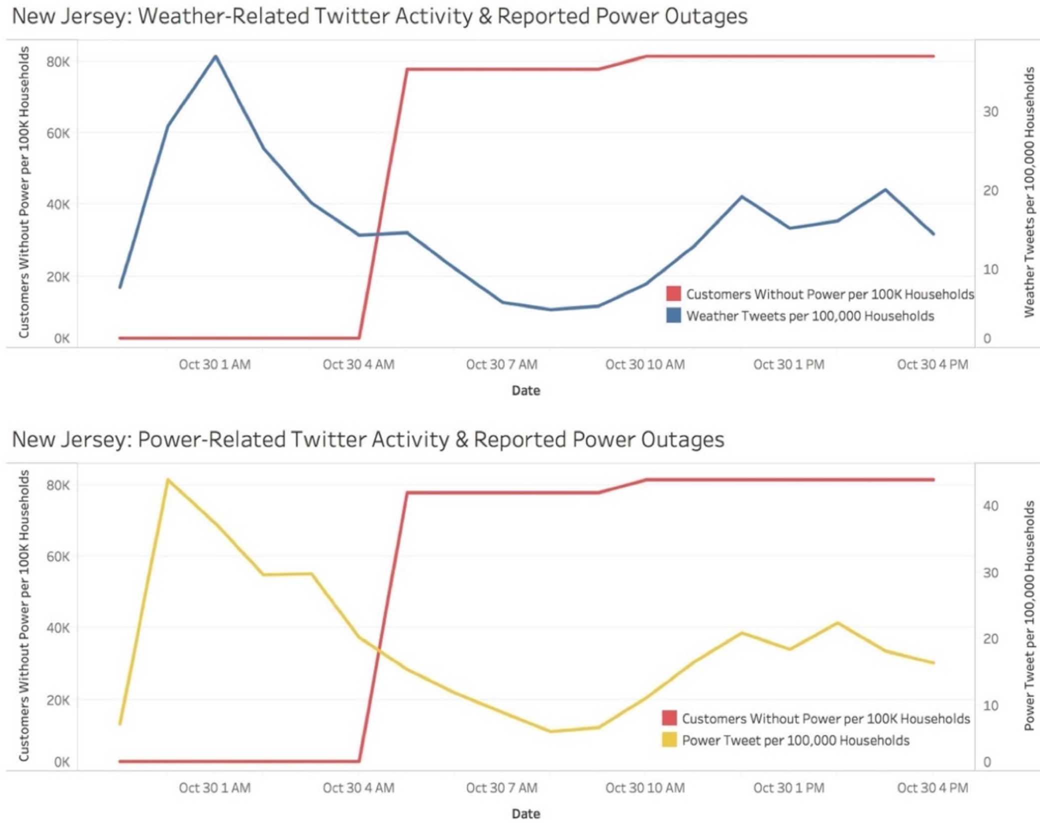

Figure 13: Timelines (in UTC) plotting the volume of weather-related Twitter activity (top) and power-related Twitter activity (bottom) against the volume of customers experiencing power outages in New Jersey.

activity coincide with a sudden increase of power outages reported at 0400 UTC. After undergoing a period of rapid volume growth, both Twitter datasets saw a drastic decrease in volume. The states experienced a minimum of weather and power-related tweets at 0800 UTC. This dip in volume corresponds to a decline in Pennsylvania power outages which occurred between 0800 and 0900 UTC. The opposite is true for New Jersey and New York. Power outages across New York and New Jersey remained constant during the decrease of tweet volume. However, New Jersey (Figure 13) and New York displayed an increase in outages and tweets during the 0900 and 1000 UTC timeframe. Across all three states, the weather and power-related tweets continued to increase until 1200 UTC, then they began to progressively decline until the end of the data collection period.

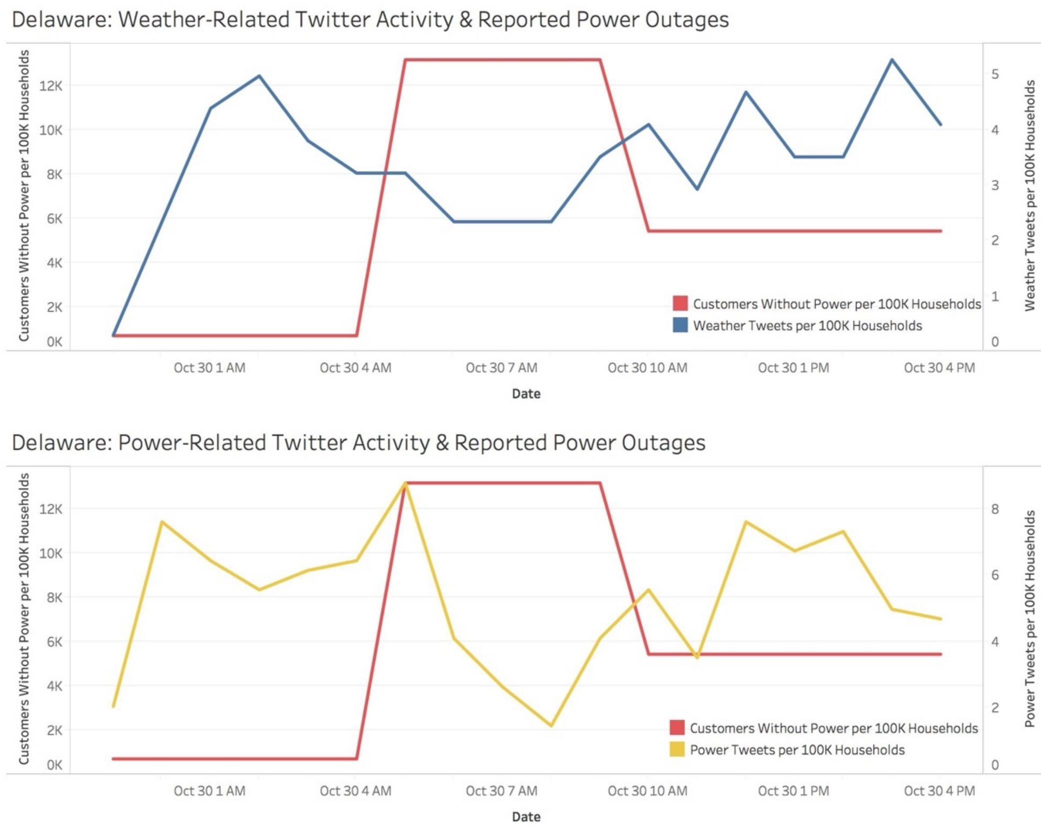

Figure 14: Timelines (in UTC) plotting the volume of weather-related Twitter activity (top) and power-related Twitter activity (bottom) against the volume of customers experiencing power outages in Delaware.

The spatiotemporal analysis for D.C. and the South-Atlantic states of Virginia, Maryland, and Delaware was nearly identical to those of Pennsylvania. All regions experienced the same peaks in power-related tweets at 0000 UTC and peaks in weather-related discussions at 0100 UTC. However, Delaware had a second peak at 0500 UTC, which corresponded to the 13,000 increase in reported power outages (Figure 14). These peaks were followed by a steep incline of power outages at 0400 UTC and plateaus until 0900 UTC. By this time, Twitter volume begun to fall until it reached a minimum at 0800 UTC. An hour later, the power and weather discussions started to steadily increase in volume. Whereas the reported power outages started to decline over the next hour before plateauing again. Chatter across both, weather and power-related datasets progressively increased over the 0800 to 1400 UTC timeframe. Then all Twitter activity declines until the end of their collection period.

The South-Atlantic states of North Carolina and West Virginia had very different relationship trends between Twitter activity and reported power outages. Twitter activity in North Carolina and West Virginia displayed a small peak in power and weather-related tweets at 0100 UTC. By 0500 UTC, West Virginia's reported power outages had increased to over 27,000 outages per every 100,000 households and remained steady until 0900 UTC. Whereas, North Carolina's reported power outages had already peaked with over 400 outages per every 100,000 households at the beginning of collection. Therefore by 0500 UTC, North Carolina was already seeing a dramatic decrease in outages. The decrease in reported outages occurred simultaneously with decreases in power and weather-related Twitter discussions. However, between 0800 and 0900 UTC chatter on Twitter begun to slightly pick back up again as the reported power outages continued its downward momentum. During this same timeframe, West Virginia was still experiencing additional power outages and increased weather and power-related discussions. West Virginia reached 35,000 outages for every 100,000 households by 1000 UTC. This volume of outages persisted until the end of the collection period. Across West Virginia and North Carolina, the volume of tweets from both datasets begun to decrease around 1200 UTC; this trend also endured until the end of the data collection.

Lastly, the spatiotemporal analysis looked at the Mid-West and Mid-Central states of Illinois, Indiana, Kentucky, Michigan, Ohio, and Tennessee. At the start of the data collection, Michigan, Illinois, and Tennessee saw a peak in weather-related tweets earlier than a spike in power-related tweets. While Kentucky, Ohio, and Indiana displayed peaks in power discussions before peaks in weather discussions. After these peaks occurred between 0000 and 0200 UTC, the weather and power discussions on Twitter slowly dwindled down. Within this timeframe, power outages in Ohio, Indiana, and Michigan started to surge at 0400 UTC. Two of these states, Ohio and Indiana, had seen surges of power-related discussions prior to storm-related conversations. On the other hand, Tennessee (Figure 15) and Illinois first saw increases in weather-related tweets and as a result, Tennessee, Illinois, and Kentucky didn’t report any power outages until 0900 UTC. However, regardless of when states begun to see power outages, all Twitter datasets displayed increases in volume from 0900 to 1300 UTC. Ohio, Indiana, and Michigan, which had reported power outages earlier than other neighboring states, saw a slight increase in power outages during this time. Finally, as the Twitter datasets begun to approach the end of the collection period their volume started to progressively decline.

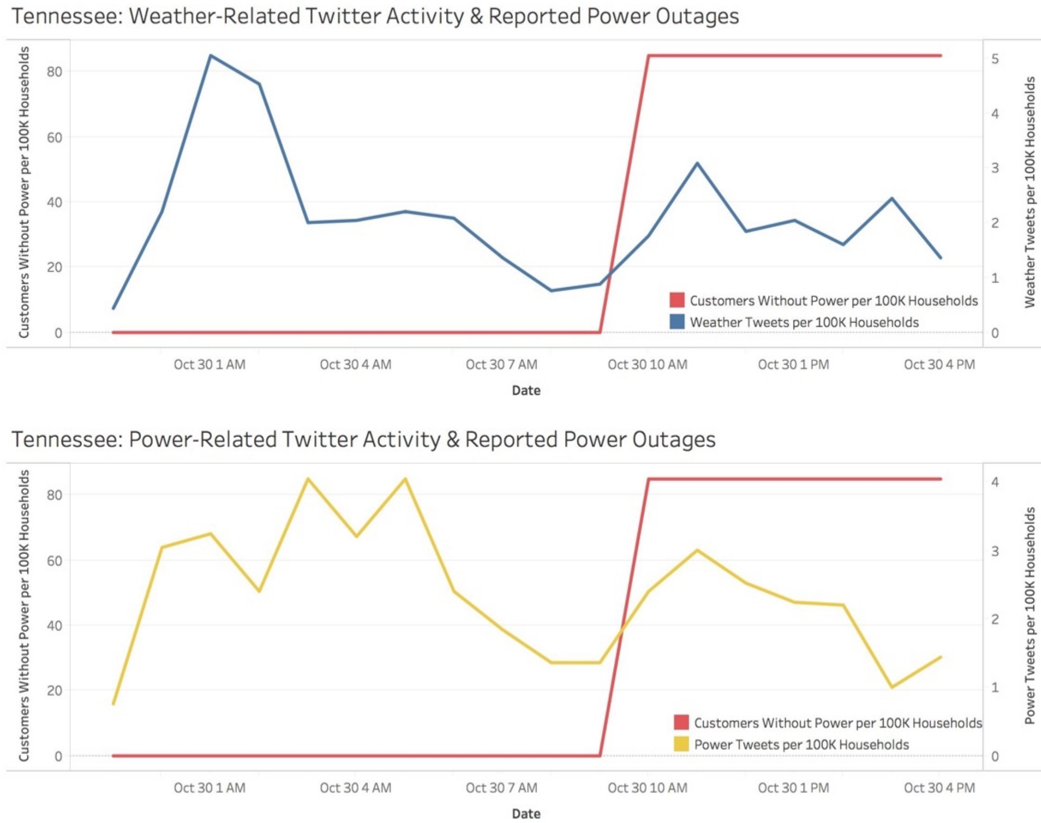

Figure 15: Timelines (UTC) plotting the volume of weather-related Twitter activity (top) and power-related Twitter activity (bottom) against the volume of customers experiencing power outages in Tennessee.

Overall, the temporal monitoring of weather and power-related tweets disclosed a relationship of when power outages are documented. It appears that the states that were closer to where Superstorm Sandy made landfall had higher volumes of power-related tweets. The rise in the volume of power-related discussions was also an indicator of a later rise in power outages. States that were further away from the track of Superstorm Sandy often saw closer relationships between weather-related discussions and reported outages. In this case, rises in the volume of weather-related tweets were frequently followed by increases in documented power outages.

6.3 Geospatial Analysis

The geospatial analysis revealed which areas contributed to discussions pertaining to power outages and how their initial volume of tweets compared to their reported number of power outages. The results below illustrate this analysis at the state and utility service territories levels for Hurricane Harvey and Superstorm Sandy.

6.3.1 Hurricane Harvey

The geospatial exploration revealed which states and service areas had the most power-related tweets per every 1,000 households and their comparison to the number of power outages for every 1,000 households. Of the fourteen utility companies, the top eight utilities with reported power outages and the top ten utilities with power-related tweets were located in Texas. This discovery is supported by the power outage data, which shows that nearly 90% of reported power outages due to Hurricane Harvey occurred in Texas. Also, the Hurricane Harvey power-related Twitter dataset confirms that roughly 94% of power-related discussions were in Texas. Figure 16 also illustrates the dramatic differences in the volume of power-related tweets and reported power outages between Texas and Louisiana.

The geospatial analysis was also completed on a smaller scale—utility service areas. Overall, the Houston area and southern Texas utility companies had the highest volume of power-related tweets and outages. Austin Energy (3268.07 tweets per 1,000 households), CenterPoint Energy (3015.89 tweets per 1,000 households), Entergy Texas

Inc. (2780.39 tweets per 1,000 households), and AEP Texas (2575.63 tweets per 1,000 households) were the service providers which had the most tweets per 1,000 households. CenterPoint Energy, Entergy Texas Inc., and AEP Texas were also among the top four

Hurricane Harvey: Twitter Activity & Power Outages

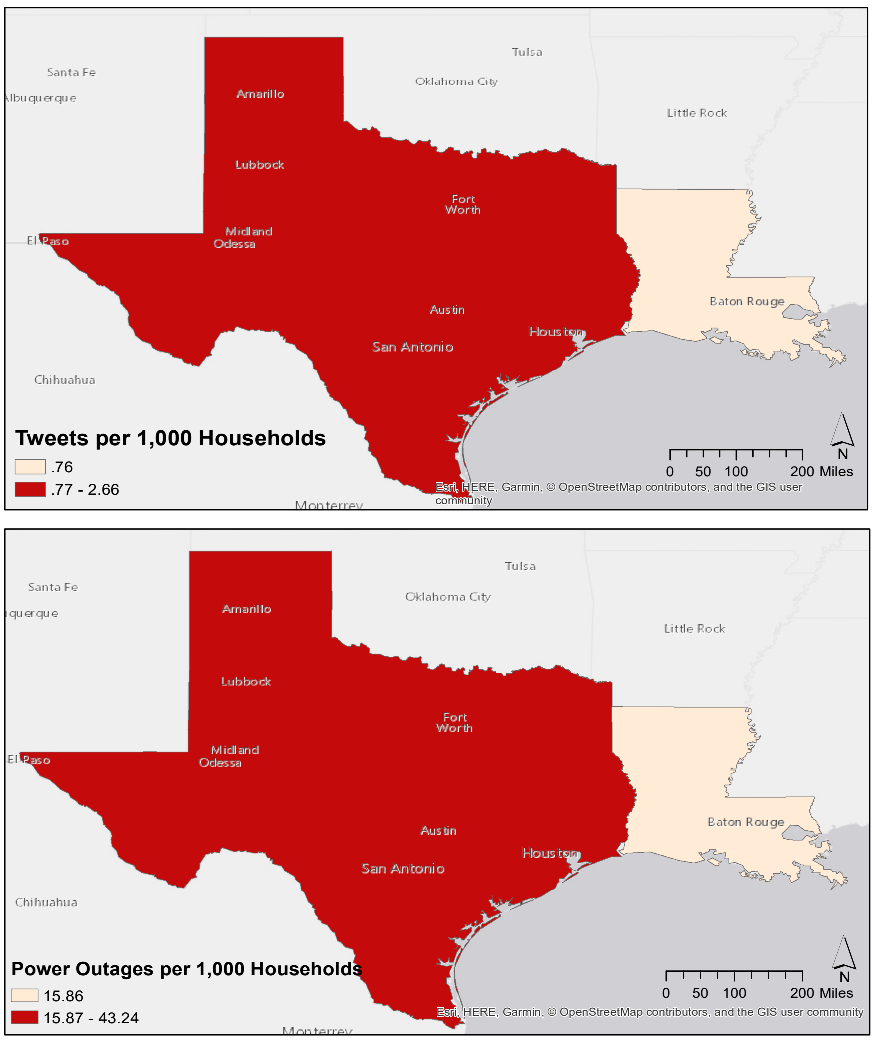

Figure 16: Choropleth map illustrating the spatial relationship between Twitter usage and power outages at the state level.

utility companies that reported outages due to Hurricane Harvey. The top utility companies reporting power outages were AEP Texas (87592.70 outages per 1,000 households), CenterPoint Energy (42818.90 outages per 1,000 households), Entergy Texas Inc. (25671.90 outages per 1,000 households), and Nueces Electric Cooperative (21293.30 outages per 1,000 households). The top six utility territories with the highest tweet volumes were within the top seven utility providers that reported outages. Also, six utilities with the lowest tweet volumes fell within seven of the utilities with the least number of power outages.

As seen in Figure 17, the highest volume of reported power outages was from the utility providers along the Gulf of Mexico's coast. The service territories with the next highest number of reported outages were from Houston area utility companies. These areas also had the highest number of power-related tweets and therefore acted as a good indicator for power outages. In contrast, large utility areas in central and northern Texas had moderate volumes of power-related tweets, but they also recorded the least number of outages. Lastly, utility companies in Louisiana that had very little power-related discussions and reported a small volume of power outages.

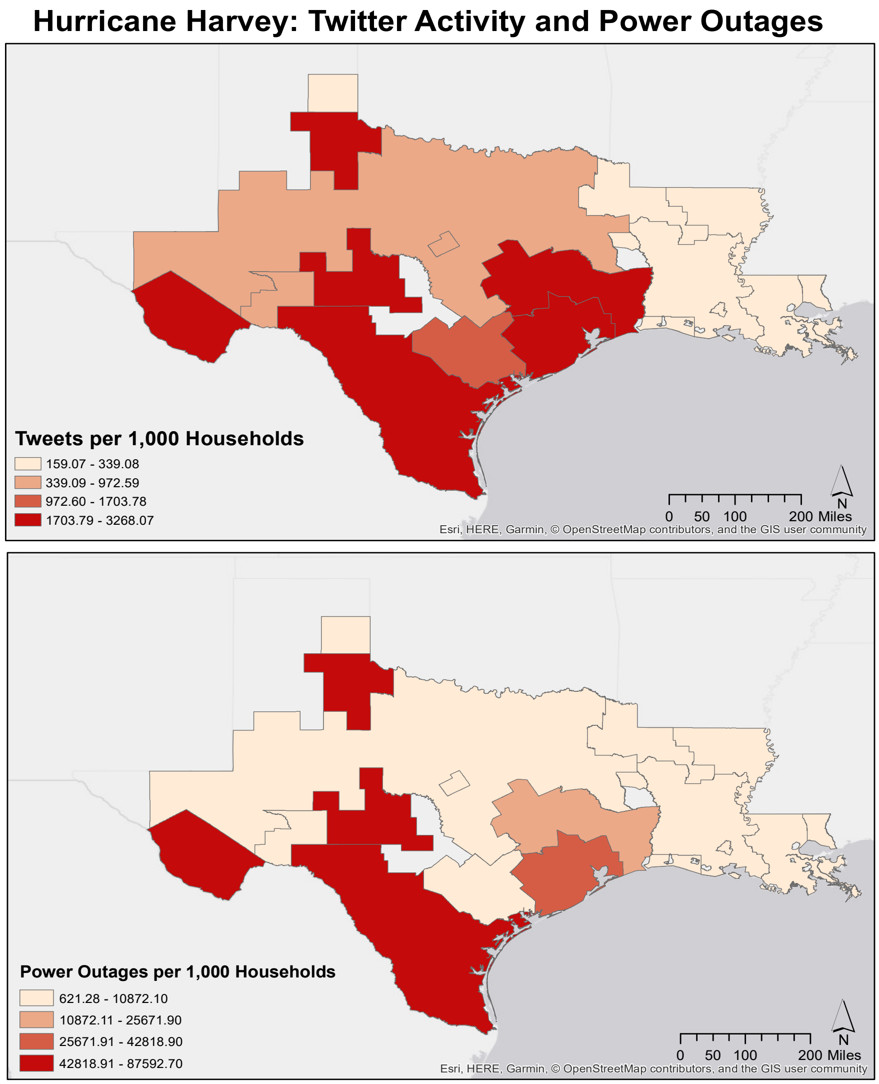

Figure 17: Choropleth map illustrating the spatial relationship between Twitter activity and power outages by utility service territories during Hurricane Harvey.

6.3.2 Superstorm Sandy

The geospatial examination revealed which states and service areas had the most power-related discussions and power outages per every 100,000 households due to Superstorm Sandy. The relationship between power-related tweets and power outages

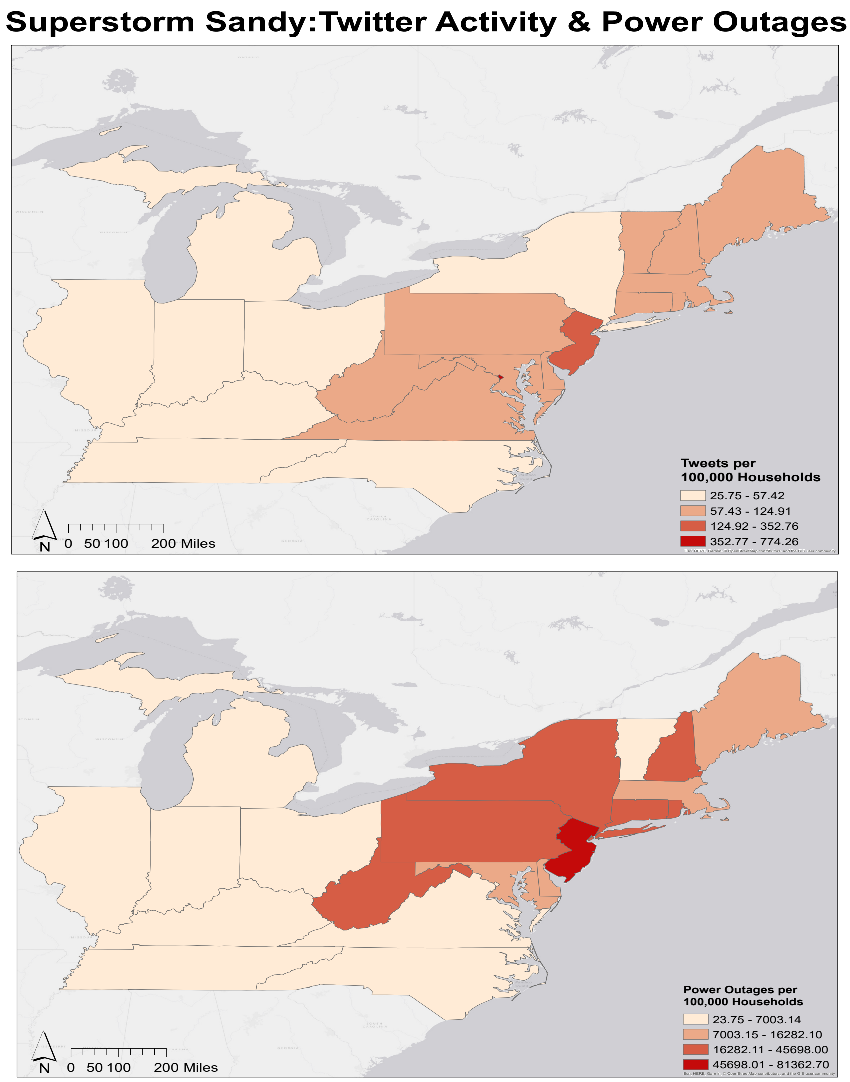

Figure 18: Choropleth map illustrating the spatial relationship between Twitter usage and power outages by states during Superstorm Sandy.

were more consistent at the state level (Figure 18). The highest volume of power-related tweets was in the District of Columbia. New Jersey had the next highest volume of power-related tweets and was also the state with the highest volume of outages. Other states with moderately light volumes of power outages were Vermont, New Hampshire, Connecticut, Rhode Island, Massachusetts, Maine, Pennsylvania, Maryland, Delaware, West Virginia, and Virginia. All of these states, except for New Hampshire, Connecticut, Pennsylvania, and West Virginia, also had light to moderate volumes of power-related tweets. While heavier volumes of reported outages were documented in New Hampshire, Connecticut, Pennsylvania, and West Virginia. New York reported very light volumes of Twitter activity, but suffered from a large number of power outages. Lastly, the lowest volumes of outages and tweets were found in Michigan, Illinois, Indiana, Kentucky, Tennessee, and North Carolina.

The same analysis was completed at a finer granularity—utility service areas. The top five utilities with the highest volume of tweets were Consolidated Edison Co-NY Inc. (312.15 tweets per 100,000 households), Atlantic City Electric Co. (213.22 tweets per 100,000 households), Jersey Central Power & Lt Co. (121.65 tweets per 100,000 households), Public Service Elec & Gas Co. (110.80 tweets per 100,000 households), and PECO Energy Co. (106.49 tweets per 100,000 households). These companies surround the highly populated areas of New York City, New Jersey, and Philadelphia. Although these areas had a large volume of tweets, they all did not exhibit a high volume in overall power outages. The five utilities with the highest volume of outages were Orange & Rockland Utilities Inc. (69899.79 outages per 100,000 households), Monongahela Power Co (53176.5 outages per 100,000 households), PECO Energy Co (47890 outages per 100,000 households), CONVEX (37322.3 outages per 100,000 households), and Long Island Power Authority (36435.10 outages per 100,000 households). The only utility company to have both a high volume of power-related tweets and reported outages was PECO Energy. Otherwise, of the 22 utility companies only six of the eleven utility companies with high tweet volumes were within the highest eleven companies with reported outages.

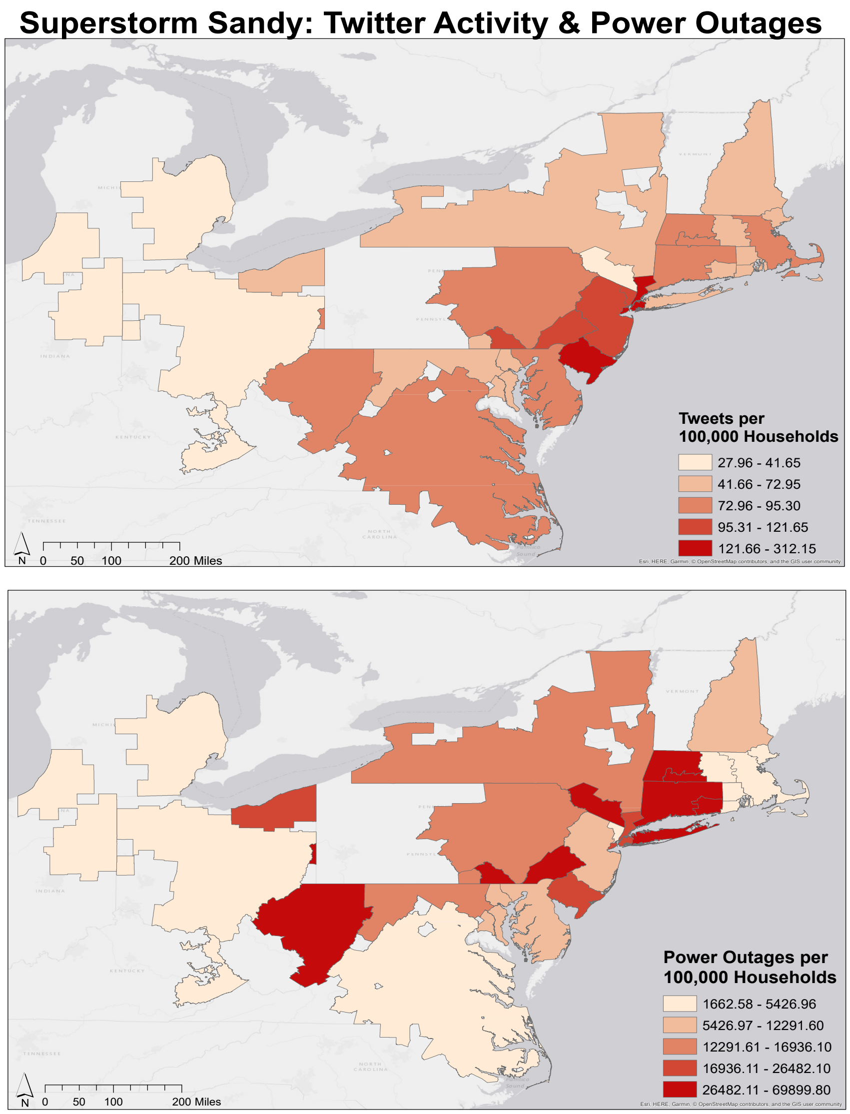

Figure 19: Choropleth map illustrating the spatial relationship of Twitter activity and power outages by service provider during Superstorm Sandy.

Using Figure 19, the relationship between power outages and tweet volume can be further viewed. The highest volume of power outages was around New York City, Philadelphia, northern West Virginia, eastern Ohio, Long Island, southern New Jersey, and western Massachusetts and Connecticut. While, the highest volume of power-related tweets were near New York City, New Jersey and Philadelphia. On the contrary, the lowest volume of power-related tweets was located just northwest of New York City and throughout parts of Ohio, Michigan, Indiana, Illinois, and Kentucky. Whereas Virginia, Rhode Island, and eastern Massachusetts and Connecticut had the lowest volume of power outages.

The utility companies in eastern Massachusetts, Rhode Island and Connecticut had a low to moderate distribution of tweets, but only had a very small volume of power outages. Western Massachusetts and Connecticut had the opposite relationship; their volume of tweets was moderate; however they had a high volume of power outages. Utility companies in Vermont were consistent with lower volumes of tweets and power outages. In New York, the New York State Elec & Gas Corp., which provides electricity across the state of New York, had a moderately light volume of tweets and a moderate volume of reported outages. Long Island also had a moderately light volume of tweets, but had a high volume of outages. Moving towards New York City, Consolidated Edison Co-NY Inc., had the highest number of power-related tweets and a high volume of outages. Yet, Orange & Rockland Utilities Inc., which is the utility provider just northwest of New York City, displayed extremes on both sides of the scale. Orange &

Rockland Utilities Inc. had the lowest number of power-related tweets and the highest volume of power outages. PECO Energy Co. also had one of the highest volumes of power outages, but it also had a corresponding high tweet volume. Other outages throughout Pennsylvania, eastern Maryland, and Delaware coincided with a moderate volume of tweets and reported outages. Northern New Jersey also had a moderate volume of outages, but it had a higher volume of tweets. The same association applied to southern New Jersey. Virginia also saw slightly higher volumes of tweets than power outages. While Maryland and northern West Virginia exhibited the exact opposite relationship. Those utility companies experienced low to moderate volumes of tweets, but actually displayed moderate to very high volumes of power outages. Lastly, utilities in Ohio, Michigan, Illinois, Indiana, and Kentucky all-around had a low volume of tweets and power outages.

7 DISCUSSION

Chapter 7 presents the discussions for each of the analyses described in chapter 5. The discussion of the results from the geosocial, spatiotemporal, and geospatial analyses are broken up into the separate sections of 7.1, 7.2, and 7.3. Each section addresses one of the research questions listed in chapter 3. Additionally, the results from the previous chapter are supplemented with information and overall themes discussed previously throughout this paper. These discussions will aid in the drawing of conclusions found in the next chapter.

7.1 Geosocial Analysis

The main discussion topics discovered in the geosocial analysis differed between Hurricane Harvey and Superstorm Sandy. As previously mentioned the Hurricane Harvey case study appeared to be more concerned with the record flooding in Houston than widespread power outages. These concerns were expressed through words such as *flood*, *flooding*, *Houston*, *help*, *victims*, and *relief* (Table 1); and expressed through hashtags, such as #harveyrelief, #harveyflood, #houstonflood, and #houstonstrong (Table 2). The word *power* didn't emerge until the top 500 words and #power emerged as the 2,076th most common hashtag. Whereas, power outages were a major discussion topic during Superstorm Sandy. The word *power* was the third most used word during Superstorm

Sandy (Table 3). Also, the words *without* and *million* were frequently used to talk about the millions of customers without power. The #blackout was also commonly used to talk about power outages during Sandy (Table 4). The variance in the volume of power outages during Harvey and Sandy is observable by the most commonly tweeted words and hashtags. In the aftermath of Sandy, over 8.5 million customers were without power [3]. While, 1.7 million customers were without power due to Hurricane Harvey [8]. The disparity in topics of concern made these storms excellent case studies because they each served as an example illustrating the relationship between Twitter activity and different magnitudes of power outages.

Another noticeable difference between Superstorm Sandy and Hurricane Harvey's geosocial results were the geographic extent. The top 20 most frequently tweeted words showed the general range of impacted areas. Superstorm Sandy's top 20 words included more locations than Hurricane Harvey's results. Within Sandy's results the locations of New York, New Jersey, New York City, Manhattan, the U.S. east coast and metropolitan areas were commonly identified by Twitter as impacted areas. While Harvey's geosocial results focused on the heavily impacted Houston, Texas region. This dissimilarity in geographic extent is because Sandy's impacts were widespread across 21 states. Whereas, Harvey's wrath was primarily felt by Texas and Louisiana. However comparably, both case studies captured the large volumes of tweets expressing signs of sympathy. Tweets offering thoughts and prayers were prevalent throughout both

Hurricane Harvey and Superstorm Sandy. Also, the #redcross was a trending hashtag during both storms.

Overall, the results of Hurricane Harvey and Superstorm Sandy's geosocial analysis were able to capture discussions of power outages. The analysis also revealed that there exists an association between Twitter activity and reported power outages. Power outages due to severe weather are observable within discussions among the Twitter community. These discussions covered various topics of the powerful storms and their dangerous aftermath, such as flooding, rain, floodwaters, damages, relief efforts and power outages. The word and hashtag frequency analysis also divulged information about the locations of life-threatening aftermath. The locations identified within the top 20 tweeted words were places which felt the largest impacts. Ultimately, the semantic analysis of tweets disclosed a linkage between disturbances in critical infrastructure and social media usage.

7.2 Spatiotemporal Analysis

In Hurricane Harvey's case study, there was a higher volume of weather-related tweets than power-related tweets. As a result, the temporal analysis displayed a closer relationship between reported power outages and weather-related tweets in Texas. The onset of reported power outages started as Hurricane Harvey made landfall on August 25th [14]. Just before the initial reports of power outages, there are small spikes in weather and power-related Twitter discussions. However, the largest spike in weather-related tweets occurs on August 28th. This is roughly 48 hours after Harvey's landfall and the first reports of power outages. This peak is associated with the flooding caused from Harvey stalling over Texas. On August 26th, Harvey begun to stall over south Texas before it gradually moved back into the Gulf of Mexico on August 28th [14]. The slow forward momentum quickly caused catastrophic flooding and therefore caused an increase in weather-related Twitter discussions. Days later, Texas' power-related tweets peaked between August 31st and September 1st. This spike in power-related discussions is possibly attributed to tweets mentioning the slow return of power. The opposite relationship was true for Louisiana; the spatiotemporal analysis illustrated a closer relationship between reported power outages and power-related tweets. The major onset of power outages occurred on August 29th. This is just hours before Hurricane Harvey made a second landfall near Cameron, Louisiana on August 30th [14]. The peak in power-related tweets comes 24 hours after the peak in reported power outages. This peak in Twitter activity could be a delayed response to the large volume of outages. Another possible explanation could be discussions of power returning because the number of reported power outages begins to gradually decline during this time. As seen in Texas, Louisiana's weather-related tweets also peaked on August 28th. This peak can be attributed to the national discussions of the historic flooding in Texas.

Superstorm Sandy's spatiotemporal analysis revealed higher volumes of power-related tweets than weather-related tweets. Within each state, the relationships between power and weather-related tweets appear to follow similar trends. Both Twitter datasets see large peaks 2-3 hours before the initial onset of power outages. Superstorm Sandy made landfall late at 2330 UTC on October 29th [3] and nearly four hours later came the first reports of power outages. These large surges in tweet volume serves as a good indicator a power outage would follow. In most states the weather and power-related Twitter datasets saw a second peak between 1000 and 1400 UTC. These peaks were generally of equal or lesser magnitude of the prior peak. This trend could serve as a sign of an outage because over time social media usage would become limited due to the inability to recharge electronics. However, in some states the second peak in Twitter activity was of greater magnitude than the first. In most states this accompanied a decrease in the magnitude of power outages. As seen in Harvey's spatiotemporal analysis, Twitter users tweeting about power being returned could be responsible for the increase in Twitter activity while there was a decrease in the volume of power outages.

Throughout both case studies, Harvey and Sandy, the daily diurnal cycle is present. The daily cycle is easily visualized in the Superstorm Sandy spatiotemporal analysis (Figure 10 and Figure 11). The dip in Twitter activity between 0400 and 1100 UTC coincides with night on the United States' east coast. As expected, Twitter activity decreases in the overnight and the early morning hours. The nightly dips can also be seen in Hurricane Harvey's spatiotemporal analysis (Figure 12, Figure 13, Figure 14, and Figure 15).

The spatiotemporal analysis discovered that Twitter has the ability to capture when customers lost power due to Hurricane Harvey and Superstorm Sandy. In Sandy and Harvey's spatiotemporal analysis, an increase in the volume of weather and power-related Twitter activity could be used as reconfirmation of an active power outage. Therefore, the spatiotemporal analysis proved that Twitter activity could serve as an indicator of a loss of electricity. Also, Superstorm Sandy's analysis revealed that social media activity that was highly concerned about power outages, could be used to identify an active power outage before traditional reporting by power and utility companies.

7.3 Geospatial Analysis

Hurricane Harvey's geospatial analysis was able to identify areas that had a high volume of Twitter activity and a high volume of reported power outages. The areas with the highest number of reported outages and power-related tweets were located in southern Texas. This region is where powerful, category 4 Hurricane Harvey made landfall between the coastal cities of Port Aransas and Rockport, Texas [14]. The highly populated area surrounding Houston also had a high volume of power-related tweets and the results reported a moderately high volume of power outages. Elsewhere in Texas and Louisiana, there was light to moderate Twitter activity. However, compared to the volume of outages reported along Texas' coastal zone, the geospatial results showed smaller volumes of power outages throughout these regions. Overall, the geospatial analysis discovered a moderately weak positive correlation between power-related Twitter activity and reported power outages (Figure 20). A regression of the logarithmic functions of reported power outages and power-related tweets resulted in a r^2 value of 0.3291 and a p-value of 0.0319. The p-value is less than the accepted alpha level of 0.05, which indicates that the result is statistically significant. Ultimately, areas with higher power-related Twitter activity tended to have a higher number of reported power outages.

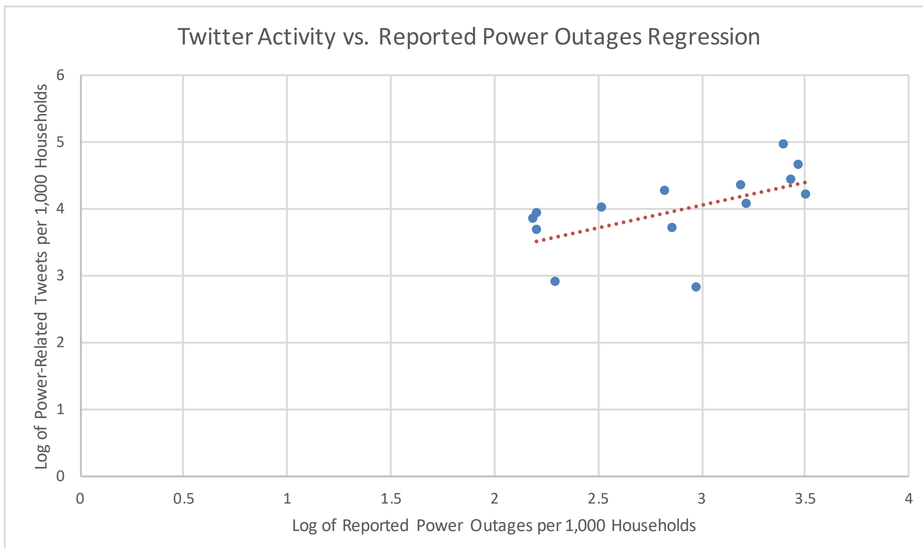

Figure 20: Regression of power-related tweets and reported power outages per 1,000 households.

To contrast, Superstorm Sandy's geospatial analysis determined that the relationship between power-related tweets and reported power outages had little to no correlation and is not statistically significant. A regression of the logarithmic values of reported power outages and power-related tweets resulted in a r^2 value of 0.0011 and a p-value of 0.8852 (Figure 21). The p-value is greater than the accepted alpha level of 0.05, which indicates that this result is not statistically significant. Although the volume of Twitter activity was a relatively good indicator for the number of outages in some areas, such as Vermont, New York City, eastern Pennsylvania, Delaware, Maryland, and parts of Ohio, Michigan, Illinois, Indiana, and Kentucky, it was a poor indicator for areas closer to the landfall region. Utilities areas laying in the coastal zones of Connecticut, Massachusetts, Rhode Island, New York, and New Jersey had more discrepancies between power outages and Twitter activity. A possible explanation for these inconsistencies could possibly be attributed to mandatory evacuations.

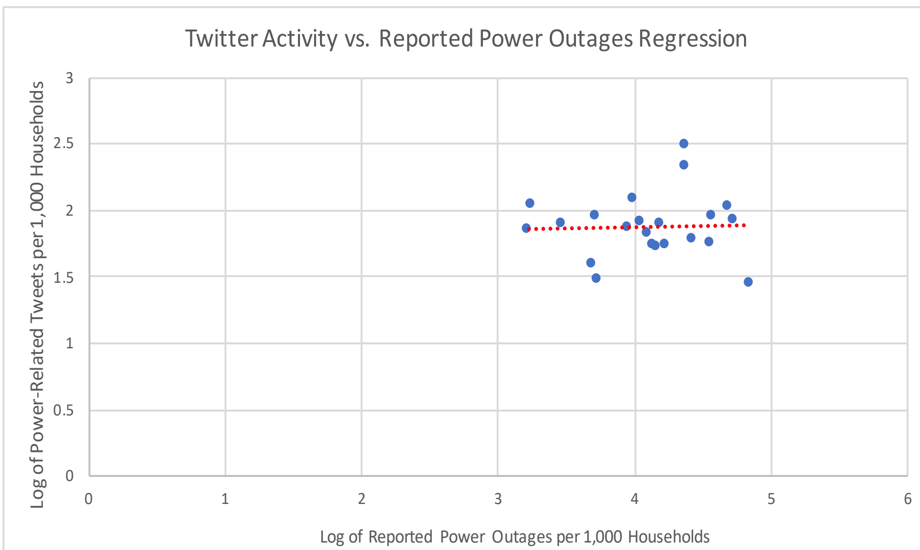

Figure 21: Regression of power-related and reported power outages per 100,000 households.

Evacuations could be a possible explanation of discrepancies between the number of tweets and number of power outages. The major variances between Twitter activity and reported outages lies in an area northwest of New York City, Long Island, and western Massachusetts and Connecticut. Not only did these areas see a lower volume of Twitter activity and extremely high numbers of power outages, but they were all affected by mandatory evacuations. Mandatory evacuations were ordered for the low laying coastal zones of lower Manhattan, Long Island, Connecticut, and New Jersey [52]. An estimated 375,000 people were evacuated from evacuation Zone A in New York City and another 360,000 people were evacuated from the Connecticut coast [52]. Evacuations of this magnitude could explain the decrease of Twitter activity in those areas and the high volumes of power outages. Large evacuations of this degree were not seen during Hurricane Harvey. Mandatory evacuations were issued for the immediate coastal counties around Harvey's possible landfall trajectory [53]. The Houston area was only issued a voluntary evacuation [53]. These evacuations ordered thousands of people to flee low lying coastal areas. The evacuations during Hurricane Harvey were at a lesser degree than Superstorm Sandy. This is why there were more discrepancies between Twitter activity and power outages in Sandy's case study as opposed to Hurricane Harvey.

Lastly, an additional discrepancy found in both case studies can be attributed to the normalization factor. This research used the number of households to normalize the datasets. However, the normalization factor does not account for the density of households. This primarily effected utility areas which covered more land area. An example from the Superstorm Sandy case study is the utility area located in northern New Jersey (Figure 19). This utility area roughly covers the entire state of New Jersey. The discrepancy arises from the fact that most of the power outages were felted by populated coastal areas, which only make up a minor percentage of the total households within the entire the utility area. As a result, the large number of reported outages are normalized by all households throughout the northern and central parts of New Jersey; and therefore, have less of an impact compared to smaller utility areas. The same principle can also be used to explain the results found in New York and Virginia from the Sandy case study (Figure 19) and in northern Texas from the Harvey case study (Figure 17). Factoring in the density of households could aid in understanding outages in larger service areas. Additionally, areas with higher household densities may have one main power line to distribute power. As a result, if a main power line loses power in a high-density area it affects more households.

Ultimately, the volume of power-related tweets preformed well as an indicator for power outages during Hurricane Harvey. The relationship between Twitter activity and reported power outages had a moderately, weak positive correlation, which was statically significant. In the Superstorm Sandy case study, the volume of power-related tweets preformed well as an indicator for power outages in utility companies with smaller land coverage. However, the relationship between outages and Twitter activity had little to no correlation and was found not to be statically significant. Overall, the performance of using Twitter activity as an indicator of power outages is affected by the magnitude of evacuations and the normalization factor.

8 CONCLUSIONS

Chapter 8 presents the final conclusions drawn from this research study. Section 8.1 will discuss the main themes and overall outcomes from each analysis and how it can impact real-world applications. Since this study does have real implications, the challenges and limitations of certain aspects, such as data and resources, are considered in section 8.2. Finally, section 8.3 introduces topics for future research studies in the areas of energy, natural disasters, and geographic information systems.

8.1 Outcomes

Overall, this research was able to determine that Twitter had the ability to capture discussions of power outages during Hurricane Harvey and Superstorm Sandy. Power outages due to severe weather events are largely, observable within discussions among the Twitter community. Therefore, this research study revealed there exists an association between reported power outages and Twitter activity. Likewise, the temporal monitoring of power and weather-related Twitter activity can serve as a secondary indicator of a loss of electricity. Hurricane Harvey's spatiotemporal results showed that an increase in the volume of weather and power-related tweets can act as reconfirmation of an active power outage within the state. Also, the Twitter community throughout Texas and Louisiana exhibited delayed reactions to power outages. Sandy's spatiotemporal results discovered that spikes in power and weather-related tweets can signify the onset of power outages 2-3 hours before traditional reporting. However, using the volume of power-related tweets to indicate locations of power outages is dependent on other factors, such as evacuations. Finally, the relationship between the volume of power-related tweets and reported power outages had a positive moderately, weak correlation in Harvey's case study. As a result, areas with higher power-related Twitter activity tended to have a higher number of reported power outages. Whereas, the same relationship had little to no correlation in Superstorm Sandy's case study.

The results from this study are expected to serve as an additional aid to help electric and utility providers quickly identify areas without electricity during future power outages. The findings from this research can also aid local governments and disaster relief organizations' funding decisions for future disasters. As well, conclusions from this study could aid in developing a standardized framework for detecting blackouts from social media platforms. A social-media-assisted framework can allow utility companies to quickly detect and prioritize areas which need power returned. Information gathered from social media could be used to locate damaged electrical equipment and to allow technicians to assess whether conditions permit for safe repairs. A standardized structure would provide utility providers with an additional avenue to work with local governments and relief organizations to quickly restore electricity. A social-media-assisted framework can also serve as another resource in case of a power grid failure due to a cyber-attack on the U.S. energy grid. The close observations of social media activity could be a valid option to detect outages when traditional equipment fail. The constant monitoring of social media can determine regions, which are without power for an extended period of time and are in need of crucial resources and supplies.

8.2 Challenges and Limitations

During this research study, there were some challenges that slowly emerged. The first challenge dealt with the consistency and accuracy of the geographic coordinates and locations. This study used a mixture of precise and imprecise coordinates. Also, this study assumes that a user tweeting about a loss of power is indeed in an area experiencing a power outage. However, extracted coordinates correspond to the location of the Twitter user and not the location mentioned within a message of a tweet. As a result, a person could be tweeting about a power outage but still have power. Another challenge dealt with the spatial distribution of tweets. Realistically, highly populated areas produce more tweets than lower populated areas. Therefore, dense urban and metropolitan areas are well accounted for, however rural areas are not as well represented. To limit this challenge, normalized tweet counts for each state and service territory were used.

Language bias was a limiting factor in this research study. Nearly 20% of languages spoken in the United States are a language other than English [49]. The most common languages spoken in the United States are English, Spanish, Chinese, and French. Approximately 13% of the population is Spanish speaking [49]. However, this study only analyzed tweets which were tweeted using the English language. The word and hashtag frequency analysis only examined English words. Also, only the English words for power outages and weather events were used to develop the individual keyword Twitter datasets. Therefore, this study has an English language bias. Also, important be noted is Twitter's age bias. There are substantial differences in social media usage by age. A Pew Research Institution 2016 report showed that 24% of Internet users use Twitter [50]. This statistic has been unchanged from a previous study. These studies show that young Americans are 3 times more likely to use Twitter than older Americans [50]. An updated 2018 report published by the Pew Institution states that 45% of 18-24 year olds, roughly 32% of 25-29 year olds, 27% of 30-49 year olds, and 14% of 50 years and over use Twitter [51]. As a result, Twitter usage is skewed towards the younger generations.

Finally, inconsistencies amid the collection of energy data between service territories and states levels was another challenge of this research study. Inconsistencies exist between power outage documentation between EIA and DOE reports. This research mitigated these discrepancies by working with the fullest and most complete reports. Also, the granularity of energy reports was a limiting factor. At most DOE published energy updates bi-daily and EIA only published the overall length of power outages. Power outage reporting could be further improved with an hourly reporting system. A dataset of the volume of power outages reported hourly would be beneficial for similar future studies.

8.3 Future Research

Overall, this research explored the relationship between Twitter usage and power outages. Future research should be conducted by applying the same research methods to past and future hurricanes to confirm and verify the results found during this case study. Future studies should explore the how the relationships between power outages and social media users' behavior change depending on the magnitude of the hurricane. This research can also be expanded by analyzing the mood of tweets. By examining the sentiment of tweets, more information about the Twitter community could be divulged. This could also help separate tweets talking positively and negatively about power outages. For example, tweets expressing happiness that power is return would be in a separate dataset than tweets expressing concern because of a loss of power. Also, if the Twitter collection period could be expanded to include the entire outage length, the positive and negative power-related Twitter datasets could be used to detect the onset and the return of power to customers. This research concept would offer a more in-depth look at the social media community during a power outage in its entirety.

Another future research idea that could build off of this study, is to see if there is a correlation between power outages and other major weather events. Other severe weather events that would make interesting case studies would be the 2012 Washington D.C. derecho, the 2017 Mid-West derecho, the March 2018 Nor'easter, and 2016 Winter storm Jonas. The same methodologies could be applied to these extreme weather events to see whether the Twitter community's response differs during different types of weather events. Also, by using various types of weather events, all with different impacts to critical infrastructure, researchers could learn about how the online community reacts when faced with different durations of outages. Ultimately, researching additional weather phenomenology could further confirm the findings of this study.

Lastly, future work could possibly continue analyzing the Harvey weather keyword dataset to see if the volume of tweets could help meteorologists validate essential forecasts. In the future, operational meteorologists could use similar procedures as additional sources of confirmation for forecasts and verification for watches and warnings. Ultimately, this can help meteorologists increase the accuracy and precision of their warnings. Also, a similar analysis could explore whether higher tweet volumes occur in areas that FEMA has declared areas of major disasters. The findings from this research can also aid local governments and disaster relief organizations during future disasters. Outcomes from this type of research could be used to help direct crucial resources and supplies to inundated areas.

REFERENCES

[1] US Department of Commerce, NOAA, National Weather Service. (2017, November 13). Hurricane Harvey Info. Retrieved April 05, 2018, from https://tinyurl.com/ybs7ukhz

[2] Harris County Flood Control District (HCFCD). (2007). Hurricane Harvey: Harris County Has Never Seen a Storm Like Harvey. Retrieved April 05, 2018, from https://tinyurl.com/y8qmzuwr

[3] US Department of Commerce, NOAA, National Hurricane Center, Blake, E. S., Kimberlain, T. B., Berg, R. J., Cangialosi, J. P., & Beven II, J. L.. (2013, February 12). Tropical Cyclone Report: Hurricane Sandy. Retrieved April 05, 2018, from https://tinyurl.com/ya62cp75

[4] US Department of Commerce, NOAA, National Weather Service. (2014, May 28). Hurricane Sandy. Retrieved April 05, 2018, from https://tinyurl.com/y9te2kh8

[5] Federal Emergency Management Agency. (2017, November 01). FEMA Fact Sheet: Mitigation Assessment Team Results – Hurricane Sandy. Retrieved April 05, 2018, from https://tinyurl.com/ydh4knyl

[6] United States Department of Homeland Security. (2017, July 11). Critical Infrastructure Sectors. Retrieved April 05, 2018, from https://tinyurl.com/ybyczucr

[7] United States Energy Information Administration. (2017, September). Independent Statistics and Analysis: Electric Power Monthly. Retrieved April 05, 2018, from https://tinyurl.com/yck96kkc

[8] Office of Electricity Delivery and Energy Reliability. (2017, September 19). Hurricanes Maria, Irma and Harvey Situation Reports. Retrieved April 05, 2018, from https://tinyurl.com/yaszlro3

[9] United States Energy Information Administration. (2013, January). Independent Statistics and Analysis: Electric Power Monthly: Electric Power Monthly. Retrieved April 05, 2018, from https://tinyurl.com/y8cjykrc

[10] U.S. Department of Energy, Office of Electricity Delivery & Energy Reliability. (2012, November 07). Responding to Hurricane Sandy: DOE Situation Reports. Retrieved April 05, 2018, from https://tinyurl.com/ybed835d

[11] US Department of Commerce, NOAA, National Weather Service. (2017, August 13). NHC Graphical Outlook Archive. Retrieved April 05, 2018, from https://tinyurl.com/y8qtjhs4

[12] US Department of Commerce, NOAA, National Weather Service. (2017, August 17). Tropical Storm Harvey. Retrieved April 05, 2018, from https://tinyurl.com/y9qs4mtd

[13] US Department of Commerce, NOAA, National Weather Service. (2017, August 19). Remnants of Harvey. Retrieved April 05, 2018, from https://tinyurl.com/ybmo5mzt

[14] The Weather Channel. (2017, August 29). Historic Hurricane Harvey's Recap. Retrieved April 05, 2018, from https://tinyurl.com/ybkpq9z2

[15] Klotzbach, P. J., & Bell, M. M. (2017, November 30). Summary Of 2017 Atlantic Tropical Cyclone Activity and Verification of Authors' Seasonal and Two-Week Forecasts. Retrieved April 8, 2018, from https://tinyurl.com/y7a4crkj

[16] Federal Emergency Management Agency. (2017, September 22). Historic Disaster Response to Hurricane Harvey in Texas. Retrieved April 05, 2018, from https://tinyurl.com/y7g6jf79

[17] US Department of Commerce, NOAA, National Weather Service, Climate.gov. (2017, September 18). Reviewing Hurricane Harvey's Catastrophic Rain and Flooding. Retrieved April 05, 2018, from https://tinyurl.com/y974r4sk

[18] The Weather Channel. (2017, October 03). Staggering Stats of Harvey's Impact in Texas Shows the Magnitude of the Deadly Storm. Retrieved April 05, 2018, from https://tinyurl.com/yde4wgpr

[19] Moravec, E. R. (2017, September 14). Texas Officials: Hurricane Harvey Death Toll At 82, 'Mass Casualties Have Absolutely Not Happened'. Retrieved April 05, 2018, from https://tinyurl.com/y97yvagp

[20] Fortune. (2017, September 3). Hurricane Harvey Damages Could Cost up to $180 Billion. Retrieved April 05, 2018, from https://tinyurl.com/ybrbyy5b

[21] Office of the Texas Governor. (2017, November 20). Governor Greg Abbott Extends Disaster Declaration for Texas Counties Impacted By Hurricane Harvey. Retrieved April 05, 2018, from https://tinyurl.com/ybmq4csx

[22] Office of the Louisiana Governor. (2017, August 27). Gov. Edwards Requests Federal Emergency Declaration. Retrieved April 05, 2018, from https://tinyurl.com/y9477e8z

[23] Amadeo, K. (2018, January 11). Hurricane Sandy Damage and Economic Impact Facts. Retrieved April 05, 2018, from https://tinyurl.com/y8j65h75

[24] US Department of Commerce, NOAA, Atlantic Oceanographic & Meteorological Laboratory, Hurricane Research Division. The thirty costliest mainland United States tropical cyclones 1900-2013 (Unadjusted). Retrieved April 05, 2018, from https://tinyurl.com/ycxobkdj

[25] Sullivan, C. (2012, October 29). 375,000 told to flee Sandy trap. New York Post. Retrieved April 05, 2018, from https://tinyurl.com/ybkzavjm

[26] Federal Emergency Management Agency. (2013, July 1). Hurricane Sandy FEMA After-Action Report. Retrieved April 05, 2018, from https://tinyurl.com/jldhybr

[27] Federal Emergency Management Agency. Disaster Declarations by State/Tribal Government. Retrieved April 05, 2018, from https://tinyurl.com/y8rolv9b

[28] Goodchild, M. F. (2007). Citizens as Sensors: The World of Volunteered Geography. GeoJournal, 69(4), 211-221. Retrieved April 05, 2018, from https://tinyurl.com/y7nyepwt

[29] Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting Ambient Geospatial Information from Social Media Feeds. GeoJournal, 78(2), 319-338. Retrieved April 05, 2018, from https://tinyurl.com/yclorexy

[30] Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). # Earthquake: Twitter as a Distributed Sensor System. Transactions in GIS, 17(1), 124-147. Retrieved April 05, 2018, from https://tinyurl.com/ycanzxzz

[31] Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a Social Network or a News Media? In Proceedings of the 19th international conference on World wide web (pp. 591-600). ACM. Retrieved April 05, 2018, from https://tinyurl.com/qh98a24

[32] Twitter. (2016, June 30). Company | About. Retrieved April 05, 2018, from https://tinyurl.com/lb576l5

[33] Boyd, D., Golder, S., & Lotan, G. (2010, January). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In System Sciences (HICSS), 2010 43rd Hawaii International Conference on (pp. 1-10). IEEE. Retrieved April 05, 2018, from https://tinyurl.com/yaptxoya

[34] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. Icwsm, 10(10-17), 30. Retrieved April 05, 2018, from https://tinyurl.com/ycpb3x5b

[35] Weng, J., & Lee, B. S. (2011). Event Detection in Twitter. ICWSM, 11, 401-408. Retrieved April 05, 2018, from https://tinyurl.com/yagkw8hs

[36] Kleinberg, J. (2002, July). Bursty and Hierarchical Structure in Streams. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 91-101). ACM. Retrieved April 05, 2018, from https://tinyurl.com/y95e29xs

[37] Fung, G. P. C., Yu, J. X., Yu, P. S., & Lu, H. (2005, August). Parameter Free Bursty Events Detection in Text Streams. In Proceedings of the 31st international conference on Very large data bases (pp. 181-192). VLDB Endowment. Retrieved April 05, 2018, from https://tinyurl.com/y9m9luqe

[38] Middleton, S. E., Middleton, L., & Modafferi, S. (2014). Real-Time Crisis Mapping of Natural Disasters Using Social Media. IEEE Intelligent Systems, 29(2), 9-17. Retrieved April 09, 2018, from https://tinyurl.com/y9w9trz7

[39] Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using Social Media to Enhance Emergency Situation Awareness. IEEE Intelligent Systems, 27(6), 52-59. Retrieved April 09, 2018, from https://tinyurl.com/y8ur5r2w

[40] Becker, H., Naaman, M., & Gravano, L. (2011). Beyond Trending Topics: Real-World Event Identification on Twitter. Icwsm, 11(2011), 438-441. Retrieved April 09, 2018, from https://tinyurl.com/yb4jkoc9

[41] Sakaki, T., Okazaki, M., & Matsuo, Y. (2010, April). Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In Proceedings of the 19th international conference on World wide web (pp. 851-860). ACM. Retrieved April 09, 2018, from https://tinyurl.com/kyn737h

[42] Jongman, B., Wagemaker, J., Romero, B. R., & de Perez, E. C. (2015). Early Flood Detection for Rapid Humanitarian Response: Harnessing Near Real-Time Satellite and Twitter Signals. ISPRS International Journal of Geo-Information, 4(4), 2246-2266. Retrieved April 05, 2018, from https://tinyurl.com/yabmrt7p

[43] Gilstrap, M., Amin, S., & DeCorla-Souza, K. (2015, July). United States Electricity Industry Primer (United States of America, U.S. Department of Energy, Office of Electricity Delivery and Energy Reliability). Retrieved April 05, 2018, from https://tinyurl.com/ycdpzkzp

[44] Campbell, R. J. (2012, August). Weather-Related Power Outages and Electric System Resiliency. Washington, DC: Congressional Research Service, Library of Congress. Retrieved April 05, 2018, from https://tinyurl.com/y6vwm2oe

[45] Bagasra, A. (2015). U.S. Patent No. 9,041,544. Washington, DC: U.S. Patent and Trademark Office. Retrieved April 05, 2018, from https://tinyurl.com/yd6xff8a

[46] DataCapable. (2016). Data Mining for Utilities. Retrieved April 05, 2018, from https://tinyurl.com/y8qqyrbx

[47] U.S. Energy Information Administration - EIA. Maps - Layer Information for Interactive State Maps. Retrieved April 05, 2018, from https://tinyurl.com/y8dfvh92

[48] US Census Bureau. (2012, September 01). TIGER/Line® Shapefiles and TIGER/Line® Files. Retrieved April 05, 2018, from https://tinyurl.com/dxtmgm9

[49] Statistical Atlas. (2015, April 28). Languages in the United States. Retrieved April 06, 2018, from https://tinyurl.com/y8muhjel

[50] Greenwood, S., Perrin, A., & Duggan, M. (2016, November 11). Social Media Update 2016. Retrieved April 06, 2018, from https://tinyurl.com/zxfjxf4

[51] Smith, A., & Anderson, M. (2018, March 01). Social Media Use in 2018. Retrieved April 06, 2018, from https://tinyurl.com/y9sdmjyr

[52] Jennifer Preston, Sona Patel and Maira Garcia. (2012, October 29). New York Times: State-by-State Guide to Hurricane Sandy. Retrieved April 09, 2018, from https://tinyurl.com/y9l2nnky

[53] McCausland, P., & Chuck, E. (2017, August 25). Hurricane Harvey prompts widespread evacuations in Texas. Retrieved April 09, 2018, from https://tinyurl.com/yb7h3cwt

BIOGRAPHY

Katelyn Thomas received her Bachelor of Science in Atmospheric Science with a specialization in Meteorology and a minor in Geographic Information Systems from George Mason University in 2016. Her research interests include applying geospatial concepts and tools to discover solutions to meteorological and environmental issues.