# USE OF GANG MEMBER SOCIAL MEDIA POSTINGS TO DETECT VIOLENT CRIME

by

Sherry Lynn Fowler

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of
Business Administration

Charlotte

2020

Approved by:

______________________________
Dr. Antonis Stylianou

______________________________
Dr. Reza Mousavi

______________________________
Dr. Shannon Reid

______________________________
Dr. Dongsong Zhang

# ABSTRACT

SHERRY L. FOWLER. Use of Gang Member Social Media Postings to Detect Violent Crime. (Under the direction of DR. ANTONIS STYLIANOU)

Many large cities in the U.S. have a problem with violent crime, some of which is committed by gang affiliates. Those individuals use social media platforms like Twitter to express messages of loss and aggression, which can grow in volume and disseminate quickly, often serving as credible signals to commit an imminent violent crime. These tweets may be useful to law enforcement and community service workers who seek to mitigate violent crime by halting the criminal activity. Thus, this research explores the feasibility of automatically finding criminal signaling of gang members on Twitter and examining the relationship between this signaling and daily crime per city. Content and dissemination features from this analysis, along with time series and other auxiliary predictors, are used to train supervised algorithms. It was discovered that several indicators point to credible aggression and credible loss in gang-affiliated social media posts, including the number of followers, user mentions, and the frequency and speed of the retweets. It was also found that credible aggression, along with several other predictors such as weather and past crime instances, were positively associated with violent crime in the subsequent period. The research shows that knowledge of these indicators has theoretical importance for understanding credible social media posts and later interactive engagement. It also has practical significance for communities to use in mitigating violent crime by finding criminal signals in the virtual space before actual crimes are committed in the physical space.

KEYWORDS:

Violence, crime, gang, social media, Twitter, word embedding, credibility, machine learning

# DEDICATION

This work is dedicated to the memory of two individuals gunned down in Chicago, IL during the writing of this dissertation: Tyquan Manney and Jabari Pittman. May these individuals and others like them never be forgotten.

# ACKNOWLEDGMENTS

Undertaking an effort like this would not be possible without assistance and motivation. I am very grateful for my faith, which has sustained me over the length of this project. I am thankful for the foundational opportunities and love of learning my parents provided early in my life. I am thankful for my wonderful immediate family members: Mark, Megan, Libby, and Graham, as they have each supported and encouraged me throughout this process. I am sincerely appreciative of my dissertation chair, Dr. Antonis Stylianou, and my committee members, Dr. Reza Mousavi, Dr. Shannon Reid, and Dr. Dongsong Zhang, who wisely and patiently advised me through the stages of this project. Lastly, this research would not have been plausible without my research assistants and all of the volunteers and employees of *Project Hood*, a non-profit organization based in Chicago, IL. These individuals live, work, and serve daily in the midst of a dangerous urban area and are seeking to end generational poverty and violence. I am truly thankful to know them.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AIC | Akaike’s Information Criteria |
| API | Application Programming Interface |
| AUC | Area Under the Curve |
| BIC | Bayesian Information Criteria |
| BDK | Black Disciples Killer |
| CA | Credible Aggression |
| CALF | Credible Aggression and Loss Tweet Frequency |
| CBOW | Continuous Bag of Words |
| CL | Credible Loss |
| CLEAR | Citizen Law Enforcement Analysis and Reporting System |
| CNN | Convolutional Neural Network |
| CPD | Chicago Police Department |
| Co-UGS | Credibility of Uses and Gratifications Signaling |
| FBI | Federal Bureau of Investigation |
| GATF | Gang-Affiliated Tweet Frequency |
| GDK | Gangster Disciples Killer |
| GPU | Graphics Processing Unit |
| HMS | Hedonic-Motivated System |
| HMSAM | Hedonic-Motivated System Adoption Model |
| IDF | Interactive Dissemination Frequency |
| IDS | Interactive Dissemination Speed |
| IUCR | Illinois Uniform Crime Reporting Code |
| KDE | Kernel Density Estimation |
| MLP | Multi-layer Perceptron |
| MSE | Mean Squared Error |
| NLP | Natural Language Processing |
| NLTK | Natural Language Toolkit |
| NOSQL | Not Only Structured Query Language |
| PACF | Partial Autocorrelation Function |
| POS | Part of Speech |
| ReLu | Rectified Linear Unit |
| REST | Representational State Transfer |
| ROC | Receiver Operating Characteristic |
| SCT | Social Cognitive Theory |
| SENTPROP | Sentiment Label Propagation |
| SQL | Structured Query Language |
| SVD | Singular Value Decomposition |
| SVM | Support Vector Machine |
| TF-IDF | Term Frequency–Inverse Document Frequency |
| U&G | Uses and Gratifications |
| URL | Uniform Resource Locator |
| VAR | Vector Autoregressive |
| VC | Violent Crime |

# CHAPTER 1: INTRODUCTION

"Sticks and stones may break my bones, but words can also hurt me" (Byrne 1994, p. 38). Words, used as signals, are wielded like weapons for some criminals, including affiliates of street gangs in America's urban areas. Words consisting of threats have the power to not only wound but also kill when used on modern social media as a precursor to the shot of a gun or the swing of a machete. From the recent medieval-style killings in Los Angeles and New York City (NY Times 2019) to the gun slayings in Charlotte, NC (Latos 2017), cities in the United States are battling the effects of violence.

Chicago, the third-largest city in the United States, has such a horrific record of violence that the "Windy City" acquired in 2015 a new moniker "Chi-raq" due to the rising violence and 4,265 murders from 2003-2012, which was approaching at the time the count of 4,410 U.S. soldiers killed in the Iraq War (Goudie 2015). A substantial portion of criminal activities is due to street gangs, with more than 1.4 million people in the United States cited as members of such gangs (FBI 2015). Chicago has 59 active gangs with 150,000 members and 2,000 smaller autonomous cliques (factions) (NBC 2018), most comprised of adolescents or men in their early twenties (Franco, Romero, and Saffell 2018). In 2015, gangs were to blame for 85% of gun murders in the city (Neyfakh 2016). Though the murder count dropped 16% in 2017 (from 771 in 2016 to 650 in 2017) (Park 2018) and 15% in 2018 (Chicago Police Department 2018), murders in Chicago were still high in 2018 and 2019. Unfortunately, this crime problem is not just limited to large cities like Chicago. In 2015, crime began to increase in cities across the United States, representing the largest increase in deadly violence in a quarter-century and reversing a twenty-year drop in violence in American cities (Rosenfeld 2016).

The rise of social media use is partly responsible for fueling this murder epidemic and cycle of urban violence. Social media have altered the street-gang culture and transformed criminal activities, spanning from coercing and cyberbullying (aggressive online behavior) to the ubiquitous propagation of gang member recruitment (Byrne 2015; King, Walpole, and Lamon 2007; Pyrooz, Decker, and Moule Jr. 2015), marking territory, moving product, and communicating directives, including commands to kill (Byrne 2015). Sela-Shayovitz (2012) reported that "the web provides support for gang activities through bragging, posting gang fights on YouTube, and making threats," (p. 393) and the Internet influences the socializing processes of gang members. Young adults and teenagers employ online tools to insult, taunt, and viciously threaten members of rival gangs to gain status, to promote their gang identity, or to share the live streaming of violent acts, also known as cyberbanging (Howell 2010; Patton, Eschmann, and Butler 2013; Tarm 2018). Cyberbanging is defined as "the phenomenon of gang affiliates using social media sites to trade insults or make violent threats that lead to homicide or victimization" (Howell 2010; Patton, Eschmann, and Butler 2013, p. A54). According to a gang-conflict mediator based in a crime-ridden neighborhood, "there is nearly always a link between an outbreak of gang violence and something online" (Tarm 2018, p.1). Known as the new graffiti, threats and taunts that street gang members traditionally meted out to rivals in the physical space now appear in a large volume on social media, with street gang members opening as many as fifty different accounts per member, according to the U.S. State Attorney General's Office. Thousands of gang members in the U.S. (more than 60%) use the popular social media platform Twitter daily (FBI 2015). In the Chicago suburb of Cicero, as much as 70% of gang conflict stems from conversations on social media (NBC 2018). Thus, social media have become a viable source for identifying gang members and monitoring criminal activities, and dissemination of social media-based intentions or events can signal pending violent crime.

These revelations are driving both Criminology and Information Systems (IS) researchers to look beyond gang activities in the physical space. Though researchers have long employed methods to process structured data, usage of methods for processing unstructured data (e.g., text and images) in support of crime watch is still lacking. The imperative need for effective social media analytics techniques that can extract useful socio-behavioral gang-related information, discover relationships between individuals and gangs and among gangs, monitor gang member activities, and predict future criminal activities for preventing and reducing crime is obvious, as the volume of social media data is simply too large to tackle manually. Many police departments attempt to sort through these social media data manually looking for gang member profiles and insights from posts, which is very tedious, time-consuming, and ineffective (e.g., NYC employed more than 300 detectives for this purpose in 2013 alone (PER Forum 2013)). Practitioners also realize they can leverage social media analytics to help combat crime, as they recognize the capacity of social media as a valuable information source. Many gang-related confrontations initiate via social media, intensify over time, and culminate in a physical crime that could have been prevented by timely notice to an authority who, in turn, could have proactively reached out to the offender or even the target victim (PER Forum 2013). Thus, if an automated process can identify Twitter messages (tweets) that credibly taunt or threaten others in the virtual space, it may quickly alert law enforcement officers or community workers to stop a future crime in the physical space.

This study can also assist future researchers due to limitations and gaps in prior social media studies. The first gap is the relative lack of quantitative studies in the Criminal Justice (CJ) literature related to gang-affiliated crime and its association with social media. While several general CJ qualitative studies exist with a focus on various reasons for, and predictors of, violent crime, other research that includes a social media focus does not test and validate the theories used in qualitative research. Further, there is a lack of specific CJ studies empirically testing *credible* gang-affiliated social media signals, as some gang-related messages involve mere bragging without intent to follow through with the criminal action (Stuart 2019). A dearth of existing IS and interdisciplinary studies that tie violent crime predictors to sound integrated theory also exists. This leads to the two key research questions and objectives in this study:

**RQ1:** Are credible taunts and threats (aggression) and expressions of loss in social media posts by street gang members effective predictors of violent crime?

**RQ2:** Are the characteristics of interactive dissemination behavior of street gangs on social media (e.g., frequency and speed of retweets) effective predictors of violent crime?

This two-stage interdisciplinary quantitative research addressed the above research questions by investigating the credible signaling of violent crimes. In the first stage, unstructured tweet content (e.g., text, emojis, hashtags, and mentions of users) was analyzed over two dimensions: loss (represented by out-of-control sadness or similar emotion due to the death or incarceration of an associate, friend, or family member) and aggression. Aggression and loss content was combined with other predictors to determine the perceived credibility of a tweet at an individual unit of analysis. In Stage 2 of the study, the individual data were aggregated to a daily city level to use as another unit of analysis, primarily because of the lack of specific individual location data associated with the tweets. These credible aggregated social media predictors were combined with additional auxiliary determinants of a sociological and temporal nature and past historical crime counts to predict the city's next day violent crime. The importance of detecting violent crime is that policymakers, law enforcement agencies, and social workers are interested to see the counts of forecasted violent crime and if the crime in a city is declining or not.

This study made several theoretical and practical contributions to the fields of Criminology and IS. Firstly, the study extended the social media literature by creating a novel approach that uses both structured and unstructured social media data to automatically classify and aggregate behaviors related to credible criminal signaling by street gang affiliates. This solution (when fully implemented) will be one step closer toward the creation of an automated social media monitoring tool to replace the time-consuming manual process currently deployed by law enforcement, community leaders, and social workers in their efforts to fight crime and improve the safety of neighborhood communities. The tool should enhance (not replace) the role of civic leaders by complementing their "on the ground" domain knowledge and facilitating a faster process by identifying questionable tweets and alerting authorities, allowing them to make the final decision of whether and when to act on the insight. The premise is that better crime predictions can improve the allocation of limited police officers and social workers, lessen wasted efforts, and assist civic workers in their attempts to proactively assuage pending crime.

Secondly, this research studied determinants of credible social media messages, as there is little definitive research about what types of expressions by aggressive gang-involved individuals are perceived as credible. Thirdly, the study integrated and extended the current criminology literature by introducing a combination of gang-affiliated violent crime determinants. Fourthly, this study not only asserted a theoretical contribution but also validated it through empirical evaluation. It tested the characteristics of interactive dissemination behavior (retweet frequency, retweet speed, @-mention frequency, and favorites (likes) frequency) of credible social media postings, expanded the major event predictor to include exogenous events (not just local ones), and substituted more granular weather averages in the predictive model. This allows researchers, law enforcement officials, and social workers to analyze overall violent crime at the city level.

The study found that several indicators point to credible aggression and credible loss in gang-affiliated social media posts, including the number of followers, user mentions, retweet frequency, and retweet speed. It was found that credible aggression, along with several other predictors, such as weather and past crime instances, were positively associated with violent crime in the subsequent period. The research shows that knowledge of these indicators has theoretical importance for understanding credible social media posts and later interactive engagement.

The rest of the paper is organized as follows. The literature is reviewed in the next section with an emphasis on violent crime and social media systems, particularly regarding the diffusion of loss-filled or aggressive messages on social media. In the subsequent section, the theoretical underpinnings that provide the basis for the development of the research hypotheses are presented followed by an overview of the research data, methodology, findings, and concluding thoughts.

# CHAPTER 2: LITERATURE REVIEW

Prevention of crime has been an important topic for researchers in an effort to assist the society by understanding causes of crime and attempting to mitigate it. While widely studied in the field of Criminology (Zhao and Tang 2018), this topic is comparatively understudied in the IS discipline, even though Internet technologies, data analytics, and social media technologies affect the study of crime. Various theories fit well in the investigation of crime by individuals. To set the context of the research, existing literature about violent crime was reviewed, including the key theories surrounding it, the data related to crime, and the various information system and social media methods and tools used to analyze the data, with the intent of detecting, studying, and mitigating crime. The paper highlights what could be missing and what this study adds to improve the existing methods.

## 2.1 Criminology and Psychology Literature

There are several theories and mechanisms used to study crime, including those related specifically to Criminology and Psychology (Table 1) and others associated with the interdisciplinary study of crime via social media technologies from the IS discipline. Some social theories, like Social Strain Theory, suggest that because the normal culture is inundated with dreams of opportunity, liberty, and affluence, if the social structure of opportunities or abilities is perceived to be unequal, some of the people disappointed will use unlawful criminal methods to attain it (Featherstone and Deflem 2003).

**Table 1: Criminology and Social Psychology Theories**

| **Name** | **Literature** | **Description** |
|---|---|---|
| Crime Pattern Theory | (Papachristos and Hughes 2015) | Suggests that geographic and temporal features influence the where and when of crime patterns. Suggests that gangs do not randomly scatter, and socio-economic influences make possible their existence in certain areas. |
| Deterrence Theory | (Chalfin and McCrary 2013) | Suggests that an increase in an offender's chances of being caught decreases crime, and that crime can be controlled with punishments that combine the proper degrees of certainty, severity, and celerity. Deterrence is a primary component of the U.S. justice system and a core principle of Classical School and Rational Choice theories. |
| Free Will Theory | (Maslow 1943) | States the belief that humans are rational and can make decisions according to each individual's own will and purposes. Under this perspective, people can understand the contrast in right and wrong moral judgments and can choose to commit criminal acts or to follow the law. |
| Rational Choice Theory | (Lazzati and Menichini 2016) | Suggests a premise that people commonly act in their self-interest and decide to commit crime after comparing the possible risks (including being caught and punished) with the rewards. |
| Routine Activity Theory | (Cohen and Felson 1979; Pyrooz, Decker, and Moule Jr. 2015; Sampson and Lauritsen 1990) | Suggests that crime occurs in the absence of competent supervisors (or formal controls) at points of space and time convergence of motivated offenders and victims. This theory is innate because connecting to the online world has become so commonplace for individuals. Other researchers contend that Routine Activity Theory may speak more to victimization than wrongful behavior (Sampson and Lauritsen 1990). The perceived freedom associated with technologies allows its exploration, but these platforms also have a dark side; that is, they expose individuals (especially youth) to online violence via anonymity and limited oversight without proper adult supervision. |
| Self-Control Theory | (Gottfredson and Hirschi 1990; Krueger et al. 2007) | Suggests that "individuals with low self-control will find crime appealing, because they are cannot see beyond the consequences of their actions." Well established as one of the strongest predictors of offending. |

**Table 1: Criminology and Social Psychology Theories (continued)**

| | | |
|---|---|---|
| Social Control Theory | (Ang 2015) | Emphasizes the importance of peer and family relationships and suggests that parents and peers play a critical role in facilitating some forms of violent behavior. Ang (2015) found an association between cyber aggression and the absence of parental supervision and poor emotional connections with parents. Researchers of several hundred adolescents found that partner-directed cyber aggression related to uncertain maternal connections and partner attachments. In gang-related crime research, the link to this theory is due to gang members learning deviant behavior from other gang members or media postings. |
| Social-Disorganization Theory | (Sampson and Groves 1989) | Suggests that physical and social environments are primarily responsible for a person’s behavioral choices and that a neighborhood with undesirable social structures (e.g., lack of jobs, poorly performing schools, empty and defaced buildings, etc.) has a higher probability of elevated crime rates. At the community level, community social disorganization can increase due to poverty, ethnicity, mobility, and disruptions within a family. This disorganization can lead to an increase in crime. |
| Social Exchange Theory | (Redmond 2015) | Suggests that social behavior often involves social exchanges where an individual is motivated to attain some valued reward for which he/she must forfeit something of value (cost), such that he/she seeks profits in exchanges where rewards are greater than the costs. Individuals may be bothered if there is no equity in an exchange or where others are rewarded more for the same costs they incurred. |
| Social Identity Theory | (Stets and Burke 2000) | Describes the idea that becoming part of different groups via membership helps construct identities. |
| Theory of Environmental Criminology | (Helsley and Zenou 2014; Lazzati and Menichini 2016) | Emphasizes the role that spatial factors play in crime location. |

At a societal level, these attempts at success through crime have increased because of the lack of police confidence and trust by some individuals (Rosenfeld 2016). Research has uncovered an emotion of hate toward police officers by criminals, and anger at law enforcement has been revealed in Twitter posts. For example, social media posts show threatening content including both text and small images of guns, bombs, or explosions directed towards law enforcement, indicating a desire to kill a cop (Balasuriya et al. 2016). This declining institutional legitimacy (a decrease in trust and legitimacy in law enforcement) can amplify violent crime, as individuals and communities are estranged from suitable means of social control and think they need to deal with problems themselves (Byers 2014; Rosenfeld 2016; Roth 2009). This general anger also causes neighbors not to cooperate with the police in crime investigations. Further, people may know who harmed others, but they tend not to turn in the names to the police due to lack of trust or fear of reprisal by others (Rosenfeld 2016). All of this is exacerbated if the "perception" (not necessarily fact-based reasoning) is that police do not follow-up or solve crimes (Leovy 2015), process the crime evidence slowly, or process arrests slowly, especially in communities of color (Sampson and Bartusch 1998). The result is a higher likelihood of law-breaking (Tyler 2006). Some researchers suggest that widely publicized uses of force by police validate the underlying belief system in these communities, especially if the publicity was over social media. However, others respond that "media accounts with ambiguous moral valuations strain the pliability of police officers as 'crime fighters' (Hirschfield and Simon 2010) to provide a valuable position from which the actual responses of police officers can be mediated to determine adherence to sanctioned reactions" (Beckett 1997; Watson 2016, p. 4).

Crime can increase even more if an insufficient number of police officers are on the streets. One study found that saturating high crime areas with police officers on foot could significantly reduce violent crime (Ratcliffe et al. 2011) and improve public perception (Dalgleish and Myhill 2004). Further, to exacerbate the problem, the decrease in public corruption convictions has been associated with an increase in crime, as has played out in some large urban areas in the U.S., like Chicago, IL, known as the most corrupt large American city (Simpson, Gradel, and Rossi 2019).

Other criminology theories counter the idea that crime is outside of a person's control due to psychological, biological, or social factors. For example, Deterrence Theory suggests that an increase in an offender's chances of being caught mitigates crime, and that crime can be controlled with punishments that blend fitting levels of certainty, severity, and celerity. Deterrence is a primary component of the U.S. justice system and a core principle of Classical School and Rational Choice theories (Chalfin and McCrary 2013). Several studies suggest that higher incarceration rates are positively correlated with crime reduction (Mac Donald 2016; Rosenfeld 2016; Travis, Western, and Redburn 2014).

In other studies, researchers have highlighted that less technology-driven policing can result in higher crime, while the correct use of "precision policing" (as it is known in New York City), combined with proper accountability and accessible crime-related information, may reduce crime. For example, the New York Police Department has pioneered its 20+ year use of CompStat, where police efforts targeting hotspots (dense areas of high crime) are associated with crime reduction (Mac Donald 2016). Investigational assessments consistently concur that directed patrols in crime hotspots can result in significant crime reductions (Braga and Bond 2008; Taylor, Koper, and Woods 2011). Structural changes (e.g., video cameras or shot-spotting sonar equipment) in hotspot crime areas can also reduce the ratio of productivity-to-risk and thus reduce crime (Braga and Weisburd 2010; Lazzati and Menichini 2016), though some studies report less efficacy in medium-sized cities of other countries (e.g., Sweden) (Gerrell 2016).

Other research suggests that community development efforts combined with effective policing have the most promise in bringing about sustainable crime reductions (Braga, Papachristos, and Hureau 2012). On its own, strong community cohesion has been associated with less crime (Ehrenfreund and Lu 2016). Gentrification and urban reforms, including mass transit options, are positively associated with crime reduction, as exemplified by Washington D.C., where tax breaks to businesses who agreed to move within the borders of the area and other urban reforms resulted in a crime decrease, according to John Roman, professor of criminology at the University of Pennsylvania (Fisher 2012).

Violence prevention programs have also been known to mitigate crime. For example, the Cure Violence model (also known as CeaseFire) (Slutkin, Ransford, and Decker 2015) is based on public health techniques at the community level. It claims to halt deadly violence proactively and to prevent its contagion by employing interrupters who interject ongoing conflicts with individuals of the highest risk to modify violent behavior and change community patterns (Slutkin, Ransford, and Decker 2015). Similarly, in New York City, the Citizens Crime Commission’s E-Responder program shows that interventions in online disputes by community members can mitigate real-world gun violence. Programs like these provide helpful tools and train workers to recognize signals of online risk, including threats and words of grief or emotional anguish. These initiatives are examples of “solutions that don’t just respond to violence but get ahead of it” (Friedrich 2017, par. 8). Resourceful programs that strive to remove illegal firearms from the streets are associated with a decrease in crime (Bruinius 2018); yet, not all urban areas have achieved the same level of success as New York City. For example, Chicago has some of the most stringent gun laws in the country; yet, on average, it has seized seven times more guns than NYC and two times more than Los Angeles (AJ+Docs 2017). Approximately six thousand illegal guns are removed from the streets each year (approximately eight thousand in 2016). These guns are usually coming from intermediaries (e.g., brokers), most originating in the state of Indiana, who put the guns in the hands of youth (AJ+Docs 2017).

Research has uncovered one other important predictor of violent crime in urban areas. There is a great consensus in the academic community that the correlation between membership in a deviant subculture, like a criminal gang, and delinquency is very pronounced for violent offenses (Thornberry et al. 2003). Thus, when individuals, including those associated with gangs, do not (or choose not to) live up to society's expectations via appropriate methods like a strong work ethic or delayed gratification, they may endeavor accomplishments through illicit means like crime. Gangs enable and promote violence by members indirectly by facilitating members' access to the distribution of drugs or directly through gang tasks like turf defense. Alternative enabling modes also exist, including gang-related murders (those stemming from gang activity but not involving gang members), gang-motivated murders (those stemming from gang activity and involving gang members), and non-gang youth homicides (Rosenfeld, Bray, and Egley 1999). Though various scholarly definitions of a gang exist (Esbensen et al. 2001), this study defines a gang as "a self-formed association of peers, united by mutual interests, with identifiable leadership and internal organization, who act collectively or as individuals to achieve specific purposes, including the conduct of illegal activity and control of a territory, facility, or enterprise" (Miller 1992, p. 21).

### 2.1.1 Signaling Theory

One of the most promising theories employed when studying criminal behavior related to gangs is Signaling Theory. Signaling Theory posits that one person can credibly convey underlying intentions via an online message (signal) to another person (DeWitt 2018; Pyrooz and Densley 2016). This theory has much to offer studies of criminology in general, and gangs in particular. Social media offers a new way to advertise and amplify signals of a gang member's anger and trespassing into another member's territory via combinations of content in videos (Lauger and Densley 2018), pictures (bloodstain images), emojis, and text. For example, the investigation of the 2012 death of a young Chicago-native rapper, Joseph Coleman ("Lil Jojo") and the last tweets that he posted connected to a video where he was yelling vulgar words to a member of a rival gang who subsequently responded, "I'ma kill you." Further investigation showed that the conflict resulting in the death originated and was carried out entirely via social media (Austen 2013). This credible messaging (signaling) underlies intentions from one person to another; thus, it undergirds this study of gang affiliate communication and its relationship to violent crime (DeWitt 2018). For example, at least 240 shootings and 24 homicides began as virtual fights in New York City, confirming that social media amplify and accelerate conflict (NY Crime Commission 2017).

Signals can take many forms. For instance, Twitter-based social media messages with video attract 10 times more engagement than those without video (Hootsuite 2020). Tweets with angry text are signals that express rage toward police officers via

#F**kDaOpps (considered the opposition) (Balasuriya et al. 2016). Grammar, punctuation, and expressions ending with K (e.g., CPDK (or CPDKKK, a more aggressive version of CPDK) is Chicago Police Department killer; BDK is Black Disciples Killer; GDK is Gangster Disciples Killer) are key to understanding the post's meaning. Text content such as, "I'll meet you by CH23," means gang members will fight at a local high school named Farragut (Patton, Eschmann, and Butler 2013), and 069 refers to the street address (6900 block of South Princeton Avenue in Englewood) of a murder location (Balasuriya et al. 2016). If a rival gang member makes a threat, the recipient thinks he must defend himself or others will perceive him as weak, making him more susceptible to violence off-line. A very provocative threat is to disrespect (diss) a recently killed gang member, often resulting in retaliation due to the insult being mass-broadcasted quickly via social media. Gang conflict mediators reveal that when affiliates are disrespected on that level, they would feel like they must act and follow-through by committing the crime (Tarm 2018).

Emojis are also important signaling contributors. For aggression, common emoji signals include the gun (pistol), especially when used in conjunction with a bomb, an angry face, a person running, a guardsman, or a police emoji in an 'emoji chain' (Balasuriya et al. 2016), the fuel pump (selling/using marijuana), and the 100 emoji. The term *emoji* means symbols rendered as tiny inline pictures, while an *emoticon* is a face or representation built mostly with conventional punctuation symbols (Owoputi et al. 2013). Thus, tweet content, whether text, emojis, or a combination of both, reveals the conveying of a message.

The notion of signaling credibility is also important for messages that may not be provocative but still carry an intention or reliable statement. Researchers who studied credibility for news-related and other tweets (Table 2) have shown that several characteristics of Twitter users tend to spread more credible information. Active users (Castillo, Mendoza, and Poblete 2011), users with many followers (also known as *indegree,* a measure of popularity) (Cha et al. 2010; Gupta, Lamba, and Kumaraguru 2013), and individuals (with many followers and followees) who create newer accounts (Castillo, Mendoza, and Poblete 2011) are considered to have higher credibility. On Twitter, following a person means being informed by (and possibly supporting) that person's tweets. Tweets with negative sentiment (Castillo, Mendoza, and Poblete 2011), tweets that include a URL (Gupta, Lamba, and Kumaraguru 2013; O'Donovan et al. 2012), and tweets with specific booster words, including words like undeniable (Mitra, Wright, and Gilbert 2017) are also considered more credible. Booster words are linguistic terms used to express assertiveness, strength in a statement, or the conviction of a likely outcome (Hyland 2002).

Scholars, including (Mitra, Wright, and Gilbert 2017), who studied tweet credibility perception during rapidly unfolding events, found that certain types of tweets are considered more credible. These include tweets with replies and longer message lengths (perhaps because the longer message lengths provide additional information and the reasoning behind the message) (Gupta, Lamba, and Kumaraguru 2013; O'Donovan et al. 2012). Additional credibility indicators are larger counts of mentions (representing the value of the tweeter's name) and retweets (representing the value of the tweet content) (Cha et al. 2010; O'Donovan et al. 2012). Other researchers, including those studying weather events like Hurricane Harvey, also affirmed retweet frequency as a credibility indicator (Yang et al. 2019).

**Table 2: Studies on Tweet Credibility**

| **Source** | **Domain** | **Variables Associated with Credibility in Tweets** |
|---|---|---|
| (Castillo, Mendoza, and Poblete 2011) | Tweets contrasting credible news topics from conversational topics | Content has negative sentiment<br>Active user<br>Followees Frequency |
| (Cha et al. 2010) | Tweets related to various topics and over various periods to determine measures of user influence | Retweet Frequency<br>Mentions Frequency<br>Followers Frequency |
| (Gupta, Lamba, and Kumaraguru 2013) | Tweets related to the Boston Marathon Blasts on April 15, 2013. | Followers Frequency<br>Replies Frequency<br>Tweet Length<br>Tweet contained a URL |
| (Mitra, Wright, and Gilbert 2017) | Tweets related to rapidly unfolding events | Retweet Frequency<br>Replies Frequency<br>Tweet Length<br>Tweets with booster words such as undeniable |
| (O'Donovan et al. 2012) | Tweets related to emergency and unrest situations | Replies Frequency<br>Tweet Length<br>Retweet Frequency<br>Mentions Frequency<br>Tweet contained a URL |
| (Yang et al. 2019) | Tweets related to the hurricane Harvey event | Retweet Frequency |

Castillo, Mendoza, and Poblete (2011) successfully used automated J48 decision tree classification techniques to contrast credible news topics from conversational topics based on Twitter content and considered tweets having many retweets with one user mention (on one tree level) as credible. These researchers also found that tweets that include positive sentiment, tweets with question marks or smiling emoticons, and situations when a significant percentage of tweets reference a user are usually more associated with non-credible information for news-related tweets. Tweet verbiage with positive sentiment but mocking an event's practicality with words such as 'ha' or with grins or joking are considered less credible, as well as tweets with much higher numbers of retweets, as this may indicate an effort to provoke collective cognition during an emergency or uncertain times (Mitra, Wright, and Gilbert 2017). Thus, though there have been several recent general studies related to social media credibility, this topic is relatively understudied in the criminology literature, and more knowledge is needed about what types of expressions are perceived as credible in street gang communities.

### 2.1.2 Network Embeddedness Theory

Network Embeddedness Theory is another relevant theory in this study of gang-affiliated criminal signaling, as it illuminates what drives an individual to share information in a social network. This theory states that network embeddedness is a shared characteristic between network users (Aral and Walker 2014; Easley and Kleinberg 2010). Originally coined by Karl Polanyi to refer to kinship relationships that define pre-market economies based on redistribution of resources (instead of open exchange), network embeddedness was later redefined to include social capital, advancement, and trust within a social network (Granovetter 1985). A primary requirement for dissemination of social network content is receivers sharing the information they obtain. Embeddedness is a key driver of sharing between senders and receivers (Aral and Walker 2014). Peng et al. (2016) studied the influence of three overlapping social media content sharing measures and found that receivers have a higher probability of sharing content, especially new content, from senders with whom a common set of followers or followees exists.

Embeddedness may exist in any type of network, including within deviant networks like gangs (Hagan 1993), whose members demonstrate and escalate embeddedness by embracing social media to quickly disseminate their content (bragging, taunting, and threatening posts), boost their gang status, share another member’s violent post, or have common followees or followers re-share their original post. Unlike past decades where rival gang members physically entered another gang's neighborhood to mark their presence with graffiti (often at substantial risk), social media allows an individual to do this without being witnessed but still advertising it as a threat. Anyone, including rival gang members, police, and others can freely view Twitter posts without consent (Patton et al. 2014). Twitter data are made freely available through its Application Programming Interfaces (APIs), making it a widely-accepted open data source in studies of social and human relationships (Leetaru et al. 2013). Unlike Facebook and Instagram (which do not allow the use of user-generated data for further aggregated analysis without the user’s consent, even if the data are publicly available), publicly available data on Twitter may be used for aggregated analysis as long as a user’s personally identifiable information is not revealed in the analysis. An ominous reason to use a social media system like Twitter is to establish a cyberbullying platform, fulfilling psychological needs to quickly communicate and be vengeful, malicious, powerful, and status seeking (Lee and Ma 2012), while also avoiding face-to-face contact. This increases embeddedness due to gang member affirmation, as the desired status (“capital") is lacking in real life due to poverty or another reason. These social media expressions show how gangs emphasize, exaggerate, and reveal violence in the physical environment (Storrod and Densley 2017).

The level of embeddedness is determined by whether the social media network is directed or not and can be described by the numbers of common followers (incoming links or users who are attracted to the principal user’s activity). In a directed network, like

Twitter, a person can follow someone without consent; however, mutual followers (a two-way link) are only established when users have mutual interest.

### 2.1.3 Uses and Gratifications Theory

A third theory affects the study of gang-affiliated violent crime and its relationship to social media. Uses and Gratifications (U&G) Theory states that psychological and social needs affect gratification needs and communication motives (Granovetter 1985; Rubin 2009a) and that various media vie for users' attention, with users selecting the one that satisfies their needs for emotional connection or status (Chen 2011; Tan 1985). U&G Theory asks what individuals do with media, instead of what media does to an individual (Swanson 1979) and explains the phenomenon of how and why people use media to gratify an addictive need and identify consequences from the need, especially those related to communication on a mass scale (Rubin 2009b). It follows that people who use Twitter the most are satisfied in some way by the experience. This connection represents informal solidarity originating from the need to belong (Maslow 1987) and the need to affiliate (Murray 1953).

Recently, researchers have employed U&G Theory successfully to study web usage (Ko 2000). Others have used it to research online games (Wu, Wang, and Tsai 2010), blogging (Chung and Kim 2008; Hollenbaugh 2010), and social media including Twitter (Johnson and Yang 2009), Instagram, Snapchat, YouTube, Facebook (Bumgarner 2007), Kik, WhatsApp (NBC 2018), and MySpace (Raacke and Bonds-Raacke 2008). U&G Theory highlights social and psychological desires and explains the phenomenon of people exploiting media to gratify needs and to identify consequences from these needs, especially those related to communication of status and power (Katz, Blumler, and Gurevitch 1974;

Rubin 2009a). It also concentrates on “what purposes or functions the media serve for a body of active receivers” (Fisher 1978, p. 1590). The theory is relevant to online media usage due to the result of Internet communication voiding the traditional sender-receiver model (Ko 2000), as individuals online can easily select the media they want to use based on meaningfulness (Singer 1998) and send and receive messages simultaneously. Thus, this theory is especially appropriate for examining Twitter, as Twitter offers the ability to communicate on a mass scale or simply between as few as two individuals (Johnson and Yang 2009). Though users have varying motivations (e.g., the need to expel negative feelings, social and entertainment needs, cognitive needs, or motivations related to affection, recognition, and status), studies using U&G Theory find that social media mitigates loneliness and gratifies an addictive urge, though this can differ by gender, location, audience, and narcissism categories.

### 2.1.4 Social Cognitive Theory

Building on U&G Theory, Social Cognitive Theory (SCT) posits that individuals learn by viewing others within the context of social interactions, experiences, and media influences (Bandura 1977). U&G Theory may not solely illuminate a user’s impetus because its premise is that a user regularly selects and uses media, whereas their involvement may be because of their previous familiarity. This necessitates the integration of “theoretical perspectives from SCT with perspectives from U&G Theory to also examine the role of prior experience” (Lee and Ma 2012, p. 332). Ormrod (2012) further explains that SCT provides a lens for interpreting, forecasting, and shifting human behavior. It helps distinguish *gratifications sought* from *gratifications obtained* for media consumption and explains behavior via the interconnection between individuals and behaviors. Thus, this theory posits that gang members learn from others *how* to increase status and gratification, furthering their embeddedness. It clarifies why gang affiliates do not seek mere gratification and status via boastful or bragging social media posts but deploy them by committing the crime due to the high reputational cost associated with not obtaining the gratification (Lee and Ma 2012).

## 2.2 Social Media Literature

In IS literature, the study of crime has included online crimes of identity fraud (Jamieson et al. 2012), data hacking (Khanapur and Patro 2015), corporate fraud (Dong, Liao, and Zhang 2018), software piracy (Siponen, Vance, and Willison 2012), contract violation in virtual markets (Pavlou and Gefen 2005), and others. Chan, Ghose, and Seamans (2016) empirically studied the effect of Internet access on racial hate crimes from 2001-2008 and found a positive association between segregated areas of the U.S. and instances of racial crimes by lone shooters. They also highlighted the offline societal challenges that can arise from an increase of online computing access.

### 2.2.1 Hedonic-Motivated System Adoption Model

An IS theory applicable to the study of criminal signaling comes from Social Media literature. In this area, Van Osch and Coursaris (2015) made a disquieting discovery in their meta-analysis of 610 scholarly papers on social media that most (almost 75%) referenced no theoretical foundation. Of the papers that highlighted a theoretical underpinning, none of the employed theories sufficiently explained the adoption of solely intrinsic or hedonic–motivated systems (HMS) such as virtual worlds, gaming, and social media systems. Addressing a similar concern, Lowry et al. (2013) earlier proposed the

HMS adoption model (HMSAM) for systems that individuals employ predominantly to fulfill an intrinsic motivation for pleasure, even deviant pleasure, more than productivity. This model also reinforces the premise that reasons gangs use social media include their intrinsic motivation to increase their sense of competency and autonomy and to satisfy their need for approval (Butler et al. 2002; Deci and Ryan 1995; Foltz 2004; McClure, Scambray, and Kurtz 2009; Ye and Kishida 2003).

In addition to sound theoretical approaches, social media literature reveals how many researchers are also using novel empirical methods to study crime data. These include participatory mapping, volunteered geographic information, big data population estimates, and the increased use of big data via social media systems like Twitter to find gang-related insights (Morselliand and Décary-Hétu 2013). Other examples include a study of Los Angeles-based street gang rivalries (Radil, Flint, and Tita 2010), crowd-sourced Twitter data to estimate mobile crime risk (Malleson and Andresen 2015), and crime, demographic, and business data with a random forest machine learning classifier to forecast non-linear threats of increased violence over census tracts in DeKalb County, GA (Bowen et al. 2018). Researchers also employ other machine learning classifiers in social media research, including logistic regression and neural networks.

The remaining research on social media usage to study gang-affiliated crime is categorized into three groups. These include location analyses, language and sentiment analyses, and automatic gang member identification.

### 2.2.2 Location Analyses

Many crime researchers have concentrated on a physical location mechanism, using theories such as Crime Pattern Theory, which suggest that geographic and temporal features influence the location and timing of crime patterns. Location analyses studies focus on finding the hotspots in urban areas from geo-tagged location data to support predictive methods of imminent crime locations. For example, Wang, Brown, and Gerber (2012) monitored social media communication to track risk behavior trends and predict geographic hotspots using a model that incorporated intelligent semantic analysis of Twitter posts, dimensionality reduction, and new feature selection through Latent Dirichlet Allocation, improving prediction performance of location-based future crime. Gerber (2014) used linguistic analysis and mathematical topic modeling to find discussion topics and integrate them into a crime prediction model. The addition of Twitter data enhanced the accuracy of the crime prediction and identified reasons for performance decreases in a Twitter-based decision support system. Similar research established the importance of identifying hotspots on specific city blocks when targeting the locale of violent crimes and shootings, as criminals often repeat crimes in the same area. In Boston, Yale University sociologists documented that fifty percent of crimes involving guns occurred on approximately three percent of blocks in specific neighborhoods. This indicates that gangs are not arbitrarily scattered, and socio-economic factors make possible their existence in certain areas (Papachristos and Hughes 2015). Because criminals and victims often follow common life patterns, intersections in those patterns may correlate to a higher probability of resultant crime. Researchers studied the representative value physical locations hold as places of group-based involvement and recall (Conquergood 1997) and the value as spaces of financial pursuits (Venkatesh 2000), while other scholars analyzed a more ominous relationship between the physical space and the gangs (Katz and Schnebly 2011). Scholars also suggest that street gangs set meaningful physical spaces (Tita, Cohen, and Engberg

2005) and fiercely defend them (Decker 1996; Horowitz 1983; Hughes and Short 2005; Suttles 1972). Thus, gang-occupied neighborhoods can be exceedingly cruel and treacherous locales (Huebner et al. 2016; Sharkey 2006).

Spatial factors can also influence the choice of crime location, in part to overcome the lack of generalizability issue with hotspot maps due to their focus on specific locations (Chainey, Tompson, and Uhlig 2008). Spatial features include distance to specific places like schools and businesses, intersections, highways, as well as other neighborhood information (Wang and Brown 2012). To include spatial influences, a researcher can model expected payoffs (Helsley and Zenou 2014). Lazzati and Menichini (2016) employed both Rational Choice Theory and Theory of Environmental Criminology in their study. Following the properties of equilibrium of an estimable model's locale that incorporated social interactions, they assumed that each decision is dependent upon other criminal choices. They used a model based on game theory and found that the best crime reduction strategy involves targeting locations with the potential of more crime, as suggested by some policing strategies. Another finding was that excessive implementations of this strategy might result in an unintentional effect of increasing crime.

### 2.2.3 Language and Sentiment Analyses

Language analysis research focuses on studying linguistic analysis or sentiment analysis to develop insights into crime patterns. For example, Wang, Brown, and Gerber (2012) used linguistic analysis, while Scrivens and Frank (2016) employed Part-of-Speech (POS) tagging to detect frequent keywords, calculate keyword sentiment value for webpages using sentiment analysis, and input those into classification models. Blevins et al. (2016) automatically processed Twitter tweets between a female gang member (whose handle was @TyquanAssassin) and others to understand the impetus of exchanges about loss spiraling out of control and resulting in aggression. They created a part-of-speech (POS) tagger from (Gimpel et al. 2011) built on a tweet-specific POS tagger from (Owoputi et al. 2013) for the gang sublanguage, a phrase table that mapped the vocabulary to typical English, and a linear-kernel Support Vector Machine (SVM) classifier algorithm to find tweets expressing grieving (loss) and aggression. Researchers have further demonstrated the nonstandard language, lack of capitalization use, and abbreviations employed on social media. For example, *“ikr smh he asked fir yo last name so he can add u on fb lololol”* translates as *“I know, right? Shaking my head. He asked for your last name so he can add you on Facebook”* (Owoputi et al. 2013). POS tags were used as classifier features, in addition to quantitative scores to represent word affect. To find and retrieve the correct word in the Dictionary of Affect in Language for each Twitter expression, they used a derived glossary to discover the traditional English words that matched to the slang terms (Whissell 2009). Yadav, Sharan, and Joshi (2014) represented text as a graph in which nodes represent linguistic entities such as words and sentences and the edges represent entity relationship.

Tian et al. (2017) used another modern approach for representing streaming text in natural language processing (NLP), fixed-length vector word representations, known as distributed (low-dimensional) representations (e.g., word embeddings). In this technique, an algorithm learns word vectors for a vocabulary by the words’ context (Mikolov et al. 2013), and then sentence-level representations are extended (Socher et al. 2013). For example, a seven-word sentence using a 150-dimensional vector would have a 7x150 matrix as input. A weighted averaging vector of all the words in the sentence represents the document. Here, the main goal is to improve the efficacy of text classification and sentiment analysis, while also employing text compression at a quick rate and preserving the statistical properties of the sample.

In recent years, researchers have applied neural network models to various NLP tasks, including Twitter text mining, with promising results. Convolutional Neural Network (CNN) algorithms, traditionally employed in computer vision, use convolution to multiply a matrix of pixels with a filter matrix or ‘kernel’ and sum the multiplication values before sliding to the next pixel and repeating the same process until all the pixels have been covered. Chang et al. (2018) used this approach with data from (Blevins et al. 2016) research to predict loss and aggression in gang-affiliated Twitter posts. These researchers did not rely on dictionaries but instead leveraged a large unlabeled location-specific dataset to automatically compute domain-specific embeddings and induce a lexicon using a CNN algorithm. They extracted a phrase table, domain-specific POS tags, and emotion features, trained on a larger dataset, and pruned the feature space to perform feature selection. They tuned the class weight for aggression optimally as “2”, for loss as “1”, and for other as “.12”. This approach allowed them to investigate the context, as well as the emotional and semantic content, of the users’ recent tweet history, including pair-wise exchanges with other like-minded individuals. It improved the results by correctly classifying some tweets as containing aggression or loss, while their baseline linear-kernel SVM model misclassified those (Chang et al. 2018).

### 2.2.4 Methods Used to Automatically Identify Gang Members

Some studies on criminal gang activity have used an architecture with Twitter that requires researchers to know the gang profiles *upfront*. That is, the algorithm does not discover them. For example, some researchers selected known public gang members, obtained their official profile ID, and collected their tweets. Patton, Eschmann, and Butler (2013) discovered that hip-hop music shared on social media targeting affiliates from rival gangs often resulted in real interaction among gang members. Rappers have IDs that are easily retrievable from online searches or certain online websites (Table 3). Though time-consuming, this task could potentially work if a researcher used a known dataset of individuals with gang-related activity on similarly dated tweets and used location-specific content to train a classifier to locate gang-affiliated individuals; however, it would not be effective across all of the Twitter usage areas.

Another way to procure gang profiles is to get permission to retrieve them from governmental databases. Some cities use a gang database based on Structured Query Language (SQL) like GangNet to capture gang members. However, Chicago recently (2019) removed its Sheriff's Office gang database due to controversy over inaccuracies and other issues. Other areas (like Maryland, Washington D.C., and Virginia) do not allow anyone except police to use the database.

However, an alternative method is required if the researcher does not know the profiles a-priori. Thus, an architecture to identify gang members using Twitter data is to text mine and look for *features* representing gang activity (Gerber 2014), since street gang members can express a mood, emotions, and other content on social media, using text, pictures, and emojis.

**Table 3: Data Sources**

| **Type of Data** | **Website / Source of Data** |
|---|---|
| Existing and verified 2014 dataset of tweet IDs of Twitter users associated with street gangs in Chicago, IL (Chang et al. 2018). Used to derive *GangActivity* (1/0) at the individual level. | https://github.com/serinachang5/contextifier |
| Twitter tweet content:<br>Used to derive aggression and loss features from both text and emojis at the individual level. | Twitter tweet |
| Retweet Frequency (*RETWEETFREQ*):<br>Defined as the retweets count per tweet (Hoang and Lim 2011; Yang and Counts 2010). | RetweetFreq (Twitter metadata) |
| Mentions Frequency (*MENTIONSFREQ*):<br>Defined as the @-mentions count (the number of times another user is mentioned in the tweet) per tweet (Hoang and Lim 2011; Yang and Counts 2010). | Custom feature (calculated) |
| Favorites (Likes) Frequency (*favorites_count*):<br>Defined as the number of times another user "liked" or "favorited" the tweet. | Favorites_count (Twitter metadata) |
| Retweeting Time (*IDHOURS*):<br>Defined as the time lag (difference) between the original tweet and the first retweet. Research has shown that retweet frequency and retweeting speed are good indicators for information sharing and messages going viral, content sentiment (positive or negative) going viral (Stieglitz and Dang-Xuan 2013), as well as user's influence (Cha et al. 2010; Kwak et al. 2010). | Custom feature (calculated) |
| Auxiliary Predictor (Control) #1:<br>Weather in the city on the day of the tweet (*AVGTEMP) (F)*):<br>The tweet date is obtained from the Twitter metadata and mapped to a national weather historical dataset (publicly available) to extract the mean daily temperature for the city on that day. It is saved as a feature in the dataset. | Twitter metadata;<br>Weather:<br>https://www.wunderground.com/history/monthly/us/il/des-plaines/KORD/date/2015-1 |
| Auxiliary Predictor (Control) #2:<br>Period Crime Rate (*HISCRIMERATE*):<br>The crime rate from the Chicago CLEAR dataset for that period is calculated, the annual population for 2013, 2014, 2015, 2016, 2017, and 2018 is extracted, and the period's crime rate per 100,000 general population is calculated. | Chicago Data (CLEAR) portal is here.<br>http://worldpopulationreview.com/us-cities/chicago-population/ |

**Table 3 Data Sources (continued)**

| | |
|---|---|
| Auxiliary Predictor (Control) #3:<br>Day of Week (*DAY*):<br>The tweet "*created_at"* field includes the tweet day of the week as the first three characters (e.g., "Thu") (Aghababaei 2017). The day (string) is calculated from this date and converted to a numeric day feature (e.g., Sun =1 to Sat =7) to employ in the prediction model. See example: "created_at": "Thu Apr 06 15:24:15 +0000 2017" | Twitter metadata *"created_at"* date and timestamp |
| Auxiliary Predictor (Control) #4:<br>Average Time of Tweet: (TIME) (HH/MM/SS):<br>The tweet "*created_at"* field includes the tweet time as the third part of the string. The tweet time and the "hour" part of tweet timestamp (*HOUR*) are extracted as an auxiliary predictor and average that time for all tweets that day. See example:"created_at": "Thu Apr 06 15:24:15 +0000 2017" | Twitter metadata *"created_at"* timestamp |
| Auxiliary Predictor (Control) #5:<br>Major Event (*MAJOREVENT*) (1/0):<br>The tweet date is extracted (year/month/date/time) and used to manually code whether a major event (external or local) occurred on that date. These data are stored in a secondary dataset, *Major Events*. | Manually annotated based on the date;<br>Coded individually (1=Yes; 0=No) |
| Auxiliary Predictors (Control Set) #6:<br>• Counts for violent crime for the seven prior periods (days) (*VCLAG1*, *VCLAG2*, *VCLAG3*, *VCLAG4*, *VCLAG5*, *VCLAG6*, and *VCLAG7*);<br>• Indicator variables to represent quarterly seasonality in the data (*Q1*, *Q2*, and *Q3*). | Created from the results of an algorithm. |
| Violent Crime Count (per period) (*VCCOUNT*) (dependent variable). | Annotated for the training set based on that day's violent crime count from the CLEAR dataset. |
| Gang slang websites are hipwiki.com, urbandictionary.com, and internetslang.com. | http://www.hipwiki.com/BDK-Gang<br>https://www.urbandictionary.com/<br>https://internetslang.com/ |
| Chicago Police end of year crime statistics – 2018. | https://home.chicagopolice.org/cpd-end-of-year-crime-statistics-2018/<br>https://www.policedatainitiative.org/ |
| Rapper websites, useful for retrieving candidate gang affiliate user names. | https://www.ranker.com/list/the-best-chicago-rappers/ranker-hip-hop |

A few researchers have already made strides in this third method of Twitter profile identification. Piergallini et al. (2014) examined data related to identifying individuals in gangs and developed methods to determine gang affiliation by researching gang graffiti style features in online Web forums. Wijeratne et al. (2016) used word vectors to improve profile identification of gang members through social media postings, while Balasuriya et al. (2016) extended this and automatically identified 400 unbiased street gang member profiles on Twitter by employing a city- and neighborhood-agnostic method instead of searching using gang names as keywords. These researchers employed commonly used U.S. hashtags (and their variations) including #FreeDaGuys and #FreeMyNigga (peer jailed gang affiliates) and #RIPDaGuys (grieving deceased gang affiliates). They used a word embedding model to map the identified features types into a smaller feature space, employed APIs to search profile descriptions, used gang hip-hop stars and names of individuals recently murdered, and found others via retweets, followers, and followees. They discovered that gangs used curse words, words related to drugs (e.g., *smoke, high,* and *hit)*, materialistic words (e.g., *got, money, make, real,* and *need)*, words in profiles (and YouTube comments) like *nigga, rip, free, f**k, money, featu, get, gang, sh*t,* and *lil,* and words connected to *gangsta* rap and *hip-hop*. Other researchers found commonly used gang words related to aggression (e.g., *angry*, *opps*, etc.) and loss (e.g., *free, RIP*) (Chang et al. 2018) (Table 4).

While rich in methodology, these studies expose the gap of an integrated theoretical approach, which correctly consolidates applicable theory fragments from each discipline to capture the real-life experiences of gang affiliates and crime, especially using social media. A lack of empirical studies that test credible social media signals for gang-affiliates also exists. Lastly, few studies have combined credible unstructured social media content with viable structured data to forecast a city’s upcoming violent crime count. This study bridged those gaps.

**Table 4: Aggression and Loss Lexicon Full Seed Sets (Chang et al. 2018)**

| **Sentiment** | **Words** |
|---|---|
| Aggression (Taunt and Threat) | “angry, opps, opp, fu, fuck, bitch, smoke, pipe, glock, play, missin, bang, smack, slap, beat, blood, bust, bussin, heat, BDK, GDK, snitch, cappin, killa, kill, hitta, hittas, shooter, tf” |
| Loss | “free, rip, longlive, LL, rest, up, restup, crying, cry, fly, flyhigh, fallin, bip, day, why, funeral, sleep, miss, king, hurt, gone, cant, believe, death, dead, died, lost, killed, grave, damn, soldier, soldiers, gang, bro, man, hitta, jail, blood, heaven, home” |

# CHAPTER 3: RESEARCH MODELS AND HYPOTHESES

Because there is a lack of prior combined theoretical grounding related to gang-affiliated criminal signaling on social media, this study integrated separate (but related) theories in its model. Reasons street gang members use social media align with Hedonic Motivation Theory because the experience fulfills a need for deviant pleasure. This causes the gang affiliate to want to prolong or repeat the pleasure, thus suggesting that the more the individual tweets (greater frequency) the more likely their need for approval could be satisfied.

Since gangs are “social networks that embed their members in deviant routines and isolate them from prosocial arenas” (Thornberry et al. 2003, p. 7), this study employed Network Embeddedness Theory, as it reveals what drives a receiver to share information or follow a user in a social network. Network embeddedness occurs within deviant networks, like gangs, who share their (and others’) posts to increase their embeddedness.

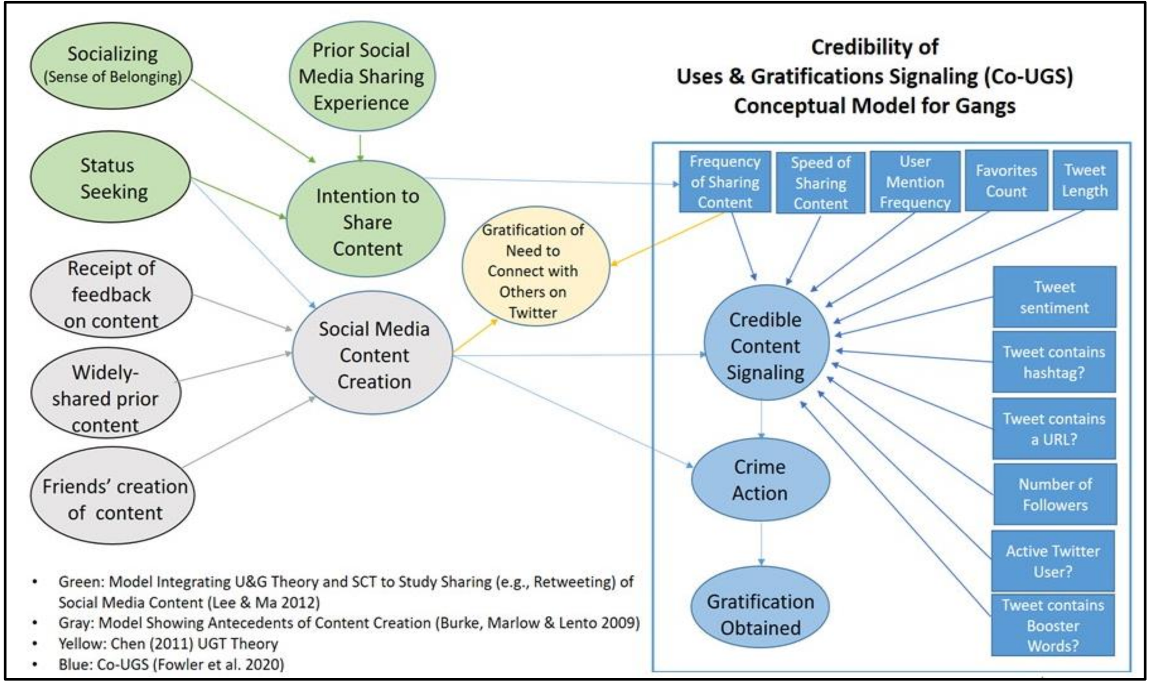
U&G Theory is particularly appropriate for studying gang-affiliated Twitter posts, as it suggests that social media gratifies an addictive psychological urge for gang members and allows them an opportunity to communicate status and power (Rubin 2009a), seek attention, and possibly receive the sought-after attention. U&G Theory undergirds the hypotheses, which are part of a broader theoretical model. Building on U&G Theory, Social Cognitive Theory (SCT) shows how tweets from gang affiliates indicate that the affiliate is not only *seeking* gratification and status via boastful social media posts but also signaling to others the *obtaining* of the gratification via the threat’s imminent deployment and the committing of the crime. SCT provides a context for understanding the general behavior of individuals, especially the interconnections between individuals. Users of social media have an impetus to post original content if their prior content is widely shared, if it receives feedback, or if they see others in their social network post content (Burke, Marlow, and Lento 2009), aligning with SCT. Lee and Ma (2012) integrated U&G Theory with SCT to explain the phenomenon of how and why people use media to gratify an addictive need, mitigate loneliness, allow status-seeking (i.e., seek to get attention), or communicate status and power. This status-seeking, regardless of the motivation, is positively associated with intention to share content mediated by prior social media sharing experience, which then positively affects the *frequency* of content sharing.

## 3.1 Conceptual Model

This research extends the conceptual model by (Lee and Ma 2012) for gangs (Figure 1). Signaling Theory, U&G Theory, and SCT Theory are combined into one framework, the Credibility of U&G Signaling (Co-UGS) Conceptual Model. The study posits that posting of original content or content sharing related to loss or aggression, combined with dissemination factors and other effective predictors, indicates the individual seeking gratification, status, or acceptance and associates positively with credible signaling of imminent deployment, increasing the reputational cost to the gang associate (DeWitt 2018; Lee and Ma 2012). Further, an intention to share the content of others positively affects the frequency of sharing content. If the content turns out to be credible due to the imminent deployment, the individual in turn will be perceived as credible (Rogers 2003), allowing him or her to attain the desired status within the social network.

The framework extends prior research by also assessing for tweet credibility and posits that the combination of aggregated gang-affiliated social media messages of aggression or loss and other variables collectively represents credible signaling to commit a violent crime. The integration of the theories allows the examination and further explanation of the increase in social media cyberbullying as signaling, as well as investigating the credibility of the social media signals.

**Figure 1: Credibility of U&G (Co-UGS) Signaling Conceptual Model**


Credibility of
Uses & Gratifications Signaling (Co-UGS)
Conceptual Model for Gangs
Socializing (Sense of Belonging)
Prior Social Media Sharing Experience
Status Seeking
Intention to Share Content
Receipt of feedback on content
Widely-shared prior content
Friends' creation of content
Social Media Content Creation
Gratification of Need to Connect with Others on Twitter
Frequency of Sharing Content
Speed of Sharing Content
User Mention Frequency
Favorites Count
Tweet Length
Tweet sentiment
Tweet contains hashtag?
Tweet contains a URL?
Number of Followers
Active Twitter User?
Tweet contains Booster Words?
Credible Content Signaling
Crime Action
Gratification Obtained
Green: Model Integrating U&G Theory and SCT to Study Sharing (e.g., Retweeting) of Social Media Content (Lee & Ma 2012)
Gray: Model Showing Antecedents of Content Creation (Burke, Marlow & Lento 2009)
Yellow: Chen (2011) UGT Theory
Blue: Co-UGS (Fowler et al. 2020)

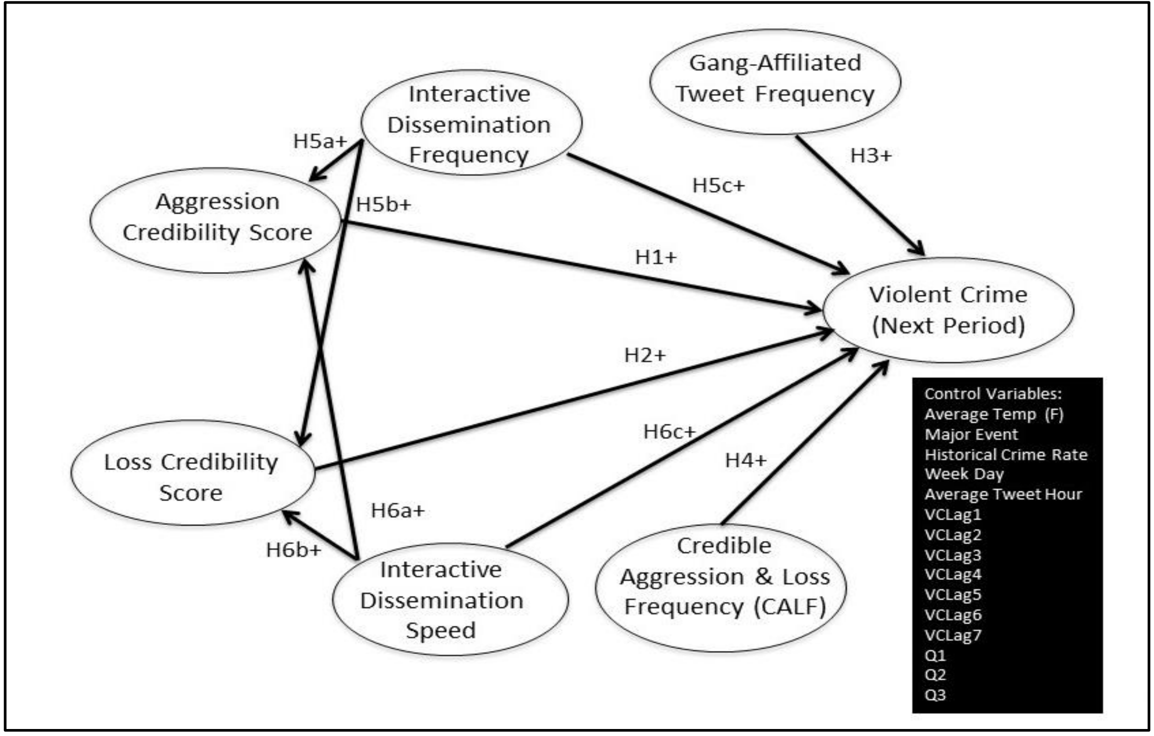
## 3.2 Research Model

Based on this theoretical integration, the research model takes different pieces of the theories, and then operationalizes and tests the constructs with a new understanding of this interdisciplinary domain. Individual tweet content is categorized to determine dimensions (themes) of aggression (taunt or threat) and out-of-control loss from another gang member's death. One instance of expressing loss is from a pensive male gang member who tweeted about his chances of dying violently, with one of his last tweets including a sad-face emoji and text "Death Gotta Be Easy Because Life is Hard" (Tarm 2018, p. 2). Blevins et al. (2016) researched tweets between a female gang member and others to understand the impetus of exchanges about loss spiraling out of control and becoming aggressive. Similar negative individual socio-behavioral signals are classified, aggregated, and used with other determinants to predict the violent crime count in a city for the next time period.

Violent crime index prediction, like the prediction of other non-deterministic signals, is difficult. For example, a precise prediction of 30 violent crime incidents in a city within 24 hours is challenging. Thus, some researchers have opted to predict the next-day crime *trend*, where the trend is defined as the direction (the sign of change) at a specific time when compared to another previous time. A positive change means that the construct has a rising trend. Aghababaei (2017) explored and found a strong correlation between content, sentiment, topics as features, and auxiliary data and a city-based crime index trend. The model annotated its training data, found collective patterns, evaluated the performance of both auxiliary features and content-based features in predicting crime rate directions, and found that the content-based features provided the best predictive power.

However, instead of predicting a crime trend, this study's model predicts violent crime instances for the next period at the city level (Figure 2). The period (time unit) in this study is defined as a 24-hour day. Though the FBI has a broader definition (FBI UCR Program), a violent crime in this study is defined as any criminal act involving force or the threat of force and composed of one of the following offenses: homicide, assault with a deadly weapon (including gun violence), and criminal sexual assault. These offenses involve the following IUCR codes for each offense: Homicide (110, 130, 141, 142); Assault with a Deadly Weapon (051A, 051B, 520, 530, 545, 550, 551-560); and Criminal Sexual Assault (261-266, 271-275, 281, 291) (FBI 2018) (Appendix A).

**Figure 2: Research Model**



Because the dissemination of social media-based intentions or events can signal forthcoming violent crime, the Co-UGS framework incorporates gang-affiliated credible signals. These signals convey underlying intentions in a message to another person via social media to advertise and amplify aggression or out-of-control loss that could escalate into aggression. This is especially true for provocative messages to other gang members who then feel they must act and subsequently follow through by committing the crime. From a social capital perspective, the social media posting is a signal of gratifying a gang members' psychological need to quickly communicate and increase status and power by threats, increasing embeddedness via gang member affirmation. From a social media perspective, gang members escalate their embeddedness when they share another member's violent post or when others re-share their original posts.

This research focuses on how tweets provide socio-behavioral signals to predict violent crime based on information observed from previous credible tweet content, interactive dissemination determinants, historical crime violent instances, and other auxiliary variables. A credible tweet is defined as one that includes content with negative sentiment from an active user (Castillo, Mendoza, and Poblete 2011) and includes additional determinants such as higher interactive dissemination frequency (retweet frequency plus favorites (likes) count), faster dissemination (retweet) speed, higher user mention counts, more followers (Gupta, Lamba, and Kumaraguru 2013), and a longer post (tweet) length. Tweet credibility is further assessed by determining whether the tweet contains a hashtag (which makes the tweet searchable), whether it contains booster words, and whether it contains a URL. A formative construct, Credible Content Signaling, and exploratory factor analysis are used to confirm the lack of correlation between indicators, making each a contributor to credibility. The relative influence of the credibility indicators on a dependent variable is determined by finding the beta weights via regression and classification.

The period's credibility scores for each aggressive tweet are summed as an input into the model and the period's credibility scores for each loss tweet are summed as another input. This study also determines whether the frequency and speed of the interactive dissemination of tweets have direct or indirect effects on violent crime. Interactive dissemination (engagement) is defined as retweeting, selecting the tweet as a favorite (also known as liking the tweet), or mentioning another Twitter user (@-mention) within the tweet content.

## 3.3 Hypotheses

Tweet signaling of aggression or loss aligns with the Credibility of U&G Signaling (Co-UGS) conceptual model for gangs because social media offers a new way to advertise and amplify signals of a gang member's anger or loss via combinations of content in emojis and text. The tweet content represents negative sentiment, which is positively correlated with credibility. The posting of credible content related to loss or aggression indicates the individual seeking gratification, status, or acceptance. Obtaining the desired gratification and status necessitates the individual committing the crime to accomplish the anticipated status within the social network and thus increases the likelihood of violent crime, justifying the positive association between the hypotheses constructs. The argument is that tweet content does not cause the crime but includes signals to predict future crime incidents. Therefore, hypotheses 1 and 2 (Table 5) are proposed as follows:

**H1:** There is a positive association between the credible social media content signaling aggression posted by street gang members and next-period violent crime at the city level; and

**H2:** There is a positive association between the credible social media content signaling loss posted by street gang members and next-period violent crime at the city level.

We also study the overall tweet frequency of gang affiliates in this model. Tweet frequency aligns with Network Embeddedness Theory because the more embedded in the gang a poster becomes, the more cost and credibility are associated with the content of each post and the more likely the deployment (DeWitt 2018). Further, tweet frequency aligns with Hedonic Motivation Theory, as individuals employ social media systems predominantly to fulfill an intrinsic motivation for pleasure, even deviant pleasure, more than productivity due to a need to increase their sense of competency, autonomy and approval. Aghababaei (2017) quantitatively confirmed this premise by evaluating the performance of social media content in predicting crime rate direction and found that content-based features and tweet frequency provided the best predictive power. Gerber (2014) also revealed an association between tweet density and crime rate at a location. Thus, hypothesis 3 is proposed as follows:

**H3:** There is a positive association between the frequency of all tweets of street gang-affiliated users and next-period violent crime at the city level.

Both Network Embeddedness Theory and Hedonic Motivation Theory also support our next hypothesis, which tests the frequency of tweets containing aggressive and loss-filled content and its relationship to violent crime. Building on the work of other researchers, we posit that the greater the number of tweets of aggression and loss by street gang affiliates, the greater the count of violent crime (Aghababaei 2017; Aghababaei and Makrehchi 2018). Therefore, hypothesis 4 is proposed as follows:

**H4:** There is a positive association between the number of tweets posted by street gang members with credible taunts and threats (aggression) and expressions of out-of-control loss and next-period violent crime at the city level.

This study also examines whether the negative sentiment in gang members' social media messages and their related interactive dissemination (engagement) behaviors indicate violent crime. The next two hypotheses are based on Network Embeddedness Theory, which explains both the original posting of social media content and the sharing of content (interactions) via retweets, favorites, and user mentions. This is because the desired increase of gang embeddedness drives content dissemination via interactions, which, in turn, increases the cost of not following through with the threat stated in the post. Interaction behavior is characterized in terms of 1) the quantity of triggered retweets, the frequency of user mentions in the tweet, and the count of favorites (likes) of the tweet from others; and 2) the speed of retweeting as the difference (time lag or delayed effect) between an original tweet and the first retweet. Retweet frequency and speed are good indicators for information sharing and content sentiment going viral (Stieglitz and Dang-Xuan 2013).

Network Embeddedness Theory illuminates embeddedness as a key driver in an individual sharing information with other users in a social network. From a cyberbulling platform, members demonstrate and escalate their embeddedness by embracing social media to quickly disseminate taunting or threatening posts and boost their gang status. This fulfills psychological needs to quickly communicate and be vengeful, malicious, or powerful, and increase social capital and embeddedness (due to gang member affirmation). The first retweet is the most important tweet because initial tweets increase exposure and thus spawn additional retweets. At five retweets, there is a 50% chance that the tweet will be retweeted another time; by ten, the chance increases to 90% (Bild et al. 2015). If a tweet is not retweeted within 60 minutes, it most likely will not be retweeted (Sysomos 2010). The average retweet speed (half-life) is most commonly cited as 18 minutes (Bray 2012). The frequency of favorites and user mentions is used, where the @ sign is used to indicate the recipient when initiating messages in a post, along with the frequency of retweets, as a proxy for interest in a topic or another gang member. Thus, hypotheses 5c and 6c are proposed as follows:

**H5c:** There is a positive association between the frequency of interactive dissemination (retweet frequency, user mention frequency, and favorites frequency) of street gang-affiliated tweets and next-period violent crime at the city level; and

**H6c:** There is a positive association between the interactive dissemination (retweet) speed of street gang-affiliated tweets and next-period violent crime at the city level.

The study also examines whether the interactive dissemination constructs are mediated (partially or fully) by credible aggression or credible loss. These next two hypotheses are based on Network Embeddedness Theory, which explains the sharing of content (interactions) and the speed of that sharing via retweets, favorites, and user mentions. These two constructs are special as they both have a dual purpose as not only predictors to credibility but also as predictors to violent crime. Thus, hypotheses 5a, 5b, 6a, and 6b are proposed as follows:

**H5a:** There is a positive association between the daily frequency of interactive dissemination of street gang-affiliated tweets and credible aggression at the city level; and

**H5b:** There is a positive association between the daily frequency of interactive dissemination of street gang-affiliated tweets and credible loss at the city level; and

**H6a:** There is a positive association between the daily average speed of interactive dissemination of street gang-affiliated tweets and credible aggression at the city level; and

**H6b:** There is a positive association between the daily average speed of interactive dissemination of street gang-affiliated tweets and credible loss at the city level.

**Table 5: Summary of Research Model Hypotheses, Theories, and Constructs/Effects**

| **Hypothesis** | **Theory** | **Construct → Effects (City Level)** |
|---|---|---|
| H1 | Co-UGS (Credible Signaling Theory, U&G Theory, SCT) | Aggression Credibility Score (per period) → Violent Crime Count (per period) |
| H2 | Co-UG (Credible Signaling Theory, U&G Theory, SCT) | Loss Credibility Score (per period) → Violent Crime Count (per period) |
| H3 | Hedonic Motivation Theory, Network Embeddedness | Gang-Affiliated Tweet Frequency (per period) → Violent Crime Count (per period) |
| H4 | Hedonic Motivation Theory, Network Embeddedness | Credible Aggression and Loss Tweet Frequency (per period) → Violent Crime Count (per period) |
| H5c | Network Embeddedness Theory | Interactive Dissemination Frequency (Retweets, Mentions, and Favorites, per period) →<br>Violent Crime Count (per period) |
| H6c | Network Embeddedness Theory | Interactive Dissemination Speed (Retweet Speed, measured as retweet time, in hours, per period) → Violent Crime Count (per period) |
| H5a | Network Embeddedness Theory | Interactive Dissemination Frequency (per period) → Aggression Credibility Score (per period) |
| H5b | Network Embeddedness Theory | Interactive Dissemination Frequency (per period → Loss Credibility Score (per period) |
| H6a | Network Embeddedness Theory | Interactive Dissemination Speed (Measured in Retweet Time, in Hours) (per period) →<br>Aggression Credibility Score (per period) |
| H6b | Network Embeddedness Theory | Interactive Dissemination Speed (Measured in Retweet Time, in Hours) (per period) →<br>Loss Credibility Score (per period) |

## 3.4 Control Predictors

Several IS and CJ researchers have added auxiliary determinants to content-based features to predict crime. These predictors included unemployment rate (Aghababaei 2017; Cohen and Felson 1979; Raphael 2001); weather (Aghababaei 2017; Anderson 1987; Chen, Cho, and Jang 2015); historical crime rates (Aghababaei 2017); day of the week (Aghababaei 2017); and emerging (events) days before and after a holiday, political election, or major sporting event (Aghababaei 2017). Aghababaei (2017) evaluated the performance of auxiliary features as well as content-based features in predicting crime rate directions and found that unemployment rate and events did not achieve high performance compared to the daily number of tweets.

Similar to prior research, this study includes historical violent crime rates and day of the week as control determinants in the model. The tweet's time (hour) of the day is captured from the tweet Timestamp. The study also extends previous models by substituting more granular weather averages (instead of aggregated monthly ones) at the city level. The researcher analyzes whether there was an external or local major event ("1"=Yes; "0"=No) at the time of the tweet. A major event is a social, political, racial, or cultural trigger that may cause friction, riots, or concern within a gang community where members hold a position for (or against) the event. Examples include a publicized police killing of an unarmed African-American individual from a local (or other) community, a local concert of a famous rapper, a popular gang member death, a political election, or a controversial criminal verdict. Major events can include holidays, major sporting events, and anniversaries of death or birthdays of gang affiliates. A major event could also include an impactful Twitter policy change, such as the April 21, 2015 prohibition of not only "threats" of violence against others but also "promoting" violence against others or the April 20, 2015 update allowing users to "direct message" a user who was not already following them. The latter Twitter policy change allows a gang member to send another gang member a direct message, which may aggravate the impact of tweets on crime.

This study extends other models by including indicator forecasting control variables. These include the violent crime counts of seven prior periods (where a period is a day), as well as seasonal indicator variables to represent the distinct quarterly seasonality in the data.

## 3.5 Additional/Supplementary Analyses: Tests of Mediation and Direct Effects

The research model posits that Interactive Dissemination Frequency and Interactive Dissemination Speed have a direct association with Violent Crime Count. Further, it suggests that there is a possible mediation role of Aggression Credibility Score between both Interactive Dissemination Frequency and Interactive Dissemination Speed to Violent Crime Count. There is a similar possible mediation role of Loss Credibility Score between both Interactive Dissemination Frequency and Interactive Dissemination Speed to Violent Crime Count. There is a theoretical justification for both Interactive Dissemination Frequency (Cha et al. 2010; O'Donovan et al. 2012; Yang et al. 2019) and Interactive Dissemination (Retweet) Speed (Cha et al. 2010; Kwak et al. 2010) as predictors to credible aggression, credible loss, and violent crime (DeWitt 2018; Lee and Ma 2012). Thus, additional tests were performed for these possibilities.

### 3.5.1 Supplemental Test for Mediation

The research includes a supplemental test which analyzes mediation (Stylianou, Subramaniam, and Niu 2019; Subramani 2004) and assesses whether the interactive dissemination frequency and interactive dissemination speed (measured as interactive dissemination time, in hours) are *direct effects* or merely *indirect effects* on violent crime in the next period. In the model, Credible Aggression (measured as Aggression Credibility Score) was the first mediating construct between Interactive Dissemination Frequency and Violent Crime Count and between Interactive Dissemination Speed and Violent Crime Count.  The second mediating construct was Credible Loss (measured as Loss Credibility Score) between Interactive Dissemination Frequency and Violent Crime and between Interactive Dissemination Speed and Violent Crime Count (the next day).

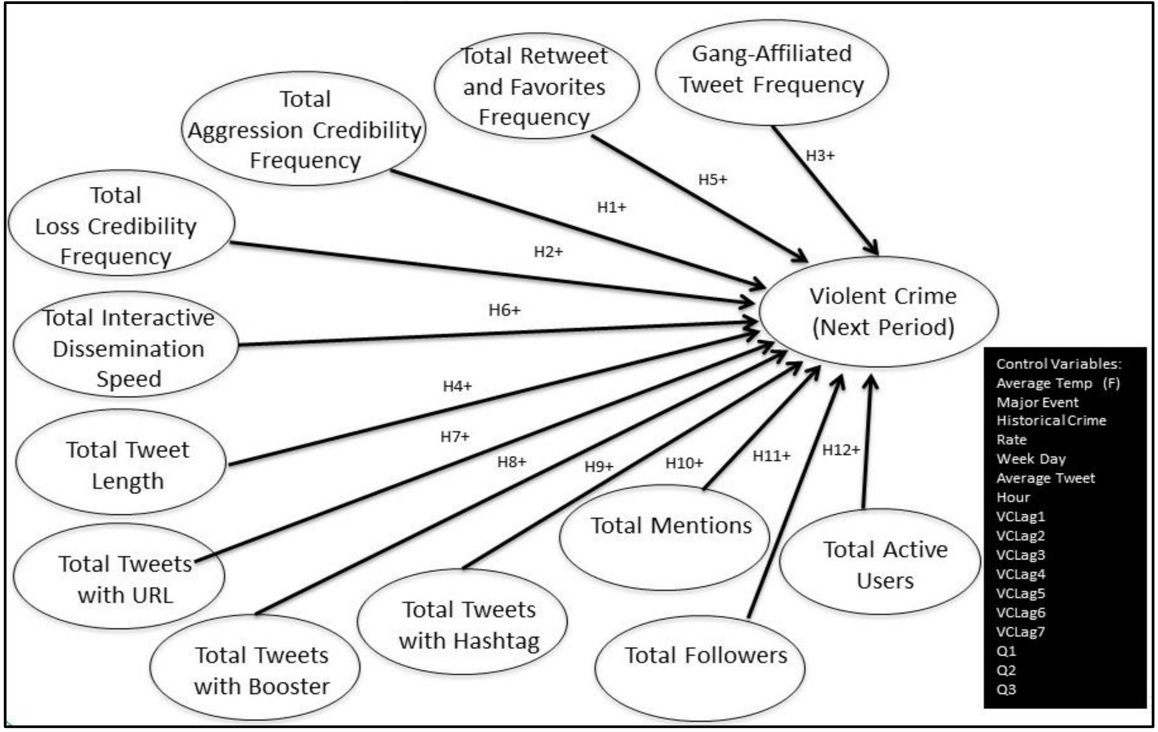
Thus, two separate analyses were conducted. The first was a partial mediation test by incorporating a direct path from the independent variable (Aggression Credibility Score) to the outcome variable, which is contrasted with a full mediation test (i.e., without the direct effect). Similar tests between the variables and Violent Crime Count (next period) were performed using Loss Credibility Score as a potential mediator, and the results were evaluated.

### 3.5.2 Direct Effects Model

Another supplemental approach to the research model was tested by incorporating a direct path from each credibility indicator to the outcome variable (Figure 3). This model was compared with the prior competing research model that included mediation. This alternative approach assessed whether all of the credibility indicators used in the Co-UGS

Conceptual Model to produce the formative Credible Content Signaling construct were better as direct effects on violent crime rather than aggregated into the credibility construct.

**Figure 3: Direct Effects Model**



Before running this model, the data were preprocessed by classifying aggression and loss in each tweet. Next, two other classifiers for credibility were employed. For the aggression credibility algorithm, nine credibility variables were used and classified each tweet as either containing Credible Aggression (1) or not (0). For the loss credibility classifier, nine credibility variables were used and classified each tweet as either containing Credible Loss (1) or not (0). The credibility factors derived from a preprocessing step. The variables were then aggregated into two sets of daily amounts (one set for aggression; the other set for loss). These variables included:

1) Total Retweet Frequency and Favorites Frequency (per period) of tweets with credible content;

2) Average Retweet Time (hour) (per period) of tweets with credible content;
3) Total User Mention Frequency (per period) of tweets with credible content;
4) Total Tweet Length (per period) of tweets with credible content;
5) Total Frequency of Credible Tweets with Booster Words (per period);
6) Total Frequency of Credible Tweets with a Hashtag (per period);
7) Total Frequency of Credible Tweets with a URL (per period);
8) Total Number of Followers of Authors of Credible Tweets (per period); and
9) Total Frequency of Active Twitter Users of Credible Tweets (per period).

These variables were inputted into an algorithm to determine their impact on violent crime. The other inputs to the violent crime predictive algorithm included the frequency of gang-affiliated tweets with credible aggression per period, frequency of gang-affiliated tweets with credible loss per period, frequency of all gang-affiliated tweets per period, prior period (time series) violent crime counts, quarterly seasonal variables, temperature, major event, tweet day of the week, tweet hour, and historical crime rate per 100,000 individuals.

# CHAPTER 4: METHODOLOGY

In this study, the researcher drew upon theories and empirical results from past research, particularly related to the fields of social media systems and social psychology. Applying these techniques to the field of criminology is a comparatively new frontier in IS research, especially in the area of social media.

The first part of this chapter discusses the general research design for the study. The second part reveals the ethical guidelines used in the research. The third part highlights the data types and data sources of the 2014-2018 data used in the study. The fourth part reveals the procedures involved in classifying the tweets of gang-affiliated posts related to loss and aggression. Further, this part reveals the methods used to combine the tweet content from these individuals with other factors to determine credible criminal signaling. This section also discusses the methods used to aggregate the credible signals to a daily level and combine these aggregated variables with various other predictor variables as input to a prediction algorithm to forecast the city's next day violent crime count. The fifth part of the chapter reveals the evaluation procedures used in the study.

## 4.1 Research Design

The research design consisted of two stages. Stage 1 involved the use of social media data at the individual level to detect credible aggression and loss by gang affiliates. Stage 2 involved the aggregation of such data into daily amounts, the testing of all research model hypotheses, and the prediction of violent crime counts at a daily city level.

This study focused on the social media platform Twitter. Twitter has smaller content than a traditional blog in both actual and aggregated file size and only allows users to exchange tweets as short messages (up to 140- or 280-characters). This enables users to share and relate via mutually interesting topics with a group of “followers” in real-time. These tweets consist of text and images and are considered unstructured data. Automatically analyzing unstructured data from social media (including Twitter) does not come without its challenges. Social media data are “exceptionally noisy and contain a great deal of grammatical variance, misinformation, and mundane chatter” (Burnap and Williams 2015, p. 225). Examples include Twitter user-generated comments like ‘hahahaha’; thus, one cannot employ typical methods for grammatical analysis like those used in traditional well-edited text varieties (e.g., newswire). The poor veracity of raw data hinders its trust by decision-makers; thus, scholars sometimes translate Twitter data to Standard American English (SAE) with a phrasebook.

In the context of studies of gang-affiliated individuals, another complicating factor is the use of gang slang on Twitter (e.g., *BDK (Black Disciples Killer), GDK (Gangster Disciples Killer),* etc.). Because gang member content includes keywords and phrases of the specific context that change quickly, researchers may employ a corpus dictionary of slang needed by text-mining algorithms, like general slang translators (NoSlang 2018b) and drug slang translators (NoSlang 2018a), to decipher social media content (Han and Baldwin 2011). Researchers have also experimented with crowd-sourced knowledge bases and gang slang websites to automatically extract gang-related slang and names, as integrating domain-specific knowledge into various machine learning models has been associated with increased performance in prior research (Sheth et al. 2017).

This study of unstructured data employed the use of NLP techniques. With NLP techniques, researchers can mine unstructured textual and graphic content, glean the textual content and signals, convert the content into features, and assess them based on their ability to provide source information to predict outcomes (Dong, Liao, and Zhang 2018). Traditional algorithms employ bag-of-words models and n-gram models, statistical language models widely used for extracting features from text and estimating the probability of each word given prior context. A bag-of-words model represents text with numerical features given a corpus (collection of texts) (C) of documents (D) and unique tokens (N) extracted out of the corpus. A list is formed from the tokens. The size of the bag-of-words matrix (M) is D x N, where each row in M contains the token frequency in the document. With Twitter data, n-grams are the contiguous sequences of n words that appear in a tweet text fragment. In this context, a dictionary could be the list of all unique tweet words and may look like the following: [*'Free', 'my', 'homie', 'or', 'Ima', 'kill'*].

The traditional and most basic method to represent text data numerically is with one-hot encoding (OHE) (or count vectorization). A traditional vector representation of a word includes a one-hot encoded vector, where "1" represents the position where the word exists and "0" stands for all other positions; thus, the vector representation of '*kill*' according to the dictionary mentioned earlier is [0,0,0,0,0,1]. Though useful, these methods create sparse large-dimensional models that cannot analyze context (words surrounding a given word). Another option to create word embeddings is via the Singular Value Decomposition (SVD) method, which outperformed other methods in research on historical English and community-specific sentiment lexicons (Hamilton et al. 2017). Lilleberg, Zhu, and Zhang (2015) showed that word vectors weighted by term frequency–inverse document frequency (TF-IDF), which mirrors the word's ranking in a document, outperformed other word embedding model variations, after training them on newsgroup posts.

In recent years, neural networks improved classification, as they learn enriched word representations from a text corpus, unlike traditional bag-of-words and n-gram models with their data sparsity issue. By representing words as dense vectors (word embeddings, or numerical representations of text) and using vector arithmetic, algorithms can determine similarities between words and other valuable features directly from context and prove beneficial in signifying the meaning of sentences in social media content. In this scenario, words like “bed” and “pillow” have shorter distances to each other than their respective distances to a word like “bus”. What traditionally could not be used with one-hot encoding can now be applied to any corpus, including social media posts, and the result (a word embedding) then used as input (often a matrix, where each row is a vector or word embedding that represents a word) to additional models such as a CNN or SVM. These improved methods allow researchers to “define a word by the company it keeps” (Brownlee 2017, p.1), (Firth 1962, p. 11). The vector space representation (distributional semantic model) clusters words (linguistic items) with similar meanings together within the space. Thus, the goal of word embedding algorithms is to find vectors for the words and their contexts in the corpus to meet some pre-defined criteria (e.g., to predict the context). Wang et al. (2016) showed that word embeddings improve short text categorization and located semantic cliques by utilizing density peaks for searching and clustering.

Prediction-based word embedding models (like those used in this study) can be performed under different neural network architectures, but two common methods include the popular Continuous Skip-gram (Skip-gram) model, which can handle less frequent words (Mikolov et al. 2013), and the Continuous Bag of Words (CBOW) method. Both are known as shallow neural networks. The CBOW architecture predicts the probability of a current target (center) *word* given the source context words (surrounding words), while the Skip-gram model “skips” the current word and predicts *context* words given a target word by considering the order of word occurrences. The result can be more than one word if the context window (number of context words) is more than one word long. The Skip-gram model can also capture two meanings for a single word (AnalyticsVidyha 2017) and represent them as two vectors, which is advantageous for words with multiple connotations. Further, Skip-gram models tuned with the negative sub-sampling parameter generally provide better performance than other methods (AnalyticsVidyha 2017). Thus, in Stage 1 of this study, word embeddings were automatically generated from the unlabeled Twitter corpus that translated the features into real (and dense) vectors cooperative for classification with machine learning algorithms. The result was later used to train a set of supervised algorithms.

An open-source Python framework (Gensim) and Word2Vec were used to produce the word embeddings. Created by researchers at Google, Word2Vec is an unsupervised algorithm that does not require human expert labeling or annotation to learn. Using cosine similarity, Word2Vec uses text as its input, and its output is a vocabulary in which each item has a vector (neural word embedding) attached to it, which is then used as input into a traditional classifier or neural network. The relative meaning (context) of similar words are closer to each other using measurable distances. Word2Vec comprises both CBOW and Skip-gram models and uses a *list of list* format for training. Each document is contained in a list, and each list holds lists of tokens of that document.

To classify the unlabeled tweets, two classifiers were generated: one for aggression and one for loss. Credibility for each tweet was then assessed using other algorithms. In Stage 2 of the study, the features from the tweet data were aggregated to a daily level, and, using various algorithms, combined with other features to detect a city's next day violent crime count.

## 4.2 Ethical Guidelines

The researcher scraped data at the individual level from publicly available Twitter posts. Though the organization's IRB exempted the study's data (due to its public nature), ethical guidelines mandated that several careful steps be taken in providing fair and just treatment of the data and the users from whom it originated. Risk was mitigated by de-identifying each tweet and removing all identifying information (e.g., @-Mentions and uniform resource locators (URLs)) before the publication of any data.

Secondly, community members based in Chicago, IL were included as domain experts in the labeling of training data and the validation of the findings, ensuring high ethical standards and safeguarding the interpretation and dissemination of insights related to violent crime (Frey et al. 2018), as on-the-ground domain experts can more accurately attest to credible gang-affiliated street language (Stuart 2019). The tweets were analyzed by two different youthful and Chicago-based gang-affiliated annotators (hired as research assistants). They used their knowledge about the text and other resources (Whitlock 2020) for emoji interpretation.

The contributions from the domain experts occurred in two key places: 1) annotation and veracity of loss and aggression tweet content for correct *Loss* and *Aggress* feature classifications; and 2) the validation of this content as credible when combined with interaction dissemination frequency, speed of retweets, and other variables. This is important, as there are risks involved with identifying aggression and loss in Twitter tweets using automatic detection methods, including possible misinterpretation and misidentification of tweets due to non-standardized language and emoji usage, mischaracterization of mere boasting and bragging as actual signaling, inaccuracies in raw data, and loss of privacy. Such misidentification could inadvertently influence the results as well as impact groups of individuals in a negative way. Thus, because of this importance, the expertise of domain experts was relied upon to ensure the highest ethical use of the data.

## 4.3 Data

The main data for this study came from Twitter and involved the text (including user mentions and hashtags) and emojis within a tweet, author, retweet frequency, tweet's creation timestamp, count of favorites (likes), and other metadata. The study used tweet IDs of known street gang affiliates in Chicago from 2014, as parts of 2014 were especially violent (Chang et al. 2018).

### 4.3.1 Validated Gang-Affiliated Tweet ID and User IDs

The study used an existing and verified dataset of 4936 tweet IDs from 279 Twitter users (Chang et al. 2018) who were associated with street gangs in Chicago, IL in 2014. From these tweet IDs, the *api.statuses_lookup* API was used to capture user IDs of 92 gang-affiliated Twitter users in the Chicago, IL area affiliated with these tweets. These included top communicators, from whom the confirmed tweets were sourced, making the data sample representative of Twitter dialogs among gang affiliates from Chicago neighborhoods during the research time. Sample bias in the data was mitigated by including statistically proportionate tweets with negative sentiment and positive sentiment. Other data sources included historical crime data collected from the Chicago City (CLEAR) portal, weather data, and data related to major events during the timeframe.

### 4.3.2 Twitter (Tweet Content) Data

Using the verified screen names, 143,700 Twitter tweets of gang affiliates from the Chicago, IL area were scraped and cleaned. Random sampling is a common approach to access streaming data, and many researchers who obtain random Twitter user names or tweets for analysis collect 1% of tweets using the REST API or the Streaming API, as these are efficacious for accessing the historical timeline of random Twitter users. However, this study used the *api.user_timeline* API, as it allows extractions of all tweets of specific gang-affiliated Twitter users in Chicago from 2014-2018.

The rationale for the study's 2014-2018 timeframe is due to the 2014 gang-affiliated tweet dataset. In general, gang text and symbols can change frequently and can vary from gang to gang and location to location. Further, gang members stay in gangs (on average) fewer than three years, though late persistent entry tenure can last up to approximately seven years (Pyrooz 2014), depending on the association level in the gang, including levels of leader, hardcore, associate, fringe, and "wanna-be" (Carlie 2002). This suggests that the tweets safely fall within a five-year timeframe from the collection time of the street gang Twitter tweet ID dataset. Further, the domain experts affirmed that the current gang-affiliated language at the time of the study was the same language used in a 2014-2016 sample of hundreds of tweets they analyzed. From the tweet content, tweet id, tweet text (including emoticons), and emojis were extracted. The data were used for the experiments and for calculating statistical significance.

### 4.3.3 Twitter Metadata

From the Twitter tweet metadata, tweet author (*screen_name*), user mentions, tweet *place*, longitude and latitude geographical *coordinates [LONG, LAT],* and *location* of the tweet source or the user profile were extracted. The number of retweets (*RetweetFreq*) and the count of @-mentions (*MentionsFreq*) were also determined. The researcher pulled the favorites count (*Favorites_Count*), determined whether the tweet had hashtags (*Hashtags* = 1/0), extracted the tweet date (*Created_at*) from the date to calculate tweet day (*Day*), and pulled the timestamp portion of the tweet's *Created_at* metadata to extract the hour (*Hour*).

### 4.3.4 City of Chicago CLEAR Dataset

The researcher extracted 2014-2018 raw and granular crime data from the City of Chicago portal and the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system, as these data reflect incidents of crime that transpired in Chicago reported by law enforcement. Due to privacy concerns of crime victims, the data display addresses at the city block level only and do not identify specific locations. Data include the following variables about each specific crime: the time (year, month, day, and hour); location (district, sector, beat, ward, community area, city block, and latitude/longitude); FBI Code; index; IUCR code; whether the crime was domestic or non-domestic; and whether the crime resulted in an arrest. Individual crime counts from Chicago via the city's data portal containing crime records on a time basis were retrieved.

From those, violent crime counts per day were aggregated. The daily crime rate per 100,000 general city population was calculated by dividing the city's aggregate daily crime counts by the annual population per city, and the results were stored in a secondary dataset.

### 4.3.5 Data for Other Control Predictors and Secondary Data

The data for the auxiliary (control) predictors came from disparate sources. These included data related to weather (temperature) and whether a major event occurred on the day of the tweet (Table 3).

## 4.4 Procedure

The study methods involved retrieving, cleaning, partitioning, calculating, and preprocessing the data as well as a series of word embedding, aggregation, classification, and prediction tasks. In this section, the data preparation steps are summarized, followed by the details of each step.

1. Retrieved the gang-affiliated tweets and retweets from 2014-2018 from the city of Chicago, IL.
2. Calculated the gang-affiliated tweet frequency (per day) and stored it as a feature.
3. Calculated retweet speed and other needed features for each tweet.
4. Cleaned, pre-processed, and tokenized the data.
5. Created a count vectorization model and domain-specific word embedding models to convert the tweet words, emojis, and seed words associated with aggression and loss to numbers.
6. Labeled the tweets in the training set with aggression or loss with the assistance of two domain experts.

7. Classified each tweet in the test set as either containing aggression (1) or not containing aggression (0) (Algorithm 1) using baseline models and neural network models, and then assessed the results.
8. Classified each tweet in the test set as either containing loss (1) or not containing loss (0) (Algorithm 2) using baseline models and neural network models, and then assessed the results.
9. Ran the best model for algorithms 1 and 2 on the unlabeled data and validated the results.
10. Calculated the tweet's Aggression Credibility Score (0-10) for tweets with aggression. Employed user mention frequency, retweet speed, tweet length, whether the tweet contain a booster words (1/0), whether the tweet contains a hashtag (1/0), whether the tweet contains a URL (1/0), the number of followers of the Twitter user, whether the Twitter user is active (1/0), and the sum of retweet frequency and favorites count (Algorithm 3).
11. Calculated the tweet's Loss Credibility Score (0-10) for tweets with loss. Employed user mention frequency, retweet speed, tweet length, whether the tweet contains a booster word (1/0), whether the tweet contains a hashtag (1/0), whether the tweet contains a URL (1/0), the number of followers of the Twitter user, whether the Twitter user is active (1/0), and the sum of retweet frequency and favorites count (Algorithm 3). Here, Algorithm 3 was modified for Loss.
12. Validated the credibility process and results with domain experts.
13. Summed the Total Aggression Credibility Score (per day) for all credible tweets containing aggression.

14. Summed the Total Loss Credibility Score (per day) for all credible tweets containing loss.
15. Used a classifier to predict Credible Aggression Frequency (1/0) (per tweet).
16. Used a classifier to predict Credible Loss Frequency (1/0) (per tweet).
17. Summed the Credible Aggression Tweet Frequency (per day) for all credible tweets containing aggression.
18. Summed the Credible Loss Tweet Frequency (per day) for all credible tweets containing loss.
19. Aggregated the other predictor and control variables (per day).

### 4.4.1 Extraction and Calculation of Individual Data (Corpus)

***Extracting Twitter Users and Tweets.*** In step 1, using a 2014 dataset of tweet IDs, Twitter users were extracted via Tweepy and the Twitter API *api.statuses_lookup*, which returns the author.id of the tweet. These were saved in a dataset. Secondly, Tweepy and the Twitter API *api.user_timeline* were used to extract tweets from 2014-2018 from each of these users. These data were saved in a dataset and automatically flagged as gang-affiliated. Based on the Twitter metadata, the author of the tweet (Twitter user id or screen_name), retweet frequency, number of followers, number of favorites, and tweet source location information were extracted. From the tweet content, the tweet text, emojis, creation date, creation timestamp, hashtags, URLs, and @user mentions were extracted.

***Extracting the Retweets.*** Tweepy and the Twitter API *api.retweets* were used to extract the dates and times of the first 100 retweets of each tweet. The *retweet_id* and the date/time were then saved in a comma-separated value dataset and sorted (first by date, and then by time in descending order) to find the date and time of the first retweet, *FirstReTweetCreated_at*. These data were then used to calculate the retweet time *IDHours* (in hours). Further, any tweets from the corpus with more than 100 retweets were deleted, as this is the maximum number limitation of the API, but more importantly, because prior research has shown that very high numbers of retweets are associated with lower tweet credibility (Mitra, Wright, and Gilbert 2017). The number of retweets, user mentions, and favorites for a tweet were summed to obtain the interactive dissemination frequency, *IDF*. The sum was normalized by converting it to a standardized score (Hoang and Lim 2011; Yang and Counts 2010).

***Calculating Retweet Speed.*** Retweet speed represents the time lag or delayed effects between the original tweet and the first retweet. The timestamp of the original tweet and the timestamp of the first retweet were retrieved to calculate the time lag, which was saved as a feature, Interactive Dissemination Time (in hours) (*IDHours)*, at the individual level. In several cases, missing data was encountered for the first retweet timestamp, which prevented the calculation of retweet speed. This was due to several factors, including constraints with the Twitter retweet API. Some tweets were not retweeted, and thus, had no retweet speed. These were marked with a zero. For these tweets that were retweeted but had missing retweet timestamp data, several options were considered for handling the missing data: delete the observation from the dataset, impute the mean retweet speed for that user's other retweets in the dataset, or mark the retweet speed as zero (valid for volumes less than 5%) (Jakobsen et al. 2017). For users with tweets that had been retweeted, the retweet speed for that user was averaged and missing values were marked accordingly. There was only one user for which the historical retweeted time average could not be calculated, so the retweet speed of those user's tweets (only 2.4% of the entire corpus) were imputed as zero across all observations (as the value could not be blank for a regression). The belief was that an assignment of zero would not be too detrimental to the overall average result across all observations. This is because the retweet time is measured in hours and results in a very small number, on average, for most tweets. Still, this action could potentially influence the credibility of each tweet. However, the risk associated with that action was determined to be better than removing the observation entirely, as it was important to keep as many observations in the dataset as possible.

***Calculating Other Needed Features.*** The tweet text was programmatically parsed to extract and calculate the mentions count, the tweet length, and whether the tweet contained a URL. Various other needed features were captured including average temperature on the day of the tweet, crime rate one year before the prediction day of violent crime, tweet day (converted from variables stored as text (e.g., Sun, Mon, Tue, etc.) to an integer of 1-7, respectively), and tweet hour (e.g., 1-24). For each day in the 2014-2018 timeframe, the researcher manually annotated whether a major event (local or non-local) occurred that could influence violent crime instances. This function was implemented by looking up the date and capturing the major event value (1/0) in a formerly scored auxiliary dataset.

### 4.4.2 Cleaning, Preprocessing, and Tokenization of Data

Common practitioner-based best practices for all preprocessing steps were used in this research, similar to other recent studies (Chang et al. 2018). Each tweet was treated separately and organized into sentences. Each sentence was then prepared using the methods below before training the word vectors.

***Preprocessing the Unstructured Twitter Text.*** The following steps were completed to preprocess all of the unstructured Twitter tweet text and emojis.

- **Removing case sensitivity.** The tweet noise was lowered by changing each letter to lowercase.
- **Replacing and removing Twitter user names (handles).** Personally identifiable information was removed, including every user mention and URL, due to privacy concerns.
- **Removing punctuation, numbers, and special characters**. Non-relevant punctuation, numbers, and special characters that help little in differentiating tweet content were removed.
- **Removing short/stop words**. Smaller words that add little value (e.g., all, the, her, his, is, are, etc.) were removed.
- **Stemming**. Stemming data is a rule-based process of feature reduction where the suffixes (e.g., “ing”, “ly”, “es”, “s”, etc.) are stripped from a word. This reduces word variations to their root word (e.g., *angry*, *anger*, *angered* each becomes *anger*) and allows the reduction of the unique words in the data without sacrificing much meaning. Though stemmed occurred in this research, the researcher appreciated the added confidence in knowing that Word2Vec places words with similar contexts close together in the vector space.
- **Tokenization**. Each tweet was reformatted and tokenized (split) into individual words (tokens). The process of tokenization allows the parsing or separating of a sequence of string text (i.e., characters and numbers) into pieces called tokens. Examples of tokens include keywords, phrases, and symbols. These become inputs

into the text mining process. Some researchers remove emoji modifiers to reduce sparsity (Blevins et al. 2016), but this research tokenized the raw data using *sent_tokenize*. Each emoji was considered an individual token unless the emoji was part of a chain.

- **Visualizing**. Visuals (e.g., word clouds) were created to explore the cleaned tweet text, gain insights, and determine the most frequently used words.
- **Trimming**. The tweets were trimmed to offer a consistent model tweet length.

***Preprocessing the Remainder of the Data (Corpus).*** The preprocessing technique, *MinMax* scaling, from the *MinMaxScaler* library of the Python *sklearn.preprocessing* library was used to normalize certain parts of the dataset. Because the data are spread across a wide range of values, which could result in various features affecting the result more than other features, normalization (rescaling to 0-1) was performed to reduce this effect. This mitigates target variable leakage, which occurs when information from outside the training dataset is used to create the model. Target variable leakage can cause the model to learn or know something that it otherwise would not know and invalidate the performance as well as causing the researcher to become overly optimistic about the model's performance. One example of this is normalizing the entire dataset, as this rescaling process knows the full distribution of data in the training dataset when computing the scaling factors. To mitigate the effects of target variable leakage in a study, a researcher can normalize within each cross-validation fold and use parameters to prepare the data on the held-out test fold on each iteration.

### 4.4.3 Creating the Word Vectors and the Domain-Specific Word Embeddings

In Stage 1 of this study, the Twitter tweet content (text and emojis) of street gang-affiliated individuals was used to train classification algorithms to predict two emotion-based binary feature labels. These include aggression (*Aggress* = 1/0), represented by taunts or threats and various threatening emojis (Table 6) and out-of-control loss (*Loss* = 1/0), represented by text and a sad face (or other similar) emojis (Table 7). The unlabeled Twitter corpus of tweets was exploited to build a domain-specific resource (word embedding) from the unlabeled corpus. Here, the importance of deriving the embedding directly from the community of the study is emphasized, as the localized community-based language is more specific than standard American English or African American English and can rapidly change.

***Build the corpus vocabulary and generator.*** To translate the tweet language local to a specific city, heterogeneous sets of features were derived from the tweet text and emojis (Balasuriya et al. 2016). Using the word embedding philosophy to define a word by the company it keeps, features (unique words) were extracted from the vocabulary of the preprocessed text and a unique numeric identifier was mapped to each feature. The algorithm used pairs, (target center word and proximate context words), a target word length of one, and surrounding context length of *2 x window_size*, where the *window_size* words include those before and after the target word.

***Build the Model Architecture and Train the Model.*** All the data collected from the Twitter dataset was fed into both a sparse word embedding (via the CountVectorization package from the Python-based Scikit library) and a dense word embedding (via the Word2Vec tool). Various models were constructed, and the results were assessed. A prediction-based Skip-gram model was constructed and trained with a negative sampling word rate of 10 sample words to perform feature learning. Skip-gram models work well with medium-sized databases (Mikolov et al. 2013) and perform better than CBOW on small datasets (Mikolov 2013). Skip-gram's objective is to predict the context of a given target-word. In general, the inputs to the algorithm are words of dimensions (*2 x window_size*) trained the model for 20 epochs (number of times the model iterates through the data). The number of dimensions of the embedding (the length of the dense vector to represent each token) were selected. The input was a token index sequence, which was mapped to a vector of various sizes (e.g., 200, 300, etc.) with a trainable embedding matrix.

**Table 6: Features/Measures for Aggression Classifier**

| **Measure Type** | **Measure Level** | **Measure Name** | **Measure Description** |
|---|---|---|---|
| Several Independent Variables (IV) | Individual | Examples could include features for Kill, Gun, Hit, Opps, Smoke, etc. | Tweet content: Text (*text*), Emojis (*emojis*[]) from tweets from a gang-affiliated user. Features retrieved from the Word Embedding process. |
| Dependent Variable (DV) | Individual | Aggression (*Aggress*) | Whether or not the content of the tweet contained aggression (binary class) (1 if the tweet content contained aggression; 0 if it did not). |

**Table 7: Features/Measures for Loss Classifier**

| **Measure Type** | **Measure Level** | **Measure Name** | **Measure Description** |
|---|---|---|---|
| Several Independent Variables (IV) | Individual | Examples could include features for Free, Sad, #RIPDaGuys, #FreeDayGuys, Sad face, etc. | Tweet content: Text (*text*), Emojis (*emojis*[]) from tweets from a gang-affiliated user. Features retrieved from a Word Embedding process. |
| Dependent Variable (DV) | Individual | Loss (*Loss*) | Whether or not the content of the tweet contained loss (binary class) (1 if the tweet content contained out-of-control loss; 0 if it did not). |

Since the average tweet has a sentence length of eleven or fewer words (Hu, Talamadupula, and Kambhampati 2013), the model was initially trained with word embeddings by setting the minimum word count to five (ignores words with a total count less than 5). Initially, the context word window size was set to five, which allows the model to consider five words before (left of) the target word and five words after (right of) the target word. Experimentation with the parameters continued, including a window size of 5 for the initial CBOW model and 10 for the initial Skip-gram model, and then adjusting down to a size of two and other numbers.

These inputs were fed to an embedding layer, which was initialized with random weights. The embedding layer size was set to *vocab_size x embed_size,* which gave compact word vectors (*1 x embed_size)* for each word. While some practices may suggest obtaining at least 250,000 unique words from the word embedding, this study's domain-specific word embedding of 143,700 documents (tweets) provided a lower number of unique words; thus, seed words were added from the lexicon to the source dataset to enhance the embedding.

After the model was trained, the embedding was accessible via the "wv" (word vector) attribute. Word2Vec's “most similar” function was used to investigate how well the word embedding models were trained. The proximity of the vectors was viewed by creating visuals.

### 4.4.4 Partitioning and Cross-Validation of the Data

There are different ways to partition (divide) the data for training and testing of the aggression and loss classifiers. For example, Chang et al. (2018) used shuffled data with a stratified sampling approach to ensure the same allocation across classes for training

(65%), validation (15%), and test sets (20%) for each cross-validation fold. For the algorithms used in this study (one classifier for aggression; one classifier for loss), the data was partitioned into both training (70%) and test (30%) sets, training (80%) and test (20%) sets, and then shuffled. This was important so that the model would not be trained using a sequential set of records and thus establish patterns using those values. This also allowed the mitigation of potential issues arising from having different distributions of labels.

Five-fold cross-validation was used to partition and test each model's ability to predict new data not used in estimating it, such that overfitting and selection bias did not occur (Cawley et al. 2010). Though several methods of cross-validation exist, five-fold cross-validation was used to split the sample into five (k) equally sized subsamples (N/k). The advantage of K-fold cross-validation is that it mattered less how the data was divided. In this approach, selection bias is not present because each data point is part of both the training set and the validation set. This method of estimating expected prediction error via error rates allowed the selection of the best fit model (one with a low error rate) while also helping to ensure the model was not overfit, a situation that occurs when data used to build the model has similar performance as performance over unseen data. This technique was insightful as it determined how the model generalized to an unknown dataset.

The number of tweets containing aggression was relatively small in the training set (reflecting the low distribution in real life). The ratio for the loss tweets in the training data was smaller, as this ratio reflects the even lower distribution of loss tweets in real life when compared to aggressive tweets. For the credibility algorithms, the same tweets containing negative sentiment (either aggression or loss) were sourced and assessed again for credibility. The remaining tweets were used as part of the larger unlabeled corpus of

Twitter data. For each algorithm, the N equaled 1,226; thus, each fold (split) held approximately 245 records. For each fold, there were five reiterations to handle the variance between runs (iterations), one of the folds for validation and the other four for training. For each row in the training set, the data was labeled as *Aggress* = “1” (if the text and emoji content expressed aggression), else it was labeled *Aggress* = “0”. Similarly, *Loss* was set to “1” (if the text and emoji content expressed out-of-control loss that could morph into aggression), else *Loss* was set to “0”.

***Validating Loss and Aggression Labels in the Training Set.*** Each tweet from the training set was validated via two raters (Chicago-based domain experts). Inter-rater reliability was tracked between the raters and the Cohen’s kappa metric was measured for sufficient to high values. This measure represents the observed agreement ((the probability of agreement based on chance) / (1-probability of agreement based on chance)). Dissimilar annotations between the two experts were designated for further review, reconciliation, and resolution by the same two Chicago-based raters.

***Feature Selection for the Aggression and Loss Classifiers.*** Several strategies are employed by researchers to assist them with feature selection in a word embedding process. For example, one can sort the word vectors in descending order and select the top valuable input words and their distances. Another extremely useful strategy to obtain features is to experiment with averaging the word embeddings for each word. This step returns a feature array with each feature labeled with a number (e.g., 0-9 for ten features). Researchers can then use Principle Component Analysis (PCA), which reduces the feature dimensions to two dimensions, and then visualize the clusters (by color-coding each one). In this study, content-based features were automatically derived from the tweet vectors, and averaging was used to cluster the results. PCA with visualization was used to view the results. The results were analyzed over both the CBOW and Skip-gram architectures.

### 4.4.5 Classifying Aggression and Loss in the Test Data Set of Gang-Affiliated Tweets

After validation of the Loss and Aggression features in the training set, the features derived from the word embedding process were used as inputs to the classifiers on the test set of the data. For these two classification tasks, algorithms ran to classify aggression and others to classify loss.

***Run the Classifiers***. For the aggression and loss classifications of Twitter content, several traditional classifiers were used, including a Logistic Regression classifier, an SVM classifier (SVC), and an SVM Linear classifier, for the baseline models. SVM is a discriminative classification technique, which uses a separating line (hyperplane, in multi-dimensional space). When provided labeled training data, it produces an optimized hyperplane, which classifies new instances by mapping a sequence of tokens to a class probability. For each tweet, the algorithm classified it as aggressive (*Aggress* = “1”) if the classifier produced the probability score above a set threshold; otherwise, it scored *Aggress* = “0”. A second classifier classified it as containing loss (*Loss* = “1”) if it produced a score above a set threshold; otherwise, it scored *Loss* = “0”.

The word embeddings also ran on a Multi-layer Perceptron (MLP) model, a neural network used for either classification or regression. Given the number of dimensions for input (*m*) and the number of dimensions for output (*o*), it trains on a dataset to learn a function Rm→Ro. An MLP is advantageous because it can ascertain a non-linear function approximation, and, unlike logistic regression, there can be multiple non-linear (hidden) layers, represented by the tuple’s length, between the input and the output layer. The tuple’s elements represent the number of nodes in the tuple's index. The neural network includes various customizable hyperparameters, including *hidden_layer_sizes* (number of layers and nodes) and the number of epochs or iterations (*max_iter*). The activation function for the hidden layers is characterized by the *activation* hyperparameter, while *solver* is the algorithm used for weight optimization across the nodes and *random_state* sets a seed for reproducing the same results. The Scikit-Learn library was used to run the MLP neural network models.

The second type of neural network used on these data was a standard word-level CNN classifier (Collobert et al. 2011; Kim 2014). The Scikit-Learn library does not offer graphics processing unit (GPU) support, so in this study, the Sequential model in Keras (an open-source library which uses Python), was used to build the CNN model. The Sequential model includes a customizable linear stack of layers for a neural network. In this approach, word vectors are circulated to a lambda layer. Experimentation with averaging the word embeddings occurred, especially with the CBOW models, as these do not consider context word order when averaged to get an average dense embedding (*1 x embed_size*). This layer is then added on top of the prior layer, which extracts features and encodes the semantic meaning of words and sentences.

Other steps included applying a 1D Convolutional layer with kernel sizes one and two (2 x 2 filter matrix), an appropriate filter size to the embedded token sequence, and the rectified linear function (ReLu). The width of the filters is usually consistent with the width of the input matrix. This is because the filters move over full rows of the matrix. The pooling step of the architecture (also known as down-sampling) cuts down the feature map dimensions by compacting the result of a small region of neurons into a single output. This action simplifies the layers. Max pooling was employed, as it often performs better than average pooling (Zhang et al. 2015). Further, in the dropout activation step, a dropout = 0.5 may help the model generalize and not overfit. Here, the dropout represents the ignoring (dropping) of randomly selected neurons during training. The algorithm will then combine and connect the max pooling output to the final single output unit and use a sigmoid activation function (Chang et al. 2018). Sigmoid is an S-shaped curve, similar to logistic regression, for binary classification. In this prediction phase, the averaged context embedding is then passed to a dense layer that classifies the sentence, thus predicting the target word. (A softmax activation function (multinomial logistic regression or normalized exponential function) requires the output to total to one to allow model output interpretations as probabilities and model predictions based on the option with the highest probability.) Each output vector's element portrays the likelihood of a specific word occurring in the context. This is matched with the actual target word, computing the loss by leveraging the categorical_crossentropy loss (where lower scores are equivalent to a more efficacious model). Backpropagation of the errors is performed to alter the embedding layer weights with each epoch. This process repeats for all (context, target) pairs for a customized set of iterations (epochs).

***Validate the Classifier Results***. After the classifiers ran, the findings were validated under five-fold cross-validation, and accuracy, precision, recall, Area under the Curve (AUC), and F1 scores were recorded. Between the baseline and neural network models, the one that achieved the best results was selected and run over the unlabeled data, including unlabeled data with hashtags. For example, *#FreeDaDommmmm* is a hashtag used in a tweet that refers to an imprisoned individual known by the Twitter user who posted the tweet, but because neural networks and SVM models function at the word level, these methods can conclude this tag as a rare or unfamiliar token (Chang et al. 2018).

### 4.4.6 Detecting and Validating Credible Content Signaling

In this study, current research was extended by testing whether tweets with loss or aggression were credible signals of forthcoming criminal action or just merely boasting or grandstanding by the gang-affiliated individual. The tweets labeled as aggressive or loss were retrieved and used as inputs to other algorithms to determine credibility.

***Creating a Scale for Credible Content Signaling.*** Via the Co-UGS conceptual model, this study posits that credible content signaling exists when tweets with negative expressions of aggression or loss sentiment, together with other dimensions of credibility, significantly explain intent to commit a crime at the individual level. More specifically, it suggests that these negatively-phrased tweets with retweet counts between 5 and 100, retweet speed within one hour, higher followers, hashtag usage, higher favorites counts, longer tweet lengths, and URL usage make the tweet more credible. It also included whether the tweet contained a booster word (seventy words shown to increase ("boost") credibility in earlier research) (Table 8). Further, it added whether the user was active or not as another dimension of credibility. Earlier research considered an active Twitter user as one who authored a tweet more than once, or one who sourced a user mention or retweet in the corpus. This research expands the metric value of (Chang et al. 2018) by using a value of 225 or more tweets in the data sample to denote an active user, though most users have hundreds of tweets.

Table 8: Measures for Credibility Algorithms

| Measure Type | Measure Level | Measure Name | Measure Description |
|---|---|---|---|
| Independent Variable (IV) | Individual | Retweet Hours (*IDHOURS*) | The time difference (in minutes) between the tweet and its first retweet. Calculated manually using *TimeStamp* from the tweet metadata and the first retweet (*FirstReTweetCreated_at*). The time of the first retweet is retrieved by a Twitter API that returns the time of all retweets associated with the tweet. The results are then sorted in descending order by time to get *FirstReTweetCreated_at*. |
| Independent Variable (IV) | Individual | Retweet Frequency (*retweet_freq*) + Favorites Count (*favorites_count*) = *(RETWEETFFREQ)* | The number of times (count) the tweet was retweeted (limited to those tweets with a retweet frequency max of 100). Extracted from the Twitter tweet metadata. Favorites count is the number of times (count) of the tweet favorites. Retrieved from the Twitter user metadata. |
| Independent Variable (IV) | Individual | User Mentions Frequency (*MENTIONSFREQ)* | The number of times (count) a Twitter user was mentioned in the text of an individual tweet. Calculated from the tweet text. |
| Independent Variable (IV) | Individual | Contains Hashtag (*CONTAINSHASHTAG*) | Whether the tweet contains a hashtag (1 (Yes) or 0 (No)). Calculated manually. |
| Independent Variable (IV) | Individual | Contains URL (*CONTAINSURL*) | Whether the tweet contains a URL (1 (Yes) or 0 (No)). Calculated manually. |
| Independent Variable (IV) | Individual | Tweet Length (*TWEETLENGTH*) | The length of the tweet text, measured in characters. Calculated manually. |
| Independent Variable (IV) | Individual | Number of followers of Twitter user (*FOLLOWERSFREQ*) | The number of followers of the Twitter user, retrieved from the Twitter metadata *followers_count*. |
| Independent Variable (IV) | Individual | Active Twitter user (*ACTIVEUSER*) | Whether the user associated with the tweet is active (1 (Yes) or 0 (No)). Calculated manually. An active Twitter user is one with 225 or more tweets in the dataset. |

**Table 8: Measures for Credibility Algorithms (continued)**

| | | | |
|---|---|---|---|
| Independent Variable (IV) | Individual | Contains a Booster Word *(CONTAINSBOOSTER)* | Whether the tweet contains a booster word. These include the following words (or stems of words):<br>"aggravat, agree, amazement, anxiously, appeared, awed, bright, brilliant, calamity, catastrophic, charismatic, clear, close, commits, contingen, darn, defeat, depending, describe, devastating, distinctive, distressed, dynamic, eager, ecstatic, established, exceptionally, exclu, express, fail, fantastic, grieve, guarantee, halfass, heartbroke, immaculate, inexplicable, intricate, loser, miraculously, miser, mishap, misses, mortified, piti, precise, promising, radiant, realize, reliability, sobbed, startl, stink, strangely, sucky, tell, tends, terrific, trouble, unanimous, undeniable, unforeseen, unique, validity, vibrant, victim, weep, while, and wonderful." |
| Dependent Variable (DV) | Individual | Credibility Aggression Score (*CREDAGGRESS*)<br>or<br>Credibility Loss Score (*CREDLOSS*)<br>or<br>Credible Aggression (*CREDAGGRESSFREQ*)<br>or<br>Credible Loss (*CREDLOSSFREQ*) | The credibility score (0-10) associated with the tweet containing aggression.<br>The credibility score (0-10) associated with the tweet containing loss.<br>The credibility (1/0) class associated with the tweet containing aggression.<br>The credibility (1/0) class associated with the tweet containing loss. |

In this part of the study, the goal was to label and calculate the probability each tweet has of exhibiting credible content, with the ultimate goal of aggregating this credible content, along with other research model measures (Table 9), and studying its association with violent crime. In the tweets where access to arrest and conviction data by a gang affiliate exists (Appendix B), the Twitter name was collected, converted to the individual's legal name, and checked to see if the individual committed a crime. For example, @TyquanAssassin is the Twitter name of Gakirah Barnes, who was a known gang leader that killed nine individuals before her death in 2014 (Blevins et al. 2016). Thus, her tweets were deemed very credible and assigned a credibility score of 10 (the maximum).

Secondly, for the instances where legal names and criminality were not available, applicable research and a structured process was used to design and create a new scale for a basis credibility score (0-10) specific to the unique environment of the study of gang-affiliated social media content. The (Boateng et al. 2018) approach was employed to create and validate the scale, which states that there are essentially three broad phases to creating a rigorous scale. These include item development, scale development, and scale evaluation. These phases are further broken down into additional steps, including identification of the domain, content and external validity, item reduction analysis, and extraction of factors.

***Identification of Domain and Content Validity.*** A domain in this context is a construct that references a conceptual variable or unobserved behavior that is one of the targets of the research. The domain was identified as social media content credibility related to aggression or loss from gang-affiliated communities in urban environments. Content validity refers to the concept that the measure should correctly assess the study's domain, or, restated, the variables should collectively measure what they are alleged to measure.

**Table 9: Research Model Measures**

| Measure Type | Measure Level | Measure Name | Measure Description |
|---|---|---|---|
| Independent Variable | Per period | Total Aggression Credibility Score (*TOTALCREDAGGRESS*) | The sum of all aggression credibility scores for the day. |
| Independent Variable | Per period | Total Loss Credibility Score *(TOTALCREDLOSS)* | The sum of all loss credibility scores for the day. |
| Independent Variable | Per period | Gang-Affiliated Tweet Frequency (*GATF)* | The count of all gang-affiliated tweets that day in the study's sample. |
| Independent Variable | Per Period | Total Credible Aggression and Loss Tweet Frequency *(CALF)* | The count of all gang-affiliated tweets containing credible aggression or loss per day in the study's sample. |
| Independent Variable | Per period | Interactive Dissemination Frequency (*TOTALIDF*) | The sum of all of the tweets retweets, mentions, and favorites (likes) for the day. |
| Independent Variable | Per period | Interactive Dissemination Hours *(AVGIDHOURS*) | The average of all of the credible tweets' retweet time (in hours) for the day. |
| Independent Variables | Per Period | Prior Periods' Violent Crime Counts (*VCLAG1*, *VCLAG2*, *VCLAG3*, *VCLAG4*, *VCLAG5*, *VCLAG6*, *VCLAG7*) | Time series data representing the prior days' violent crime count(s). |
| Independent Variables | Per Period | Quarterly Seasonal Indicators (*Q1*, *Q2*, *Q3)* | Binary data (1=yes; 0=no) representing quarterly seasonal indicators for quarters 1-3. Quarter 4 is represented by the baseline of zeros for Q1, Q2, and Q3. |
| Independent Variable | Auxiliary (control) | Average Temperature (*AVGTEMP*) | The mean daily temperature (F) in the city on the tweet date. |
| Independent Variable | Auxiliary (control) | Day of Week (*DAY*) | The day of the tweet, plus one. Examples include 1 (Sunday), 2 (Monday), 3 (Tuesday), etc. Calculated from the tweet *created_at.* |
| Independent Variable | Auxiliary (control) | Average Daily Tweet Time (*HOUR*) | The average hour for all gang-affiliated tweets in that day's sample. Examples are 1 (1:00 AM), 2 (2:00 AM), and 3 (3:00 AM), through 24 (12 PM). |

**Table 9: Research Model Measures (continued)**

| | | | |
|---|---|---|---|
| Independent Variable | Auxiliary (control) | Major Event (*MAJOREVENT*) | Whether a major event historically occurred on the day after the date of the tweet (1=Yes; 0=No). |
| Independent Variable | Auxiliary (control) | Daily Historical (Last Year's) Crime Rate *(HISCRIMERATE)* | The aggregated daily crime count in the city per 100,000 General Population (364 days before the tweet date, or restated as, one year before the day of the predicted crime count). This is calculated by dividing by the city population that year into the count of daily crimes in the city (e.g., if a tweet date is 5/5/2015, the model predicts the violent crime count for 5/6/2015). The daily crime rate per 100,000 general population on 5/6/2014 is extracted. These crime counts are aggregated from the dataset found via the city's data portal, and the population values are extracted from worldpopulationreview.com. |
| Dependent Variable | Per period | Violent Crime Count (per period) (*VCCOUNT*) | The violent crime count predicted for the day after the tweet. |

Evaluation by expert and target population judges is the common means to assess content validity. Thus, not only was the past work of social media researchers utilized, but an external domain expert was also employed to judge and confirm whether the variables using for perceived credibility were indeed valid.

***Item Reduction Analysis and Extraction of Factors.*** For item reduction analysis, the credibility indicators from the Co-UGS conceptual model were used. These credibility indicators included characteristics of the Twitter user, including whether they were active and how many followers they had. Other indicators involved data from the tweet itself, including the tweet length, retweet count, favorites count, retweet speed, user mentions count, and whether the tweet contained a booster word, hashtag, or URL. Because each tweet in this part of the study already contained negative sentiment, credibility was analyzed by using ten (of the eleven) indicators as input features to the credibility algorithm. These included 1) retweet speed, 2) retweet frequency, 3) favorites count, 4) user mention frequency, 5) whether the tweet contained a booster word, 6) whether the tweet contained a hashtag, 7) tweet length, 8) whether the tweet contained a URL, 9) the number of followers of the Twitter user, and 10) whether the Twitter user was active.

Correlation was analyzed between the factors to determine association, and an Exploratory Factor Analysis was used for relating similar factors together and reducing dimensionality. The factor analysis findings revealed that two of the credibility indicators, retweet frequency and favorites (likes) count, cleanly loaded on the first principal component with a high positive correlation. Bartlett's Test of Sphericity, using a verified Varimax rotation, confirms that the test is chi-square distributed. This test asks whether the variables in the factor correlation matrix, as a whole, differ significantly from zero (also shown in the identity matrix, where there is no correlation). Because the significance p-value is very low (.000), there is confidence that the variables are correlated to each other.

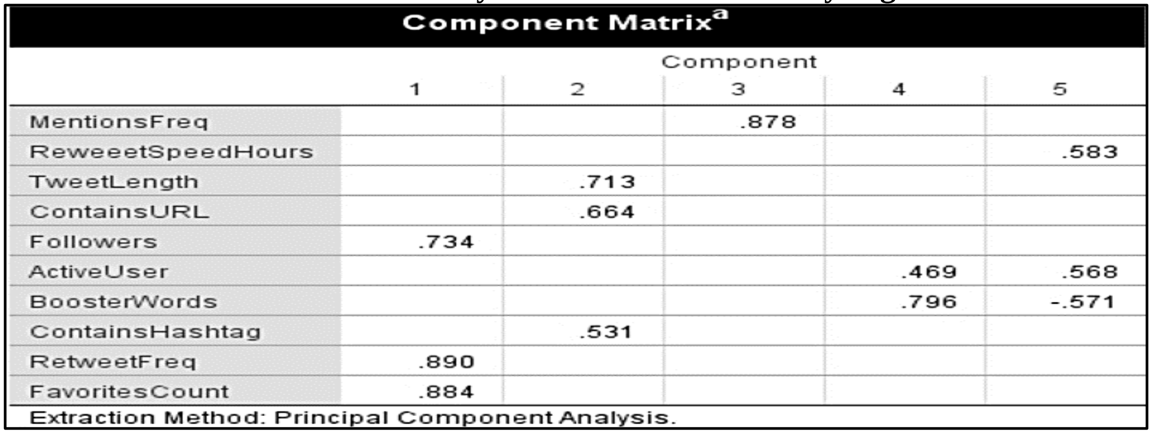
The study’s sample size is greater than 200, making it acceptable for the test. This is further demonstrated by the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy, which suggests that any value more than 0.5 is sufficient, but a value over 0.6 is preferred. The KMO value is 0.631; thus, this test is confirmed. The extraction of nine factors explains 97.64% of the total variance in the factor analysis. The scree plot indicates that five of the nine are above the eigenvalue of one; thus, by default, five factors would automatically extract.

Thus, because of these findings, the ten candidate credibility indicators were collapsed down to nine, as two of them essentially reveal the same type of information. This new construct is Retweet and Favorites Frequency (*RetweetFFreq*). The data from each of these indicators were summed to represent one aggregate indication of the retweet and favorites dissemination of the tweet. Even though the mentions frequency is part of the study's Interactive Dissemination Frequency construct, it loaded on a different component and thus, it remained as a separate credibility component. Further, while the Followers variable also loaded cleanly on Component 1, it was separated into a unique dimension, as it was not highly correlated with the other two variables in Component 1 and because interactive dissemination frequency needed to be separate for the research model algorithm.

The component matrix (Table 10) shows the correlations between the factor and the component. The retweet frequency and favorites count variables are loading cleanly on principal component 1. All other factors were kept separate, as they were not highly correlated to each other and because nine is a reasonable number of variables for the regression. The correlation matrix findings (Table 11) also revealed that retweet frequency and favorites count were highly correlated (r=.757) and could thus introduce multicollinearity into the model, affirming the decision to combine them.

***Annotating the Credibility Values in the Training Set.*** With the scale factors established and validated, the actual values in the training data were ready to be annotated. Using a straight average approach, a method was designed to label the basis credibility score as follows. If the sum of retweet frequency and favorites count equaled six or more, the basis credibility score increased by two, as it involved two separate and equally-weighted variables combined into one. If the retweet speed was within one hour, the basis credibility score increased by one. If the tweet length was more than 11 characters (the average), the basis credibility score increased by one.

**Table 10: Factor Analysis Results for Credibility Algorithm**

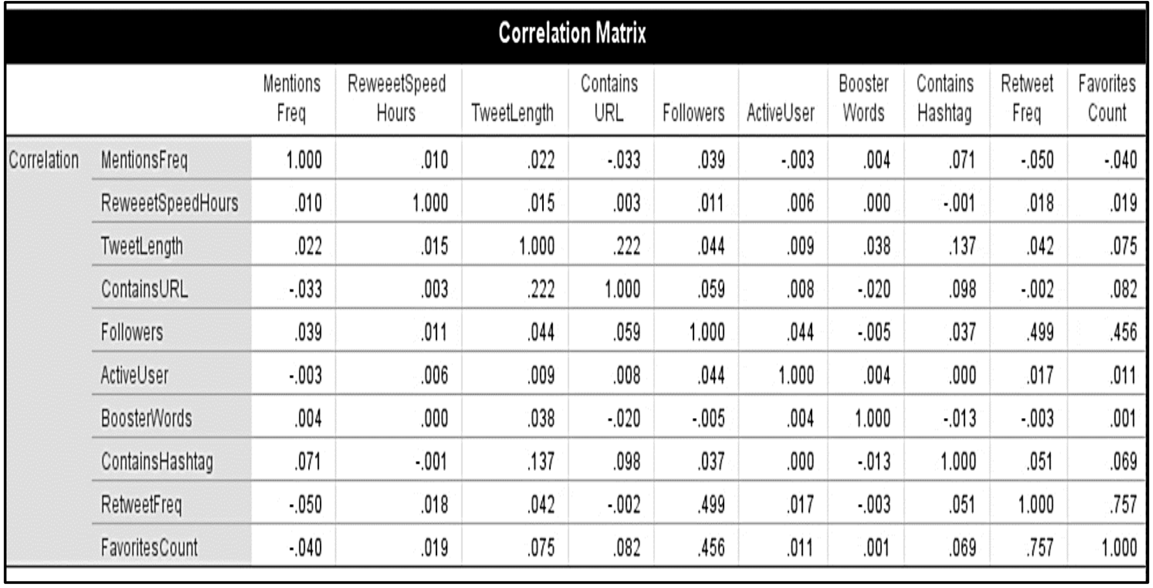


Component Matrix[a]

| | Component | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| MentionsFreq | | | .878 | | |
| ReweeetSpeedHours | | | | | .583 |
| TweetLength | | .713 | | | |
| ContainsURL | | .664 | | | |
| Followers | .734 | | | | |
| ActiveUser | | | | .469 | .568 |
| BoosterWords | | | | .796 | -.571 |
| ContainsHashtag | | .531 | | | |
| RetweetFreq | .890 | | | | |
| FavoritesCount | .884 | | | | |

Extraction Method: Principal Component Analysis.

**Table 11: Factor Analysis Correlation Results for Credibility Algorithm**



Correlation Matrix

| | | Mentions Freq | ReweeetSpeed Hours | TweetLength | Contains URL | Followers | ActiveUser | Booster Words | Contains Hashtag | Retweet Freq | Favorites Count |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Correlation | MentionsFreq | 1.000 | .010 | .022 | -.033 | .039 | -.003 | .004 | .071 | -.050 | -.040 |
| | ReweeetSpeedHours | .010 | 1.000 | .015 | .003 | .011 | .006 | .000 | -.001 | .018 | .019 |
| | TweetLength | .022 | .015 | 1.000 | .222 | .044 | .009 | .038 | .137 | .042 | .075 |
| | ContainsURL | -.033 | .003 | .222 | 1.000 | .059 | .008 | -.020 | .098 | -.002 | .082 |
| | Followers | .039 | .011 | .044 | .059 | 1.000 | .044 | -.005 | .037 | .499 | .456 |
| | ActiveUser | -.003 | .006 | .009 | .008 | .044 | 1.000 | .004 | .000 | .017 | .011 |
| | BoosterWords | .004 | .000 | .038 | -.020 | -.005 | .004 | 1.000 | -.013 | -.003 | .001 |
| | ContainsHashtag | .071 | -.001 | .137 | .098 | .037 | .000 | -.013 | 1.000 | .051 | .069 |
| | RetweetFreq | -.050 | .018 | .042 | -.002 | .499 | .017 | -.003 | .051 | 1.000 | .757 |
| | FavoritesCount | -.040 | .019 | .075 | .082 | .456 | .011 | .001 | .069 | .757 | 1.000 |

For the existence of a hashtag, a booster word, a URL, or at least one user mention, the basis credibility score increased by one for each. If the Twitter user was active or had more than 100 followers, the basis credibility score increased by one for each. Finally, from the tweet content, any use of song lyrics or threatening language targeting rival gang members or disrespecting (a *diss*) of a recently killed gang member were analyzed, as these also make the tweet more credible. If the language in the tweet content disrespected a dead gang member or threatened a rival gang member, the basis credibility score increased by one for each. The domain expert validated this approach.

***Validating the Credibility Feature in the Training Set.*** Domain experts validated the results externally by manually reviewing and confirming a sample of the tweets' basis credibility scores based on their domain knowledge about perceived credibility. Inter-rater reliability was calculated between the two experts related to credibility in 99 tweets. Of those, there was 95.77% inter-rater reliability (Cohen's Kappa (K)) in the ratings, 100.00% sensitivity, and 95.00% specificity. There were two tweets where the raters did not agree. The dissimilar annotations between the experts were designated for further review, reconciliation, and resolution by the same two Chicago-based raters.

The basis credibility score was also used to label the credibility classes in the training data for both aggression and loss, as these were needed to classify the tweets for the study's Direct Effects Model. If the aggression credibility score was 0-4.4 (as a cut-off number 4.5 or more rounds to five mathematically), the credible aggression class (CREDAGGRESSFREQ) was labeled as "0" in the training data. If the aggression credibility score was 4.5-10, the credible aggression class was labeled as "1" in the training data. Similarly, if the loss credibility score was 0-4.4, the credible loss class (CREDLOSSFREQ) was labeled as "0" in the training data, whereas, if the loss credibility score was 4.5-10, the credible loss class was labeled as "1" in the training data.

***Detecting and Validating Perceived Credibility in the Unlabeled Corpus.*** With the tweets marked with credible aggression or loss, the credibility algorithms ran over the training and test datasets using the nine credibility indicators. Thus, the *Aggress* feature or the *Loss* feature together with the credibility features was used to determine the credibility of each gang-related tweet and indicate that the author of the tweet has the potential to commit a violent crime. A Multiple Linear Regression algorithm ran to predict the aggression credibility score, and another ran to predict the loss credibility score of each tweet with negative sentiment (Table 16). A Logistic Regression algorithm executed to predict the credible aggression class and another to predict the credible loss class of each tweet with negative sentiment (to use with the Direct Effects Model) (Table 17). For each run, the results were compared and the performance evaluated using various metrics, including $R^2$, F-Statistic, and statistical significance metrics. The best algorithm for both regression and classification was selected to score the tweet credibility data and provide verified individual results.

### 4.4.7 Aggregating the Credible Signals and Other Variables

In Stage 2 of the study, those individual aggression credibility scores and loss credibility scores were aggregated to form new constructs in the research model. Daily sums for all of the aggression credibility scores were aggregated giving the total daily aggression credibility score for each day. Daily sums for all of the loss credibility scores were aggregated giving the total loss credibility score for each day. To produce the needed values for the Direct Effects Model (Table 12), similar aggregation steps were performed.

**Table 12: Direct Effects Model Measures**

| **Measure Type** | **Measure Level** | **Measure Name** | **Measure Description** |
|---|---|---|---|
| Independent Variable | Per period | Total Aggression Credibility Frequency (*TOTALCREDAGGRESSFREQ*) | The sum of all aggression credibility frequencies for the day. |
| Independent Variable | Per period | Total Loss Credibility Frequency *(TOTALCREDLOSSFREQ)* | The sum of all loss credibility frequencies for the day. |
| Independent Variable | Per period | Gang-Affiliated Tweet Frequency (*GATF)* | The count of all gang-affiliated tweets that day in the study's sample. |
| Independent Variable | Per period | Total Retweet and Favorites Frequency (*TOTALRETWEETFFREQ*) | The sum of all of the retweets and favorites (likes) from credible tweets with aggression or loss for the day. |
| Independent Variable | Per period | Average Interactive Dissemination Hours *(AVGIDHOURS*) | The average of all of the credible tweets' retweet time (in hours) for the day. |
| Independent Variable | Per period | Total Tweet Length (*TOTALTWEETLENGTH*) | The sum of tweet lengths of all credible tweets for the day. |
| Independent Variable | Per period | Total Tweets with URL (*TOTALURLS*) | The frequency sum of all credible tweets with a URL for the day. |
| Independent Variable | Per period | Total Tweets with Booster (*TOTALBOOSTERS*) | The frequency sum of all credible tweets with a booster word for the day. |
| Independent Variable | Per period | Total Tweets with Hashtag (*TOTALHASHTAGS*) | The frequency sum of all credible tweets with a hashtag for the day. |
| Independent Variable | Per period | Total Mentions Frequency (*TOTALMENTIONSFREQ*) | The frequency sum of all credible tweets' mentions for the day. |
| Independent Variable | Per period | Total Followers (*TOTALFOLLOWERS*) | The sum of followers of all Twitter users of credible tweets for the day. |
| Independent Variable | Per period | Total Active Users (*TOTALACTIVEUSERS*) | The sum of active Twitter users of credible tweets for the day. |

**Table 12: Direct Effects Model Measures (continued)**

| | | | |
|---|---|---|---|
| Independent Variables | Per Period | Quarterly Seasonal Indicators (*Q1, Q2, Q3)* | Binary data (1=yes; 0=no) representing quarterly seasonal indicators for quarters 1-3, relative to quarter 4. Quarter 4 is represented by the baseline of zeros for Q1, Q2, and Q3. |
| Independent Variables | Per Period | Prior Periods' Violent Crime Counts (*VCLAG1*, *VCLAG2*, *VCLAG3*, *VCLAG4*, *VCLAG5*, *VCLAG6*, *VCLAG7*) | Time series data representing the prior days' violent crime count(s). |
| Independent Variable | Auxiliary (control) | Average Temperature (*AVGTEMP*) | The mean daily temperature (F) in the city on the tweet date. |
| Independent Variable | Auxiliary (control) | Daily Historical (Last Year's) Crime Rate *(HISCRIMERATE)* | The aggregated daily crime count in the city per 100,000 General Population (364 days before the tweet date, or restated as, one year before the day of the predicted crime count). (See Table 9 for calculation information.) |
| Independent Variable | Auxiliary (control) | Day of Week (*DAY*) | The day of the tweet, plus one. Examples include 1 (Sunday), 2 (Monday), 3 (Tuesday), etc. Calculated from the tweet *created_at*. |
| Independent Variable | Auxiliary (control) | Average Daily Tweet Time (*HOUR*) | The average hour for all gang-affiliated tweets in that day's sample. Examples are 1 (1:00 AM), 2 (2:00 AM), and 3 (3:00 AM), through 24 (12 PM). |
| Independent Variable | Auxiliary (control) | Major Event (*MAJOREVENT*) | Whether a major event historically occurred on the day after the date of the tweet (1=Yes; 0=No). |
| Dependent Variable | Per period | Violent Crime Count (per period) (*VCCOUNT*) | The violent crime count predicted for the day after the tweet. |

All gang-affiliated tweets per day were totaled giving Gang-Affiliated Tweet Frequency and all credible gang-affiliated tweets containing aggression or loss per day giving Credible Aggression and Loss Tweet Frequency per day. From tweets containing credible loss or aggression, the Retweet Frequency, Mentions Frequency, and Favorites Frequency per tweet (IDF) were summed. All tweet-level sums were then totaled per day to determine Total Interactive Dissemination Frequency. From tweets containing credible loss or aggression, the Retweet Time (in hours) per day was averaged to determine Average Interactive Dissemination Hours (AVGIDHOURS). The auxiliary (control) variables for all tweets with credible aggression or loss was also aggregated. These include the day of the week (1-7) of the tweet (plus one to obtain the prediction day) and the average daily tweet time in hours (HOUR). Using several auxiliary datasets, the day was looked up to determine if a major event occurred on the prediction day, the average city temperature in Fahrenheit on the tweet day, and the historical (one year earlier from the prediction day) daily crime rate per 100,000 on the day. All measures used in the research model are listed in Table 9.

### 4.4.8 Creating a Control Variable for Prior Period Violent Crime Count(s)

Historical (time series) crime data were employed as control variables to potentially improve the violent crime prediction. To accomplish this, the violent crime count variable was regressed on seven of its lags (one or more past period instances) in an autoregressive AR(p) model (where p is the number of lags) as one or more of the control variables (Appendix D). One possible occurrence in this study’s model is that violent crime count (the dependent variable) may also affect the volume and dissemination of gang-related credible tweets, indicating a relationship in a bi-directional fashion. Though not assessed theoretically in the Literature Review, nor added as part of the research model, this reverse impact could have occurred in the crime data. Thus, a model may be more efficacious if it captured the lagged values of the dependent variable as well as the current and lagged values of other independent (exogenous) variables. The use of a vector autoregressive (VAR) model could assist in this effort. VAR models assume the exogenous variables also depend on lagged values of the endogenous variable and treat each variable in the model as if it influences the other variables equally; that is, they treat each variable as endogenous. VAR models describe the evolution of the model's variables in reaction to a shock (innovation or impulse) in at least one of the variables. The impulse response function allows researchers to trace a shock within time in a potentially noisy system of equations and analyze the possible backward effects in a model.

However, there are several vital considerations when using VAR models. Firstly, the researcher must test to see if the dependent variable (e.g., violent crime count (per period)) is autoregressive. Secondly, the researcher should also perform Granger Causality and Cointegration tests for the independent variables to verify which have a bi-directional relationship to violent crime count. Thirdly, both the endogenous variable (DV) and the exogenous variables should also be stationary in a VAR model, so assessing for stationarity in the independent variables is needed. Due to this premise of stationarity, the researcher must consider how to design the model to account for any existing seasonality in the data, along with possible trends (Appendix C), and employ differencing to make the data stationary. Fourthly, VAR models are atheoretical; that is, they do not impose a theoretical structure on the equations. They also assume each variable influences every other variable in the system; thus, the direct and reasonable interpretation of the estimated coefficients is problematic. Therefore, if the interpretation of the estimated coefficients is needed, as in this research, the researcher should perform and evaluate an additional baseline model using the model variables with time-series data. For example, one could account for quarterly seasonality in the data by using use (k-1) dummy variables to represent the quarters (where k = 4) and use a control group to test it by running the model first without the control variables, then again with the variables to assess the results and see whether the seasonal dummy variables added predictive power. It is for all of these reasons that a future dedicated empirical test is recommended for the relevance of a VAR model in these crime data as future research (see Chapter 6). Instead, for this study, an autoregressive test was performed on the dependent variable (violent crime count (per period)) and included indicator variables to represent the quarterly seasonality in the violent crime data.

To represent the sixth control, past time series data were captured for violent crime counts from the City of Chicago crime portal for the seven prior periods (days) (*VCLAG1, VCLAG2, VCLAG3, VCLAG4, VCLAG5, VCLAG6,* and *VCLAG7*). Customized k-1 quarterly seasonal variables (*Q1*, *Q2*, and *Q3*) were also created using binary variables. For example, for a tweet with a tweet date of 6/24/2016, the quarterly seasonal indicator variables are Q1= “0”, Q2= “1”, and Q3= “0”, as June falls into quarter 2. These features and their respective values for each tweet were written to the dataset.

### 4.4.9 Predicting the Violent Crime Count for the Next Day

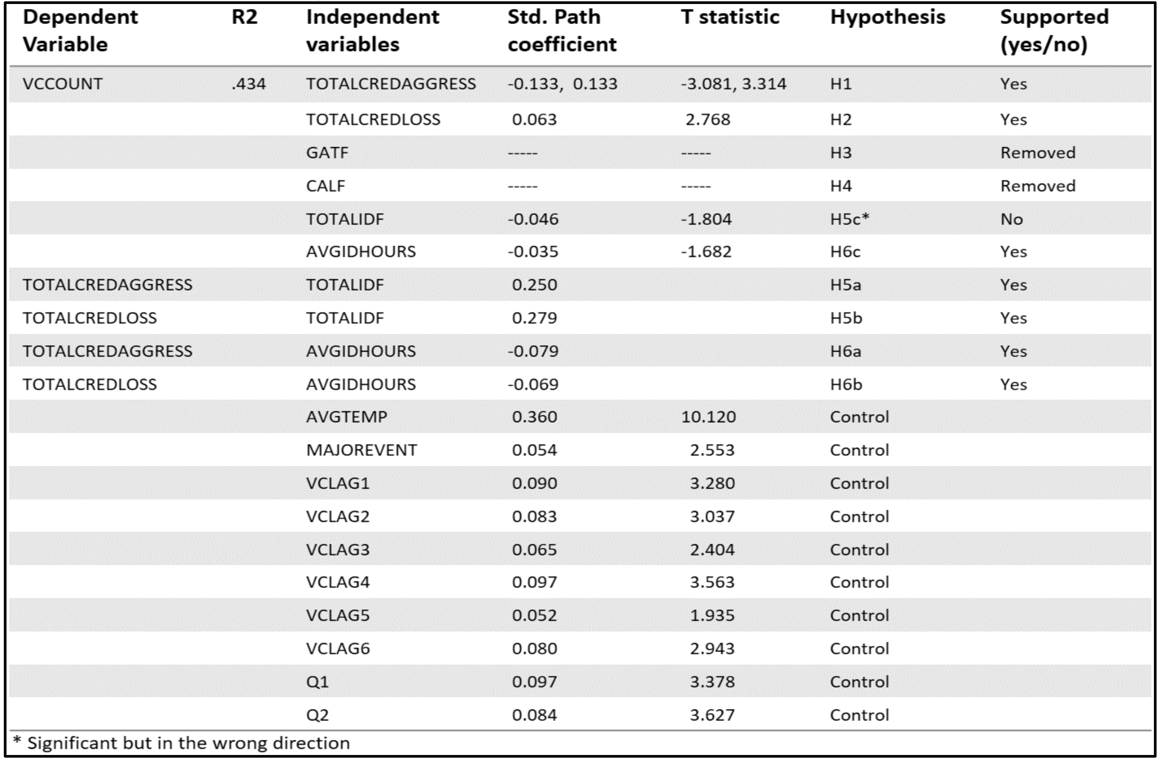
In this part of the study, the crime model of (Aghababaei 2017), who used historical crime rates, day of the week, the frequency of daily tweets, weather, unemployment rate, and emerging events with content-based features to predict crime was improved. The improved model predicted actual violent crime counts instead of crime trend direction changes. For the predictors, the researcher included the first three determinants of the (Aghababaei 2017) model, substituted more granular (daily) temperature data, and added interactive dissemination frequency, interactive dissemination (retweet) speed, tweet hour, and forecasting variables (e.g., prior violent crime period counts and quarterly seasonal indices). Further, the researcher analyzed whether there was a predictive *external* major event at tweet time, like a social, political, racial, or cultural trigger that may cause friction, riots, or concern within a gang community.

***Annotating the Dependent Variable in the Training Set.*** The data in the training set was labeled for this prediction algorithm based on actual historical violent crime counts in Chicago. These data were retrieved from the City of Chicago portal.

***Detecting Violent Crime with the Research Model Algorithm.*** With the dependent variable labeled, the aggregated daily measures and auxiliary determinants as controls were combined to predict the violent crime count (*VCCOUNT)* for the next day at the city level in the dataset. A hierarchical multiple regression was then performed to estimate the coefficients of the independent variables and the control variables to predict the violent crime count for the next day (Table 13). The sixth control variable (represented, in part, as a series of prior periods' violent crime counts) tested an autoregressive prior period (lag) to see if including these lags of prior violent crime count added to the predictive power of the model. Three dummy binary (yes = "1"; no = "0") indicator variables were also added, *Q1, Q2,* and *Q3*, to account for the quarterly seasonality in the data.

Error metrics and significance between various models were assessed, and the best one was selected to provide daily summative violent crime count insights (per period) in addition to the individual insights from the credibility algorithm (after the study's first stage).

**Table 13: Regression Results for Research Model (Standardized)**



| Dependent Variable | R2 | Independent variables | Std. Path coefficient | T statistic | Hypothesis | Supported (yes/no) |
|---|---|---|---|---|---|---|
| VCCOUNT | .434 | TOTALCREDAGGRESS | -0.133, 0.133 | -3.081, 3.314 | H1 | Yes |
| | | TOTALCREDLOSS | 0.063 | 2.768 | H2 | Yes |
| | | GATF | ----- | ----- | H3 | Removed |
| | | CALF | ----- | ----- | H4 | Removed |
| | | TOTALIDF | -0.046 | -1.804 | H5c* | No |
| | | AVGIDHOURS | -0.035 | -1.682 | H6c | Yes |
| TOTALCREDAGGRESS | | TOTALIDF | 0.250 | | H5a | Yes |
| TOTALCREDLOSS | | TOTALIDF | 0.279 | | H5b | Yes |
| TOTALCREDAGGRESS | | AVGIDHOURS | -0.079 | | H6a | Yes |
| TOTALCREDLOSS | | AVGIDHOURS | -0.069 | | H6b | Yes |
| | | AVGTEMP | 0.360 | 10.120 | Control | |
| | | MAJOREVENT | 0.054 | 2.553 | Control | |
| | | VCLAG1 | 0.090 | 3.280 | Control | |
| | | VCLAG2 | 0.083 | 3.037 | Control | |
| | | VCLAG3 | 0.065 | 2.404 | Control | |
| | | VCLAG4 | 0.097 | 3.563 | Control | |
| | | VCLAG5 | 0.052 | 1.935 | Control | |
| | | VCLAG6 | 0.080 | 2.943 | Control | |
| | | Q1 | 0.097 | 3.378 | Control | |
| | | Q2 | 0.084 | 3.627 | Control | |

* Significant but in the wrong direction

### 4.5 Evaluation

Stage 1 of this study included a comprehensive analysis, which systemically evaluated the efficacy of each proposed algorithm, checking for robustness, generalizability, and applicability. Because social media data can contain rumors, robustness was checked via the study's domain knowledge experts. The predictive power of the models was tested on a holdout sample to assess for generalizability. To check for applicability, the practical contributions and implications of each algorithm was validated with domain knowledge experts. The algorithms were also evaluated for credibility from the tweets expressing aggression or loss, and the research model predictors were assessed for next day violent crime count similarly.

### 4.5.1 Stage 1 Evaluation

The performance of the classifiers was trained and evaluated under a five-fold cross-validation scheme. For the five-fold cross-validation experiment, several evaluation metrics for the ‘aggression’ and ‘non-aggression’ classes were reported using the count of true positives (*tp*), false positives (*fp*), true negatives (*tn*), and false negatives (*fn*). These metrics and their respective calculations are:

*Precision = tp / (tp + fp)*

*Recall = tp / (tp + fn)*

*F1score = 2 ∗ (Precision ∗ Recall) / (Precision + Recall)*

The area under the receiver operating characteristic (ROC) curve (AUC) and accuracy were also reported. Accuracy is the fraction of correct predictions; that is, the correct predictions as the numerator and the total number of data points as the denominator. The two possible averages comprise the macro average (the unweighted mean per label) and the weighted average (the support-weighted mean per label). The unweighted averages were reported.

A similar method was used for the loss training set. For both aggression and loss, various baseline classifiers were ran and the results were compared to the neural network models. By analyzing the error metrics, confidence increased that a model can perform adequately and give accurate predictions on a new set of records. This evaluation demonstrated the use of pre-trained word embeddings to test and see if the precision of supervised learning models improved for aggression sentiment and loss sentiment. For the credibility prediction models, the F1 score, significance, and R$^2$ metrics were used to assess the fit of the models.

### 4.5.2 Stage 2 Evaluation

In Stage 2 of the study, several predictive algorithms were tested using the aggregated features from Stage 1, combined with tweet frequency of loss and aggression, tweet frequency of gang affiliates, time-series data, and other control predictors to predict the violent crime count for the next day. To evaluate the best research model, the F-Statistic and its significance, Pearson’s r for correlation, VIF and Pearson’s r for multicollinearity, variable p-values for significance, and R$^2$ metrics were used to assess the fit between the research model and the alternate Direct Effects Model. The model with the best results was then selected, and the optimum technique was implemented to perform the prediction of the dependent variable, Violent Crime Count (next day), at the city level.

# CHAPTER 5: FINDINGS

The findings are reported by study stages. In Stage 1, the findings of the word embedding models, the aggression classifiers, the loss classifiers, the credibility score regression algorithms, and the credibility classifiers are discussed. In Stage 2, the results from the various crime model regression analyses are revealed.

## 5.1 Stage 1 Findings

Stage 1 involved the preprocessing of data, creation and analysis of the domain-specific word embedding models, prediction and classification of aggression and loss in each gang-associated tweet, and prediction and classification of the credibility for each tweet.

### 5.1.1 Word Embedding Process

***Word Embedding Creation.*** After preprocessing the tweets, three different word embedding models were run. For each, the hyperparameters were tuned, and experimenting occurred with the models in a variety of ways. These included a baseline CountVectorizer model, a CBOW word embedding model, and a Skip-gram word embedding model. CountVectorizer (using *scipy.sparse.csr_matrix)* converted the tweets to a matrix of token counts, producing a sparse representation of the tweet content counts. The number of features is equal to the vocabulary size found by analyzing the data. The CBOW and Skip-gram models produced a dense representation of the tweet contents.

***Word Embedding Evaluation.*** Each embedding was retrieved and a word cloud created. The results showed that the words in the gang-affiliated text corpus were full of curse words, derogatory words, and non-standard English terms. The word cloud showed that several words appeared more frequently than others in the corpus. These included *sh*t, b**ch, mf, got, money, yall, d*mn, amp, love, n*gga, today, stats, dat, hoe, dont, da, lol, f**k,* and *man.*

Word2Vec's *distance* command was used to compare the semantic similarity using cosine similarity between words to cluster the synonyms referring to the same tweet feature from the documents (tweets) of the corpus. The higher the cosine value, the closer the two words relate. The cosine distance between pairs of words was measured in both dense word embedding architectures. For example, the 'most similar' words and emojis to the word *kill* was assessed. In one of the Skip-gram architecture models, the most similar words included ('write', 0.804), ('murk', 0.777), ('drown', 0.754), ('impress', 0.753), ('punch', 0.752), ('join', 0.745), ('disappear', 0.742), ('paint', 0.742), ('clap', 0.741), and ('visit', 0.741). The Skip-gram word embedding results were derived from the following parameters: *min_count* = 1, *size* = 100, *window* = 5, *iter*=20, *negative* =10, and *sg* = 1. This Skip-gram word embedding of more than 143,700 documents produced 64,970 unique words using the Python code: *print (len(list(w2vs.wv.vocab)))*. A CBOW model produced the following most similar words to the word *kill*: ('write', 0.825), ('cross', 0.796), ('meet', 0.791), ('smack', 0.790), ('steal', 0.789), ('blow', 0.788), ('murk', 0.788), ('fight', 0.777), ('catch', 0.772), and ('sell', 0.762).

Words and emojis most similar to the word *gun* from the CBOW architecture included ('pole', 0.816), ('pipe', 0.783), ('lick', 0.751), ('dirt', 0.718), ('pistol', 0.715), ('bag', 0.714), ('case', 0.713), ('40', 0.71), ('heat', 0.695), and a gun emoji (🔫, 0.693). Words and emojis most similar to the word *gun* from the Skip-gram architecture included ('pistol', 0.804), ('drum', 0.789), ('pipe', 0.776), ('necklace', 0.771), ('snatch', 0.766), ('pole', 0.765), ('tie', 0.762), ('nina', 0.762), ('muscle', 0.762), and ('burner', 0.760). These results emphasized that the domain-specific language from the gang community is, as suspected, very different from normal English meanings of the same words.

### 5.1.2 Training Data Labeling and Inter-Rater Reliability Analysis

To prepare the training data for the aggression and loss classifiers, two domain experts were employed to label an initial set (batch) of training data. A subsequent inter-rater reliability analysis was run on the labeling results. The experts read each tweet, viewing both the text and emoji for signals of aggression or loss. For example, they labeled most tweets with a gun emoji as aggressive. Inter-rater reliability between the two domain experts related to labeling aggression and loss in 626 tweets was calculated. Of the initial 626 tweets, for aggression, there was 57.38% inter-rater reliability (Cohen's Kappa (K)) in the ratings, 78.4% sensitivity, and 82.1% specificity. For loss, there was 65.29% inter-rater reliability (Cohen's Kappa (K)) in the ratings, 73.4% sensitivity, and 94.7% specificity. For aggression, there were 119 tweets where the raters did not agree; for loss, there were 50 tweets with disagreement. These were sent back to the two domain experts for reconciliation. Out of 626 labeled tweets, the first batch counts were 190 for aggression and 79 for loss. After assessing these results, the low frequency in each class, particularly the loss class, necessitated that more labeled tweets be obtained.

In the next set (batch) of tweets, more than 600 additional tweets were added, giving a total set of 1226 tweets in the labeled training data. Again, the two domain experts were employed to label the additional training observations, and an inter-rater reliability analysis on the labeling of the entire 1226 tweets was conducted. Inter-rater reliability between the two domain experts related to labeling aggression and loss was calculated and documented using the same procedure stated above.

Of the 1226 tweets labeled by the two domain experts, for aggression, there was 64.15% inter-rater reliability (Cohen's Kappa (K)) in the ratings (which is considered moderate/substantial agreement), a sensitivity of 82.46%, while specificity was 89.75%. For loss, there was 71.25% inter-rater reliability (Cohen's Kappa (K)) in the ratings (very good/substantial agreement), a sensitivity of 80.74%, while specificity was 95.60%. A Kappa result value ≤ 0 indicates no agreement, 0.01–0.20 indicates no to slight agreement, 0.21–0.40 reflects fair agreement, 0.41–0.60 is moderate agreement, 0.61–0.80 is substantial agreement, and 0.81–1.00 is almost perfect agreement (McHugh 2012), thus the metrics are both acceptable. Out of 1226 tweets, the counts were 211 for aggression and 1015 non-aggressive tweets. There were 135 tweets expressing loss and 1091 tweets with no loss-filled content. In all of the training data, for aggression, there were 141 tweets where the raters did not agree and 73 tweets for loss where a disagreement occurred. These were sent back to the two domain experts for reconciliation.

### 5.1.3 Algorithm 1 (Aggression Classification)

With the labeled data ready, Algorithm 1 (a classifier) was employed, which represents the classification of aggression of each gang-associated tweet. The annotated (labeled) tweets were split into training/test partitions of 80% and 20%. Three word embeddings were used separately as input features into several preliminary "baseline" classifier models for Aggression. These baseline algorithms included Logistic Regression (LR), Support Vector Machine Classifier (SVC), and Linear SVC. The best baseline model result for aggression using a dense word embedding was the LR model that achieved an

AUC (macro average) of .61,  accuracy of .76, weighted average precision of .77, weighted average recall of .76, and weighted average F1score of .77, with five-fold validation scores of .780, .797, .748, .695, and .760.  With the moderate accuracy, precision, recall, and AUC level (as compared to an AUC of .50 for a worthless model), these results are acceptable This model used the Skip-gram word embedding with a window size of two, a minimum word count of five, and a size of 400.

The best baseline model was then compared to the results of several neural network models, including CNN models and Multi-layer Perceptron (MLP) models. The MLP neural network model exhibited superior results over the best baseline model. Several different hyperparameters were tested (as necessary) with this classifier, and the best results came from tuning the parameters with *hidden_layer_sizes* = (150, 100, 50), *max_iter* (epochs) = 300, *activation* = ReLu, *solver*=adam, and *random_state* = 1. The rectified linear unit (ReLu or relu) activation function retrieves a neuron's output and maps it to the highest positive value or zero (if negative output) used for the hidden layer only. The best MLP model using the Skip-gram word embedding included an AUC (macro average) of .69, accuracy of .69, macro average precision of .47, macro average recall of .51, and macro F1 score of .73, using a window size of 10, a minimum word count of 5, and a size of 300. The best MLP model using the CountVectorizer technique included an AUC of .67, accuracy of .82, weighted average precision of .80, weighted average recall of .82, and weighted average F1 score of .81. The unweighted (macro) averages for the MLP model for loss, included an AUC of .673, accuracy of .82, precision of .73, recall of .67, and F1 score of .69. See Table 14 for the results on all of the Aggression classification algorithms. With its similar AUC score and higher accuracy, this model was chosen for the study.

**Table 14: Aggregession Classifiers and Results**

| Model Type | ACC | F1 | AUC | Recall | Precision | Vector Type/ Parameters | Train/ Test Split |
|---|---|---|---|---|---|---|---|
| MLP | .69 | .733 | .687 | .51 | .47 | Skip-gram, Iter=20, min_count = 5, size = 300, window = 10, negative =10 | 80/20 |
| MLP | .82 | .69 | .673 | .67 | .73 | CountVectorizer, hidden_layer_sizes=(150, 100, 50), max_ iter=300, activation = 'relu', solver = 'adam', random_state = 1 | 80/20 |
| Linear SVC | .83 | .70 | .669 | .67 | .77 | CountVectorizer, hidden_layer_sizes=(150, 100, 50), max_ iter=300, activation = 'relu', solver = 'adam', random_state = 1 | 80/20 |
| MLP | .68 | .680 | .660 | .489 | .45 | Skip-gram, Iter=20, min_count = 2, size = 100, window = 10, negative =10 | 80/20 |
| MLP | .627 | .730 | .655 | .0541 | .08 | CBOW, Window = 5, Size = 120, mincnt=2, neg=10, Iter=10, Batch size = 120 | 80/20 |
| MLP | .642 | .629 | .651 | .14 | .46 | Skip-gram, Window = 5, Size = 150, min_ cnt=2, neg=10, Iter=10 | 80/20 |
| MLP | .600 | .623 | .645 | .357 | .44 | Skip-gram, Window = 5, Size = 120, min_ cnt=2, neg=10, Iter=10, Batch size = 120 | 80/20 |
| MLP | .65 | .701 | .640 | .689 | .50 | Skip-gram, Iter=20, min_count = 5, size = 450, Window = 10, negative =10 | 80/20 |
| MLP | .652 | .672 | .637 | .815 | .595 | Skip-gram, Window = 2, Size = 400, min_ cnt=5, neg=10, Iter=20, Norm | 80/20 |
| MLP | .644 | .609 | .636 | .06 | .20 | CBOW, Window = 5, Size = 300, mincnt=2, neg=10, Iter=10 | 80/20 |

**Table 14: Aggression Classifiers and Results (continued)**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| LR | .83 | .66 | .631 | .63 | .79 | CountVectorizer, hidden_layer_sizes=(150, 100, 50), max_iter=300, activation = 'relu', solver = 'adam', random_state = 1 | 80/20 |
| CNN | .600 | .636 | .620 | .862 | .53 | CBOW, window =5, min_count =5,size = 12, iter=20, neg=10 | 80/20 |
| LR | .76 | .770 | .610 | .76 | .77 | Skip-gram, Iter=20, min_count = 5, size = 400, window = 2, negative =10 | 80/20 |
| MLP | .558 | .602 | .571 | .125 | .25 | CBOW, Window = 5, Size = 250, mincnt=2, neg=10, Iter=10 | 80/20 |
| MLP | .514 | .547 | .553 | .135 | .26 | Skip-gram, Window = 5, Size = 300, min_cnt=2, neg=10, Iter=10 | 80/20 |
| MLP | .620 | .285 | .523 | .194 | .38 | Skip-gram, Window=7, min_count = 5, size=120, iter=2,neg =10) | 80/20 |
| CNN | .785 | .039 | .505 | | .33 | Skip-gram, Iter=100, Batch size=28, Test, Unnorm | 80/20 |
| SVC | .789 | .44 | .500 | .50 | .39 | CountVectorizer, hidden_layer_sizes=(150, 100, 50), max_iter =300, activation = 'relu', solver = 'adam', random_state = 1 | 80/20 |
| CNN | .801 | .043 | .494 | | .13 | Skip-gram, Iter=250, Batch size=20, Test, UnNorm | 80/20 |
| CNN | .788 | .024 | .494 | | .11 | Skip-gram, Iter=100, Batch size=25, Test, UnNorm | 70/30 |
| CNN | .752 | .031 | .493 | | .14 | Skip-gram, Iter=200, Batch size=20, Test, Norm | 80/20 |
| CNN | .750 | .028 | .492 | | .13 | Skip-gram, Iter=270, Batch size=20, Test, UnNorm | 80/20 |

Reported from Test Set Results (with seed words)

### 5.1.4 Algorithm 2 (Loss Classification)

Algorithm 2, a classifier, represents the classification of loss of each gang-associated tweet. The researcher first ran baseline classifier models for loss. These baseline algorithms included Logistic Regression (LR), Support Vector Machine Classifier (SVC), and Linear SVC. The best baseline model result for loss was an LR model that achieved an AUC of .662, an accuracy of .78, recall of .66, precision of .78, and an average F1 score of .70 using the CountVectorizer embedding.

The baseline models were then compared to the results of the neural models. The best MLP model for loss with the Skip-gram word embedding included an AUC (macro average) of .637, accuracy of .602, precision of .60, recall of .815, and F1 score of .671, using a window size of 2, a minimum word count of 5, and a size of 400. Several different hyperparameters were tuned (as necessary) with this classifier, including results from the models used with CountVectorizer, with the hyperparameters of *hidden_layer_sizes* = 150, 100, 50, *max_iter* (epochs) = 300, *activation* = relu, *solver*=adam, and *random_state* = 1. This study employed the best MLP model for loss, which included an AUC of .663, accuracy of .88, weighted average precision of .88, weighted average recall of .88, and weighted average F1 score of .88. The unweighted (macro) averages for the MLP model for loss, included an AUC of .663, accuracy of .88, precision of .68, recall of .66, and F1 score of .67. With the excellent accuracy of this model, as well as the moderate F1, AUC, precision, and recall levels, these results are acceptable. See Table 15 for the Loss classifier model test set results.

**Table 15: Loss Classifiers and Results**

| Model Type | ACC | F1 | AUC | Recall | Precision | Vector Type/Parameters | Train/ Test Split |
|---|---|---|---|---|---|---|---|
| MLP | .88 | .67 | .663 | .66 | .68 | CountVectorizer, hidden_ layer_ sizes=(150, 100, 50), max_iter=300, activation = 'relu', solver = 'adam', random_state = 1 | 80/20 |
| LR | .78 | .70 | .662 | .66 | .78 | CountVectorizer, hidden_layer_sizes=(150, 100, 50), max_iter=300, activation = 'relu', solver = 'adam', random_state = 1 | 80/20 |
| MLP | .602 | .671 | .637 | .815 | .60 | Skip-gram, Window = 2, Size = 400, mincnt=5, neg=10, Iter=20 | 80/20 |
| LR | .87 | .87 | .610 | .87 | .86 | Skip-gram, Window = 2, Size = 400, mincnt=5, neg=10, Iter=20, Norm | 80/32 |
| Linear SVC | .78 | .81 | .590 | .78 | .85 | Skip-gram, Window = 2, Size = 400, mincnt=5, neg=10, Iter=20, Norm | 80/20 |
| MLP | .955 | .22 | .586 | .054 | .08 | CBOW, Window = 2, Size = 120, mincnt=3, neg=10, Iter=20, Batch size = 120 | 80/20 |
| SVC | .902 | .57 | .555 | .56 | .83 | CountVectorizer, hidden_layer_sizes=(150, 100, 50), max_iter=300, activation = 'relu', solver = 'adam', random_state = 1 | 80/20 |
| MLP | .922 | .00 | .500 | .357 | .44 | Skip-gram, Window = 2, Size = 120, mincnt=3, neg=10, Iter=20 | 80/20 |
| SVC | .90 | .90 | .500 | .85 | .81 | Skip-gram, Window = 2, Size = 400, mincnt=5, neg=10, Iter=20, Norm | 80/20 |

Reported from Test Set Results (with seed words)

### 5.1.5 Algorithm 3 (Credibility Regressions)

Algorithm 3 represents the prediction of credibility scores in each tweet with negative (aggression or loss) sentiment. The theoretical conceptual Co-UGS model shows eleven indicators that affect the credibility of a tweet. The empirical model starts with ten indicators, as all of the tweets in the model already exhibit negative sentiment. Further, to determine the association between these ten indicators, while also attempting to perform dimensionality reduction, the results from the factor analysis were used before running Algorithm 3.

#### 5.1.5.1 Aggression Credibility Regressions

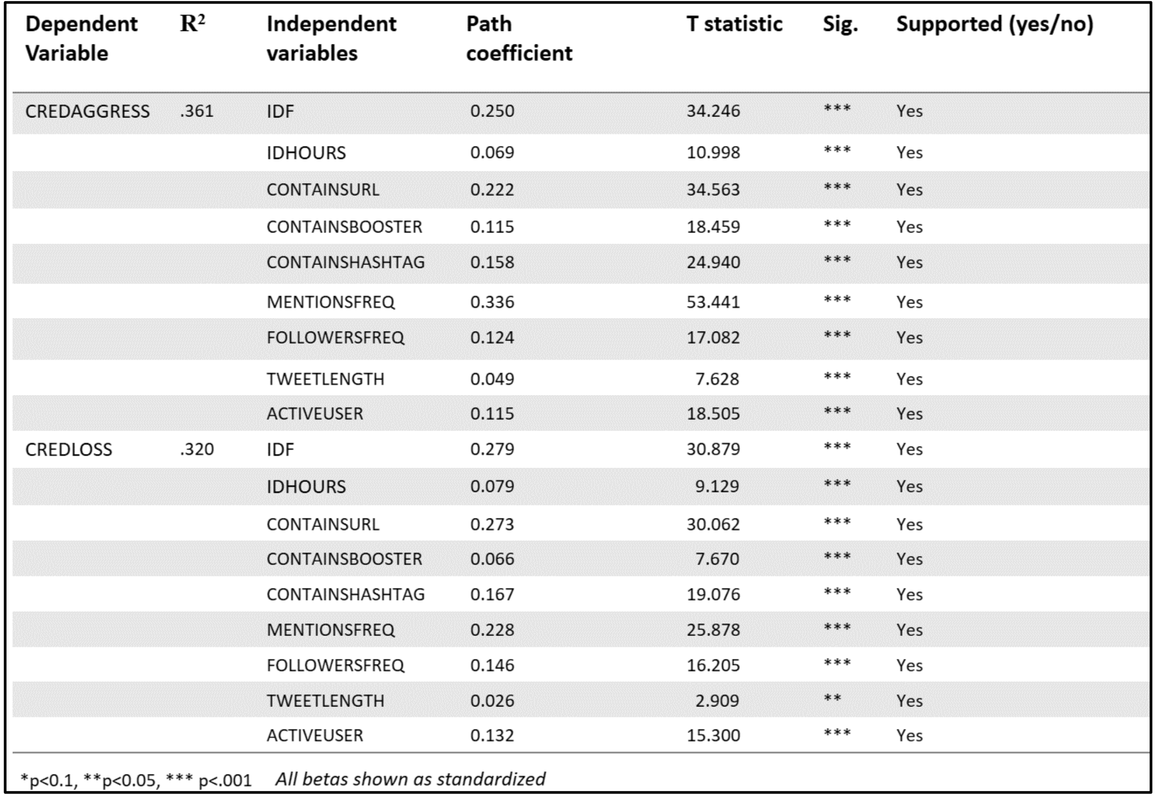
The credibility model for aggressive tweets was a multiple linear regression model that used aggression credibility score as the dependent variable and the nine credibility indicators as the independent variables. The results of this analysis show that, together, all nine independent variables have a significant strong link with aggression credibility score which, together, makes them significant predictors of aggression credibility score, such that as each of them increases by one unit, the aggression credibility score increases as well. Appendix D shows the results for all regressions used in the study.

Of the 16,466 tweets with aggression from Algorithm 1 (11.5% of the 143,700 tweets), 3,200 of them were randomly selected to train the model. The model exhibited no multicollinearity. The model reached significance in predicting aggression credibility score ($F=134.53$, $p < .001$), denoting that 27.5% of the variability in aggression credibility score was explained by the independent variables. The aggression credibility score was predicted significantly by all nine independent variables with positive standardized coefficient beta values. The standardized results of this analysis showed that the number of followers has a significant link with aggression credibility score (β= .124, t = 17.082, p < .001), which makes it a significant predictor of credible aggression. For each additional follower, the aggression credibility score of the tweet increases as well. Retweet and Favorites Frequency was also significant (β =.250, t=34.246, p<.001) and positively associates with credible aggression. Aggression Credibility Score was also significantly predicted by Mentions Frequency (β = .336, t = 53.441, p < .001), Contains Booster Word (β = .115, t = 18.459, p < .001), Retweet Time (hour) (β = .077, t = 5.108, p < .001), Tweet Length (β = .056, t = 3.651, p < .001), Active User (β = .036, t = 2.353, p < .001), and Contains URL (β = .035, t = 2.284, p < .001). As a result, all of these variables are confirmed, and this result suggests a significant effect in the same direction.

The algorithm then ran over the entire set of train and test data. This model performed better and reached significance in predicting an aggression credibility score with no multicollinearity. The model denoted that 36.1% of the variability in aggression credibility score, as a whole, is explained by the independent variables (F-Score = 1033.90, p < .001), as reported here with the positive standardized coefficient beta values. The results of this analysis showed that Mentions Frequency (β = .336, t = 53.441, p < .001) has the highest beta weight to credible aggression, while the Retweet and Favorites Frequency (IDF) was also significant (β = .250, t = 34.246, p < .001) and positively associates with credible aggression. Whether the tweet contains a URL (β = .222, t = 34.563, p < .001) and a hashtag (β = .158, t = 24.94, p < .001) significantly explains credible aggression. The number of followers also has a significant link with aggression credibility score (β = .124, t = 17.082, p < .001), which makes it a significant predictor of credible aggression. The Aggression Credibility Score was also significantly predicted by Contains

Booster Word (β = .115, t = 18.459, p < .001), Retweet Time (hour) (IDHOURS) (β = .069, t = 10.998, p < .001), Active User (β = .115, t = 18.505, p < .001), and Tweet Length (β = .049, t = 7.628, p < .001). As a result, all of these variables are confirmed, and this result suggests a significant effect in the same direction for each of them. The results from the standard coefficient beta values are shown in Table 16.

**Table 16: Credibility Score Regression Analyses and Standardized Results**



| Dependent Variable | $R^2$ | Independent variables | Path coefficient | T statistic | Sig. | Supported (yes/no) |
|---|---|---|---|---|---|---|
| CREDAGGRESS | .361 | IDF | 0.250 | 34.246 | *** | Yes |
| | | IDHOURS | 0.069 | 10.998 | *** | Yes |
| | | CONTAINSURL | 0.222 | 34.563 | *** | Yes |
| | | CONTAINSBOOSTER | 0.115 | 18.459 | *** | Yes |
| | | CONTAINSHASHTAG | 0.158 | 24.940 | *** | Yes |
| | | MENTIONSFREQ | 0.336 | 53.441 | *** | Yes |
| | | FOLLOWERSFREQ | 0.124 | 17.082 | *** | Yes |
| | | TWEETLENGTH | 0.049 | 7.628 | *** | Yes |
| | | ACTIVEUSER | 0.115 | 18.505 | *** | Yes |
| CREDLOSS | .320 | IDF | 0.279 | 30.879 | *** | Yes |
| | | IDHOURS | 0.079 | 9.129 | *** | Yes |
| | | CONTAINSURL | 0.273 | 30.062 | *** | Yes |
| | | CONTAINSBOOSTER | 0.066 | 7.670 | *** | Yes |
| | | CONTAINSHASHTAG | 0.167 | 19.076 | *** | Yes |
| | | MENTIONSFREQ | 0.228 | 25.878 | *** | Yes |
| | | FOLLOWERSFREQ | 0.146 | 16.205 | *** | Yes |
| | | TWEETLENGTH | 0.026 | 2.909 | ** | Yes |
| | | ACTIVEUSER | 0.132 | 15.300 | *** | Yes |

*p<0.1, **p<0.05, *** p<.001 *All betas shown as standardized*

### 5.1.5.2 Loss Credibility Regression

A similar model executed for credible tweets that contained loss. The credibility model for tweets containing loss was a multiple linear regression model that used loss credibility score as the dependent variable and the nine credibility loss indicators as the independent variables. The results of this analysis show that, together, all nine independent variables have a significant strong link with loss credibility score which, together, makes them significant predictors of loss credibility score, such that as each of them increases, the loss credibility score increases as well. The standardized coefficient beta results of this analysis showed that the sum of retweet frequency and favorites count has a significant link with loss credibility score (per period) ($\beta = .24$, $t = 14.034$, $p < .05$), which makes it a significant predictor of credible loss; once interactive dissemination frequency increases, the loss credibility score increases as well. As a result, this variable is confirmed, and this result suggests a significant effect in the same direction. Similarly, the remaining other variables were tested, all of which suggested a significant positive effect. These include Followers ($\beta = .20$, $t = 11.790$, $p < .05$), Mentions Frequency ($\beta = .139$, $t = 9.098$, $p < .05$), Contains URL ($\beta = .121$, $t = 7.808$, $p < .05$), Contains Hashtag ($\beta = .103$, $t = 6.769$, $p < .05$), and Contains Booster Word ($\beta = .045$ $t = 2.987$, $p < .05$). The model reached significance in predicting Loss Credibility Score ($F= 127.12$, $p < .05$), as revealed in the ANOVA results. The model, as a whole ($p < .05$), explains that 20.3% of the variability in loss credibility score is explained by the independent variables.

The algorithm then ran over the entire set of train and test data. This model performed better and reached significance in predicting Loss Credibility Score ($F=477.78$, $p < .001$) with no multicollinearity. The model denoted that 32.0% of the variability in loss credibility score, as a whole, is explained by the independent variables ($p < .001$). The loss credibility score was predicted significantly by all nine independent variables, as reported here with the positive standardized coefficient beta values. The standardized results of this analysis showed that the variable Retweet and Favorites Frequency (IDF) was significant ($\beta = .279$, $t = 30.879$, $p < .001$) and positively associates with credible loss (ranking the highest). Loss Credibility Score was also significantly predicted by Contains URL ($\beta =$

.273, t = 30.062, p < .001), Mentions Frequency (β = .228, t = 25.878, p < .001), Contains Hashtag (β = .167, t = 19.076, p < .001), Followers (β = .146, t = 16.205, p < .001), Active User (β = .132, t = 15.3, p < .001), Retweet Time (hour) (IDHOURS) (β = .079, t = 9.129, p < .001), Contains Booster Word (β = .066, t = 7.670, p < .001), and Tweet Length (β = .026, t = 2.909, p < .01). As a result, all of these variables are confirmed, and this result suggests a significant effect in the same direction. The results from the standard coefficient beta values are shown in Table 16.

### 5.1.6 Algorithm 4 (Credibility Classifications)

An alternative to the research model, the Direct Effects Model, used credible aggression frequency (instead of aggression credibility score) and credible loss frequency (instead of loss credibility score) to predict violent crime counts (next period). Thus, two classifiers executed to prepare the data for the Direct Effects Model. To label the training data for these two classifications, each tweet in the training data was classified as credible (1) or not credible (0) by using the basis credibility scores created earlier from the nine credibility indicators. In the training data, 3,500 tweets were used to train the classifiers. In all, the classifiers were used to predict 9168 tweets.

#### 5.1.6.1 Aggression Credibility Classifications

Using logistic regression, the Credible Aggression classifier was tested on the training data. When analyzing the logistic regression results, overall significance (Cox & Snell $R^2$ = .347; 89% accuracy, p < .001) was found. All variables were positively associated with credible aggression, such that with one unit increase in each, the logit of credible aggression increases by the beta amount. Reported betas are Retweet and Favorites

Frequency (β=2.899, Exp(β)=18.16, p<.001), Mentions Frequency (β=.506, Exp(β)=1.659, p<.001), Contains URL (β=.198, Exp(β)=1.219, p<.05), and Followers (β=1.108, Exp(β)=3.028, p<.001). Other variables include Contains Booster Word (β=.386, Exp(β)=1.471, p<.001), Contains Hashtag (β=.560, Exp(β)=1.750, p<.001), Active User (β=.672, Exp(β)=1.959, p=.999), Retweet Time (hour) (β=.050, Exp(β)=1.051, p=.248), and Tweet Length (β=.061, Exp(β)=1.063, p=.223).

The variables Active User, Tweet Length, and Retweet Time (hour) were initially not statistically significant. Retweet Time (hour) became significant with the removal of four outliers in the data (tweets with a retweet time more than three standard deviations away from the mean). Similarly, when three outlier tweets (with a large number of emojis) were handled, Tweet Length became significant. The results were better after the adjustments of those outliers (Cox & Snell $R^2$ = .349, 88.9% accuracy, p < .001). Reported standardized betas are Retweet and Favorites Frequency (β=2.837, Exp(β)=17.063, p<.001), Mentions Frequency (β=.502, Exp(β)=1.651, p<.001), Contains URL (β=.186, Exp(β)=1.204, p<.01), Followers (β=1.116, Exp(β)=3.053, p<.001), Contains Booster Word (β=.384, Exp(β)=1.468, p<.001), Tweet Length (β=.205, Exp(β)=1.227, p<.05), Retweet Time (β=.739, Exp(β)=2.094, p<.01), and Contains Hashtag (β=.547, Exp(β)=1.727, p<.001).

However, because the variable Active User was still not significant (β=.669, Exp(β)=1.952, p=.999) due to the low distribution of inactive users in the data (which this study defined differently from other research), Active User was removed from the model, and the regression reran. Results were reported after this removal (Cox & Snell $R^2$ = .349, 88.9% accuracy, p < .001). Reported standardized beta and significance values are Retweet and Favorites Frequency (β=2.837, Exp(β)=17.071, p<.001), Mentions Frequency (β=.502, Exp(β)=1.652, p<.001), Contains URL (β=.186, Exp(β)=1.204, p<.01), Followers (β=1.117, Exp(β)=3.056, p<.001), Contains Booster Word (β=.384, Exp(β)=1.469, p<.001), Tweet Length (β=.206, Exp(β)=1.228, p<.01), Retweet Time (IDHOURS) (β=.740, Exp(β)=2.096, p<.01), and Contains Hashtag (β=.544, Exp(β)=1.723, p<.001).

The standardized interpretation of these findings suggests that if a tweet with aggression contains a URL, the odds ratio associated with a one standard deviation increase, produces, on average, a 1.2 increase in the log odds of tweet credibility. Restated, the tweet is 1.2 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 1.19 (standard deviation) times (determined after running a simple logistic regression model), if alone. If the tweet contains a hashtag, it is 1.72 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 1.43 (standard deviation) times more likely, if no other variables exist. If the tweet contains a booster word, it is 1.47 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 1.3 (standard deviation) times more likely, if alone. For each unit increase in mentions frequency, the tweet is 1.65 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables or 1.49 alone. For each unit increase in retweet and favorites frequency, the tweet is 17.1 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 37.82 alone. For each additional follower, the tweet is 3.06 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 4.12 alone. For each increase in tweet length, the tweet is 1.23 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables or 1.67 alone. For each additional hour it takes to send the first retweet of a tweet, the tweet is 2.1 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables or 3.26 alone.

### 5.1.6.2 Loss Credibility Classifications

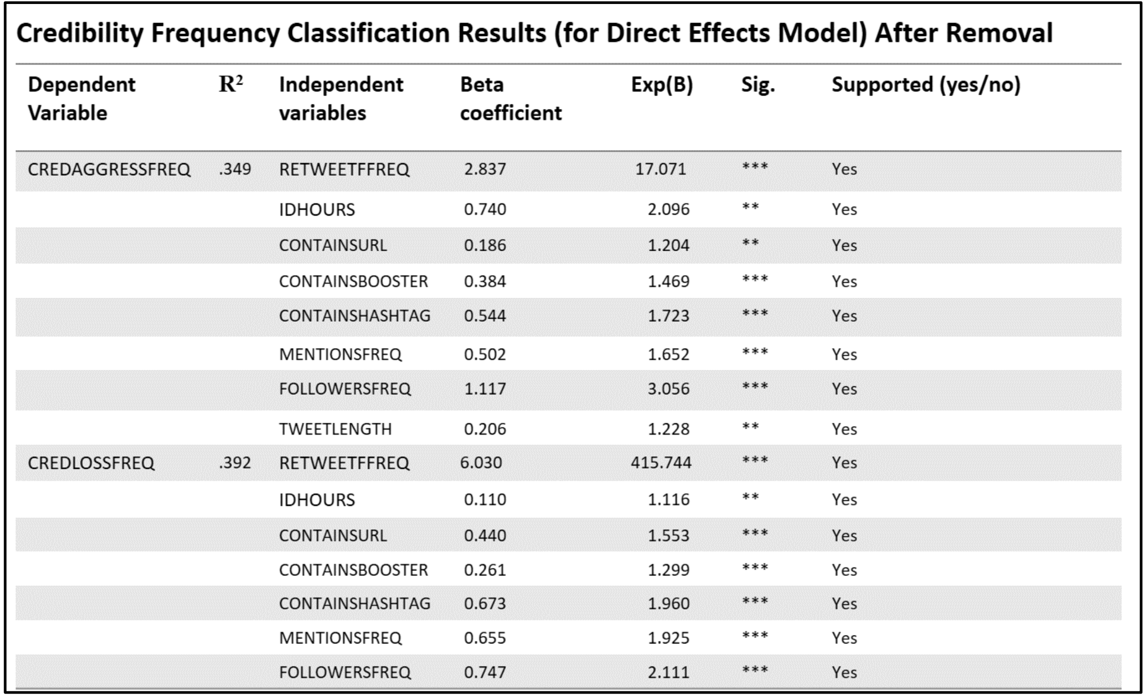
Using logistic regression, the Credible Loss classifier ran on the training data for the tweets classified with loss. The logistic regression was significant and reported Cox & Snell $R^2$ = .392, with an 89.8% accuracy. All variables were positively associated with credible loss, such that with one unit increase in each, the logit of credible loss increases by the beta amounts. Reported betas are Retweet and Favorites Frequency (β=6.027, Exp(β)=414.332, p<.001), Mentions Frequency (β=.654, Exp(β) =1.924, p<.001), Contains URL (β=.439, Exp(β)=1.552, p<.001), and Followers (β=.747, Exp(β)=2.110, p<.001). Other variables include Contains Booster Words (β=.261, Exp(β) =1.299, p<.001), Contains Hashtag (β=.674, Exp(β) =1.962, p<.001), Active User (β=.454, Exp(β) =1.572, p=.999), Retweet Time (Hr) (β=.110, Exp(β) =1.116, p<.05), and Tweet Length (β=.001, Exp(β) =1.001, p=.983).

When testing this classifier, the variables Active User and Tweet Length were initially not statistically significant. However, unlike with the aggression classifier, outliers were not the issue. Though sufficiently correlated to the Credible Loss class (without the risk of multicollinearity) and significant in the simple logistic regression, Tweet Length became non-significant when other independent variables were added to the multiple logistic regression. The variable Active User was not significant due to the low distribution of inactive users in the data. Thus, these variables were removed, one by one, and the regression run again.

This resulted in the significance of all variables with the same regression evaluation metric values. Reported betas are (in order of importance) Retweet and Favorites Frequency (β=6.030, Exp(β)=415.744, p<.001), Followers (β=.747, Exp(β)=2.111, p<.001), Contains Hashtag (β=.673, Exp(β) =1.960, p<.001), Mentions Frequency (β=.655, Exp(β) =1.925, p<.001), Contains URL (β=.440 Exp(β)=1.553, p<.001), Contains Booster Words (β=.261, Exp(β) =1.299, p<.001), and Retweet Time (Hrs) (β=.110, Exp(β) =1.116, p<.01).

To interpret this, the findings suggest that if the tweet with loss contains a URL, the odds ratio associated with a standard deviation increase, produces, on average, a 1.6 increase in the log odds of tweet credibility. Restated, the tweet is 1.6 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 1.465, if alone. If the tweet contains a hashtag, it is 1.96 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 1.61, if no other variables exist. If the tweet contains a booster word, it is 1.3 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 1.15 alone. For each unit increase in mentions frequency, the tweet is 1.925 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables or 1.2, if alone. For each unit increase in retweet and favorites frequency, the tweet is 416 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 277, if alone. For each additional follower, the tweet is 2.11 (standard deviation) times more likely to be credible, controlling for individual differences in the other predictor variables, or 2.5 alone (Table 17).

**Table 17: Credibility Frequency Classification Analyses and Standardized Results**



**Credibility Frequency Classification Results (for Direct Effects Model) After Removal**

| Dependent Variable | $R^2$ | Independent variables | Beta coefficient | Exp(B) | Sig. | Supported (yes/no) |
|---|---|---|---|---|---|---|
| CREDAGGRESSFREQ | .349 | RETWEETFFREQ | 2.837 | 17.071 | *** | Yes |
| | | IDHOURS | 0.740 | 2.096 | ** | Yes |
| | | CONTAINSURL | 0.186 | 1.204 | ** | Yes |
| | | CONTAINSBOOSTER | 0.384 | 1.469 | *** | Yes |
| | | CONTAINSHASHTAG | 0.544 | 1.723 | *** | Yes |
| | | MENTIONSFREQ | 0.502 | 1.652 | *** | Yes |
| | | FOLLOWERSFREQ | 1.117 | 3.056 | *** | Yes |
| | | TWEETLENGTH | 0.206 | 1.228 | ** | Yes |
| CREDLOSSFREQ | .392 | RETWEETFFREQ | 6.030 | 415.744 | *** | Yes |
| | | IDHOURS | 0.110 | 1.116 | ** | Yes |
| | | CONTAINSURL | 0.440 | 1.553 | *** | Yes |
| | | CONTAINSBOOSTER | 0.261 | 1.299 | *** | Yes |
| | | CONTAINSHASHTAG | 0.673 | 1.960 | *** | Yes |
| | | MENTIONSFREQ | 0.655 | 1.925 | *** | Yes |
| | | FOLLOWERSFREQ | 0.747 | 2.111 | *** | Yes |

## 5.2 Stage 2 Findings

In Stage 2 of this study, all values from the tweets with credible aggression and credible loss were aggregated into daily results and used, along with other aggregated independent variables and aggregated control variables, to predict the city violent crime count (next period). Here, the main research model (including a test for mediation) and the alternative Direct Effects Model were employed.

### 5.2.1 Research Model

The research model predicted violent crime counts for the next time period (day). It used independent variables of aggression credibility score, loss credibility score, gang-affiliated tweet frequency, credible aggression and loss tweet frequency, interactive dissemination frequency, and interactive dissemination speed. The model also controlled for average daily temperature (F), historical crime rate, weekday, average daily tweet time (hour), whether a major event occurred on the tweet day, prior periods' violent crime counts, and quarterly seasonal indicators.

To prepare the data, the detected credible aggression or credible loss were aggregated into daily amounts. Over the five years of data, this provided 1,311 rows of data. Of these, the dependent variable, violent crime count (next day), was automatically labeled with historical data sourced from the City of Chicago data portal. Algorithms then ran over the data, using various simple and multiple linear regressions.

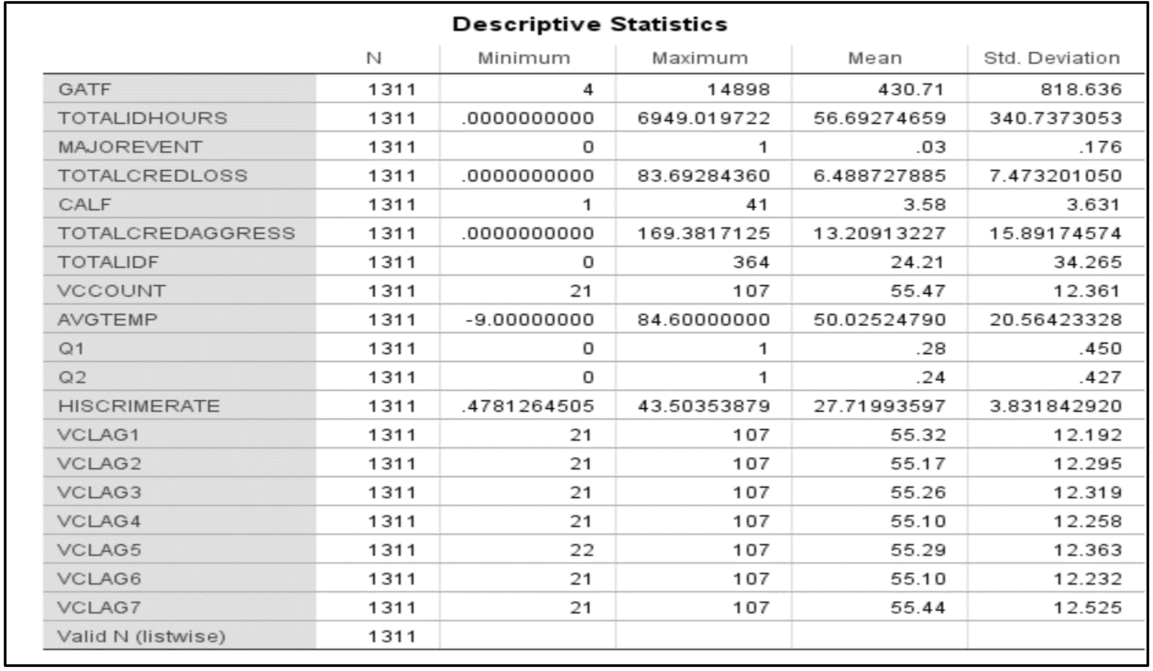
### 5.2.1.1 Descriptive Statistics and Correlation Analysis

Table 18 shows the descriptive statistics of all variables used in the model. Violent crime count (VCCOUNT) was on average 55.47 instances (SD = 12.36). Total Aggression Credibility Score (per period) (TOTALCREDAGGRESS) was 13.21 (SD = 15.89), while Total Loss Credibility Score (per period) (TOTALCREDLOSS) was 6.49 (SD = 7.47). On average, the Gang-Affiliated Tweet Frequency (GATF) (per period) was at a moderate level (M = 430.71, SD = 818.64). Total Interactive Dissemination Frequency (TOTALIDF) (per period) was 24.21 (SD = 34.27), while Average Interactive Dissemination Speed (measured in Retweet Time, in hours) was 56.69 (SD = 340.74).

The results of the correlation analysis revealed that there was a negative correlation between Total Aggression Credibility Score (per period) and Violent Crime Count (per period) (r = -0.169, p < .01) and a slight positive correlation between Total Loss Credibility Score (per period) and Violent Crime Count (per period) (r = 0.003). Gang-Affiliated

Tweet Frequency is negatively associated with violent crime (r = -0.024), as is Credible Aggression and Loss Frequency (CALF) (r = -0.124, p < .01). Total Interactive Dissemination Frequency is negatively associated with Violent Crime Count (r = -0.149, p < .01), while Average Retweet Time (AVGIDHOURS) is also negatively related to Violent Crime Count (r = -0.070, p < .05) (Table 19).

**Table 18: Research Model Descriptive Statistics**

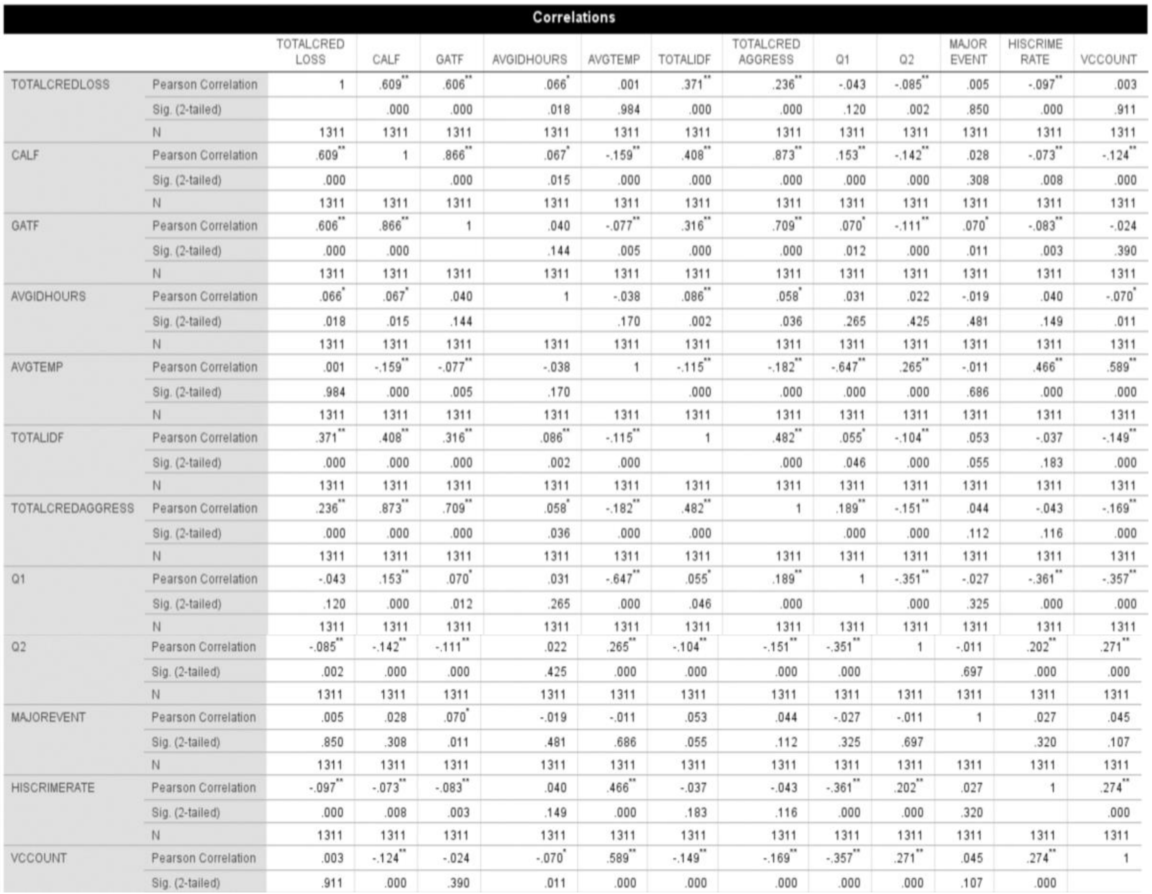


Descriptive Statistics

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| GATF | 1311 | 4 | 14898 | 430.71 | 818.636 |
| TOTALIDHOURS | 1311 | .0000000000 | 6949.019722 | 56.69274659 | 340.7373053 |
| MAJOREVENT | 1311 | 0 | 1 | .03 | .176 |
| TOTALCREDLOSS | 1311 | .0000000000 | 83.69284360 | 6.488727885 | 7.473201050 |
| CALF | 1311 | 1 | 41 | 3.58 | 3.631 |
| TOTALCREDAGGRESS | 1311 | .0000000000 | 169.3817125 | 13.20913227 | 15.89174574 |
| TOTALIDF | 1311 | 0 | 364 | 24.21 | 34.265 |
| VCCOUNT | 1311 | 21 | 107 | 55.47 | 12.361 |
| AVGTEMP | 1311 | -9.00000000 | 84.60000000 | 50.02524790 | 20.56423328 |
| Q1 | 1311 | 0 | 1 | .28 | .450 |
| Q2 | 1311 | 0 | 1 | .24 | .427 |
| HISCRIMERATE | 1311 | .4781264505 | 43.50353879 | 27.71993597 | 3.831842920 |
| VCLAG1 | 1311 | 21 | 107 | 55.32 | 12.192 |
| VCLAG2 | 1311 | 21 | 107 | 55.17 | 12.295 |
| VCLAG3 | 1311 | 21 | 107 | 55.26 | 12.319 |
| VCLAG4 | 1311 | 21 | 107 | 55.10 | 12.258 |
| VCLAG5 | 1311 | 22 | 107 | 55.29 | 12.363 |
| VCLAG6 | 1311 | 21 | 107 | 55.10 | 12.232 |
| VCLAG7 | 1311 | 21 | 107 | 55.44 | 12.525 |
| Valid N (listwise) | 1311 | | | | |

All assumptions were assessed for linear regression in the research model. The research model outcome variable is not completely normally distributed, indicated by the variable histogram (Appendix E) and the Shapiro-Wilk Test of Normality. This is most likely due to the presence of a few outliers in the data. Further, there was a need to transform some of the independent variables due to the existence of either extreme outliers or a non-linear relationship between two variables. As mentioned earlier, only four observations that were outliers were removed due to an extreme value in retweet speed, while three others were transformed due to extreme values in tweet length. All observations represent real and accurate values in the data. Some variables, however, were transformed due to a curvilinear relationship with Violent Crime Count (next period).

**Table 19: Research Model Correlation Matrix**



Correlations

| | | TOTALCRED LOSS | CALF | GATF | AVGIDHOURS | AVGTEMP | TOTALIDF | TOTALCRED AGGRESS | Q1 | Q2 | MAJOR EVENT | HISCRIME RATE | VCCOUNT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TOTALCREDLOSS | Pearson Correlation | 1 | .609** | .606** | .066* | .001 | .371** | .236** | -.043 | -.085** | .005 | -.097** | .003 |
| | Sig. (2-tailed) | | .000 | .000 | .018 | .984 | .000 | .000 | .120 | .002 | .850 | .000 | .911 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| CALF | Pearson Correlation | .609** | 1 | .866** | .067* | -.159** | .408** | .873** | .153** | -.142** | .028 | -.073** | -.124** |
| | Sig. (2-tailed) | .000 | | .000 | .015 | .000 | .000 | .000 | .000 | .000 | .308 | .008 | .000 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| GATF | Pearson Correlation | .606** | .866** | 1 | .040 | -.077** | .316** | .709** | .070* | -.111** | .070* | -.083** | -.024 |
| | Sig. (2-tailed) | .000 | .000 | | .144 | .005 | .000 | .000 | .012 | .000 | .011 | .003 | .390 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| AVGIDHOURS | Pearson Correlation | .066* | .067* | .040 | 1 | -.038 | .086** | .058* | .031 | .022 | -.019 | .040 | -.070* |
| | Sig. (2-tailed) | .018 | .015 | .144 | | .170 | .002 | .036 | .265 | .425 | .481 | .149 | .011 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| AVGTEMP | Pearson Correlation | .001 | -.159** | -.077** | -.038 | 1 | -.115** | -.182** | -.647** | .265** | -.011 | .466** | .589** |
| | Sig. (2-tailed) | .984 | .000 | .005 | .170 | | .000 | .000 | .000 | .000 | .686 | .000 | .000 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| TOTALIDF | Pearson Correlation | .371** | .408** | .316** | .086** | -.115** | 1 | .482** | .055* | -.104** | .053 | -.037 | -.149** |
| | Sig. (2-tailed) | .000 | .000 | .000 | .002 | .000 | | .000 | .046 | .000 | .055 | .183 | .000 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| TOTALCREDAGGRESS | Pearson Correlation | .236** | .873** | .709** | .058* | -.182** | .482** | 1 | .189** | -.151** | .044 | -.043 | -.169** |
| | Sig. (2-tailed) | .000 | .000 | .000 | .036 | .000 | .000 | | .000 | .000 | .112 | .116 | .000 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| Q1 | Pearson Correlation | -.043 | .153** | .070* | .031 | -.647** | .055* | .189** | 1 | -.351** | -.027 | -.361** | -.357** |
| | Sig. (2-tailed) | .120 | .000 | .012 | .265 | .000 | .046 | .000 | | .000 | .325 | .000 | .000 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| Q2 | Pearson Correlation | -.085** | -.142** | -.111** | .022 | .265** | -.104** | -.151** | -.351** | 1 | -.011 | .202** | .271** |
| | Sig. (2-tailed) | .002 | .000 | .000 | .425 | .000 | .000 | .000 | .000 | | .697 | .000 | .000 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| MAJOREVENT | Pearson Correlation | .005 | .028 | .070* | -.019 | -.011 | .053 | .044 | -.027 | -.011 | 1 | .027 | .045 |
| | Sig. (2-tailed) | .850 | .308 | .011 | .481 | .686 | .055 | .112 | .325 | .697 | | .320 | .107 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| HISCRIMERATE | Pearson Correlation | -.097** | -.073** | -.083** | .040 | .466** | -.037 | -.043 | -.361** | .202** | .027 | 1 | .274** |
| | Sig. (2-tailed) | .000 | .008 | .003 | .149 | .000 | .183 | .116 | .000 | .000 | .320 | | .000 |
| | N | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 | 1311 |
| VCCOUNT | Pearson Correlation | .003 | -.124** | -.024 | -.070* | .589** | -.149** | -.169** | -.357** | .271** | .045 | .274** | 1 |
| | Sig. (2-tailed) | .911 | .000 | .390 | .011 | .000 | .000 | .000 | .000 | .000 | .107 | .000 | |

The assumption of no multicollinearity was met with the study variables, as the correlation metric for most variables was (-0.70 < r < 0.70), and none of the VIF values were above 10. However, there were two exceptions. Credible Aggression and Loss Frequency (CALF) introduced multicollinearity into the model with Total Aggression Credibility Score (TOTALCREDAGGRESS) and with Gang-Affiliated Tweet Frequency (GATF). The second exception was between GATF and TOTALCREDAGGRESS. Thus, CALF and GATF were removed from the multiple regression model. Finally, the assumptions of linearity and homoscedasticity were met, as the scatterplots of the standardized residual on the standardized predicted value did not curve or exhibit an undesirable pattern in most variables (exceptions noted below).

### 5.2.1.2 Simple Linear Regression Analyses

The first hypothesis was initially tested utilizing a simple linear regression that employed Violent Crime Count (next period) as the dependent variable. It used the daily Total Aggression Credibility Score as the independent variable (and its square to account for the curvilinear association). The standardized results of this analysis show that credible aggression has a significant link with violent crime count (per period), initially negative to a point ($\beta = -.345$, $t = -6.774$, $p < .001$), and then curving toward the positive ($\beta = .208$, $t = 4.077$, $p < .001$) at a score of approximately 75, which makes it a significant predictor of violent crime count. If the total aggression credibility score increases, violent crime count initially decreases but eventually curves at approximately 75 and increases. This result suggests a significant effect, ultimately in the same direction, explaining 4.1% ($R^2$) of the variance in violent crime count in the data. As a result, this variable is confirmed, and this result suggests a significant eventual effect in the same direction.

The second hypothesis was initially tested using a simple linear regression that employed Violent Crime Count (next period) as the dependent variable and Total Loss Credibility Score as the independent variable. The results of this analysis show that total loss credibility score (per period) does not have a significant link with violent crime count (per period) ($\beta = .003$, $t = .112$, $p > .05$), which does not make it a significant predictor of violent crime count alone.

The third hypothesis was initially tested through a linear regression that used Violent Crime Count (next period) as the dependent variable and Gang-Affiliated Tweet

Frequency (per period) (GATF) as the independent variable. The standardized results of this analysis show that Gang-Affiliated Tweet Frequency has a significant link with Violent crime count (per period), initially negative to a point (β = -.266, t = -4.834, p < .001), and then curving toward the positive (β = .249, t = 5.332, p < .001) at a score of approximately 6,500, which makes it a significant predictor of violent crime count. If the total gang-affiliated tweet frequency increases, violent crime count initially decreases but eventually curves at approximately 6,500 and increases. The results of this analysis show that gang-affiliated tweet frequency (per period) has a significant link with violent crime count (per period), which makes it a predictor of violent crime count. As a result, this variable is confirmed, and this result suggests a significant eventual effect in the same direction.

The fourth hypothesis was initially tested through a simple linear regression that used Violent Crime Count (next period) as the dependent variable and total Credible Aggression and Loss Tweet Frequency (CALF) (per period) as the independent variable. The standardized results of this analysis show that CALF (per period) has a significant link with violent crime count (per period) (β = -.124, t = -4.51, p < .001), which makes it a significant predictor of violent crime count; once credible aggression and loss tweet frequency increases, violent crime count decreases.

The fifth hypothesis was initially tested utilizing a simple linear regression that employed Violent Crime Count (next period) as the dependent variable and Total Interactive Dissemination Frequency (retweets, mentions, and favorites) (per period) as the independent variable. The standardized results of this analysis show that interactive dissemination frequency (per period) has a significant link with violent crime count (per period) (β = -.149, t = -5.44, p < .001), which makes it a significant predictor of violent crime count; once interactive dissemination frequency increases, violent crime count decreases. This explains 2.2% ($R^2$) of the variance in the model.

The sixth hypothesis was initially tested using a simple linear regression that employed Violent Crime Count (next period) as the dependent variable and Interactive Dissemination (retweet) Hours (per period) as the independent variable. While the construct theoretically has a positive association with violent crime count, the construct with the average retweet time was measured in hours. Theoretically, retweet time (in hours) has a negative association with violent crime counts. The standardized results of this analysis show that average retweet time (in hours, per period) has a significant negative link with violent crime count (per period) ($\beta = -.070$, $t = -2.55$, $p < .05$), which makes it a significant predictor of violent crime count; once retweet time increases (resulting in a decrease in speed), violent crime decreases. Thus, looking at the conceptual model, this means that as interactive dissemination speed increases, violent crime count increases as well. As a result, this variable is confirmed, and this result suggests a significant effect for interactive dissemination speed in the same direction as violent crime count.

#### 5.2.1.3 Research Model Hierarchical Multiple Regression Analysis

In the research model, four of the six of the variables used in the simple linear regressions (whether initially confirmed or not) are combined into a multiple regression to further assess them. The independent variables included Total Aggression Credibility Score (per period), Total Loss Credibility Score (per period), Total Interactive Dissemination Frequency (per period), and Average Interactive Dissemination Hours (per period). CALF and GATF were removed due to multicollinearity.

Further, the test controls for several confounding variables and their impact on the model constructs. This test was executed via a Hierarchical Multiple Regression that used Violent Crime Count (next period) as the dependent variable, and Average Temperature (F), Historical Crime Rate per 100,000 on tweet day, Weekday, Average Tweet Time (hour), Major Event on tweet day, and prior seven periods' violent crime counts (including seasonality indicator variables) as the control variables.

The results of this analysis show that, together, Total Aggression Credibility Score, Total Loss Credibility Score, Total Interactive Dissemination Frequency, Average Interactive Dissemination Hours, and several of the control variables have a significant link with the next day's violent crime count. The control variables include Average Temperature, Major Event, violent crime lags (VCLAG1 through VCLAG6), and two of the quarterly seasonality indicators (Q1 and Q2). The resulting regression was significant (F(15,1295)=66.212, p<.001), with an $R^2$ of .434 (Appendix D).

***Independent Variables.*** With the exception of Total Interactive Dissemination Frequency, all of the rest of the significant associations were also positive (accounting for the change in sign between the constructs of Average Interactive Dissemination Speed and its measure, Average Interactive Dissemination Hours). Together, this makes them significant predictors of violent crime count, such that as each of them increases by one unit, all other variables held equal, violent crime count increases as well.

Total Aggression Credibility Score has a curvilinear (initially negative, then positive) relationship with violent crime count. Violent crime count (per period) was predicted by Total Aggression Credibility Score ($\beta$ = -.133, t = -3.081, p < .001), which, though initially negative, curved to the positive ($\beta$ = .133, t = 3.314, p < .01). After controlling for average daily temperature, a major event on the tweet day, prior period violent crime counts, and seasonality, for every one standard deviation increase in the credible aggression, violent crime count decreased initially by .133, but later increased by .133. Violent crime count (per period) was also predicted by Total Loss Credibility Score (β = .063, t = 2.768, p < .001). After controlling for average daily temperature, a major event on the tweet day, prior period violent crime counts, and seasonality, for every one standard deviation increase in the credible loss, violent crime count increased by .063. Violent crime count (per period) was also predicted by Average Interactive Dissemination Hours (β = -.035, t = -1.682, p < .001). After controlling for average daily temperature, a major event on the tweet day, prior period violent crime counts, and seasonality, for every one standard deviation increase in AVGIDHOURS, violent crime count decreased by 0.035 (suggesting that Interactive Dissemination Speed, however, increased by that amount). As a result, the hypotheses for these predictor variables (H1, H2, and H6c) are confirmed, as this result suggests a significant effect in the same direction.

Violent crime count (per period) was predicted by Total Interactive Dissemination Frequency (β = -.046, t = -1.804, p < .001). After controlling for average daily temperature, a major event on tweet day, prior period violent crime counts, and seasonality, for every one standard deviation increase in TOTALIDF, violent crime count decreased by .046. As a result, the hypothesis for this predictor variable (H5c) is not confirmed, as this result suggests a significant effect in the opposite direction.

***Control Variables***. The results of this analysis show that average temperature (F) has a strong positive link with violent crime count (per period) (β = .360, t = 10.120, p < .001), which makes it a significant predictor of violent crime count; once average daily temperature increases, violent crime count increases as well. As a result, this control variable is confirmed, as this result suggests a significant effect in the same direction.

The results of this analysis show that if a major event occurred on the date, violent crime count (per period) increased ($\beta = .054$, $t = 2.553$, $p < .05$). As a result, this control variable is confirmed, as it suggests a significant effect in the same direction.

The dependent variable, violent crime count, is autoregressive in these data, and (in an initial test), seven prior lags (one week) were significant. These seven lags were also assessed for multicollinearity, but none was found. The results of this regression analysis also show that violent crime count lags 1-6 each have a significant positive link with violent crime count (per period) which makes each of them a significant predictor of violent crime count. For lag 1, if the prior day’s violent crime increases, the violent crime count increases as well ($\beta = .090$, $t = 3.280$, $p < .01$). For lag 2, if the prior second day’s violent crime increases, the violent crime count increases as well ($\beta = .083$, $t = 3.037$, $p < .01$). For lag 3, if the prior third day’s violent crime increases, the violent crime count increases as well ($\beta = .065$, $t = 2.404$, $p < .05$). For lag 4, if the prior fourth day’s violent crime increases, the violent crime count increases as well ($\beta = .097$, $t = 3.563$, $p < .01$). For lag 5, if the prior fifth day’s violent crime increases, the violent crime count increases as well ($\beta = .052$, $t = 1.935$, $p < .05$). For lag 6, if the prior sixth day’s violent crime increases, the violent crime count increases as well ($\beta = .080$, $t = 2.943$, $p < .05$). As a result, these six control variables, representing violent crime count lags, are confirmed, as this result suggests a significant effect in the same direction as violent crime count for each.

Further, two of the quarterly seasonality control variables, Q1 and Q2, were positively and significantly associated with violent crime count. For Q1, if the tweet day falls in January, February, or March, the violent crime count increases ($\beta = .097$, $t = 3.378$, $p < .001$), relative to Q4 (the baseline quarter). For Q2, if the day falls in April, May, or June, the violent crime count increases ($\beta = .084$, $t = 3.627$, $p < .05$), relative to Q4, but not by as high of a factor as in Q1. As a result, these two control variables, Q1 and Q2, are confirmed, as this result suggests a significant effect in the same direction as violent crime count for each.

The variable Q3 was negatively correlated to violent crime count but not significant, so it was removed from the model. The beta value for Q4 is captured in the intercept constant. The remaining control variables were not significant, were not confirmed, and were removed from the model.

### 5.2.2 Supplemental Test for Mediation

The supplemental test for mediation assessed whether the total daily interactive dissemination frequency and the average daily interactive dissemination hours were *direct effects* or merely *indirect effects* on violent crime in the next period. The supplemental assessment for mediation was tested in various steps utilizing several multiple linear regressions, some of which employed violent crime count (per period) as the dependent variable (VCCOUNT). The independent variables included Total Aggression Credibility Score (per period) (TOTALCREDAGGRESS), Total Loss Credibility Score (per period) (TOTALCREDLOSS), Total Interactive Dissemination Frequency (TOTALIDF), and Average Interactive Dissemination Hours (AVGIDHOURS), measured in time (hour).

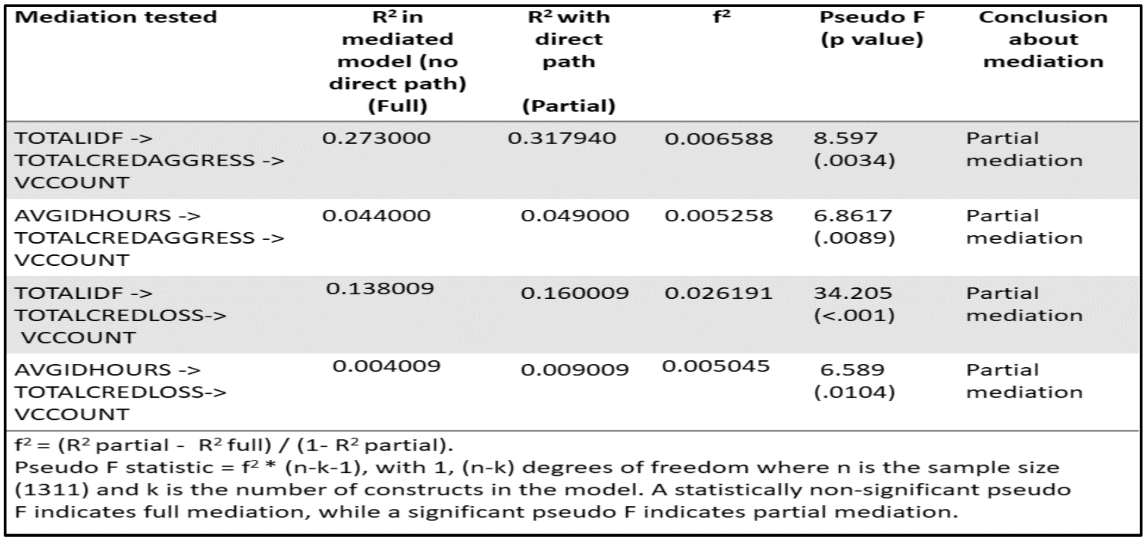
To evaluate the results for each test, $R^2$ was calculated in the mediated model (no direct path) (which represents full mediation), the $R^2$ with the direct path (which represents partial mediation), and the $f^2$ statistic, where $f^2 = (R^2 \text{ partial} - R^2 \text{ full}) / (1 - R^2 \text{ partial})$. The pseudo F statistic, where pseudo F statistic = $f^2 * (n-k-1)$ was also computed. The sample size (n) is 1,311. The degrees of freedom are 1, (n-k), where k represents the number of constructs in the model. For the paths that used TOTALCREDAGGRESS, five was the value of k (to account for the square term of aggression credibility score); however, four constructs were employed for the paths that used TOTALCREDLOSS. A statistically non-significant pseudo F indicates full mediation, while a significant pseudo F indicates partial mediation.

The conclusion from the test results was that Total Aggression Credibility Score partially mediated the relationship between Total Interactive Dissemination Frequency and Violent Crime Count and between Average Interactive Dissemination Hours and Violent Crime Count. The direct paths between Total Interactive Dissemination Frequency and Violent Crime Count and between Average Interactive Dissemination Hours and Violent Crime Count were statistically significant. It was also discovered that the Total Loss Credibility Score partially mediated the relationship between Total Interactive Dissemination Frequency and Violent Crime Count and the relationship between Average Interactive Dissemination Hours and Violent Crime Count. As a result, the hypotheses for these predictor variables (H5a, H5b, H6a, and H6b) are confirmed, as this result suggests a significant effect in the same direction. The results of this test are listed in Table 20.

A hierarchical multiple regression with the dependent variable of violent crime count was executed. The model reached significance in predicting violent crime count (per period) ($F = 66.212$, $p< .001$). The control variables explained 42.5% of the model, while 43.4% variability is explained, as a whole, by the entire model. The $R^2$ change is 0.9%;

that is, an additional 0.9% was achieved in the outcome, even when controlling for various control variables (p<.001). The MSE is 5791.85. Figure 4 shows the path analysis results.

**Table 20: Mediation Test Results**



| Mediation tested | $R^2$ in mediated model (no direct path) (Full) | $R^2$ with direct path (Partial) | $f^2$ | Pseudo F (p value) | Conclusion about mediation |
|---|---|---|---|---|---|
| TOTALIDF -> TOTALCREDAGGRESS -> VCCOUNT | 0.273000 | 0.317940 | 0.006588 | 8.597 (.0034) | Partial mediation |
| AVGIDHOURS -> TOTALCREDAGGRESS -> VCCOUNT | 0.044000 | 0.049000 | 0.005258 | 6.8617 (.0089) | Partial mediation |
| TOTALIDF -> TOTALCREDLOSS-> VCCOUNT | 0.138009 | 0.160009 | 0.026191 | 34.205 (<.001) | Partial mediation |
| AVGIDHOURS -> TOTALCREDLOSS-> VCCOUNT | 0.004009 | 0.009009 | 0.005045 | 6.589 (.0104) | Partial mediation |

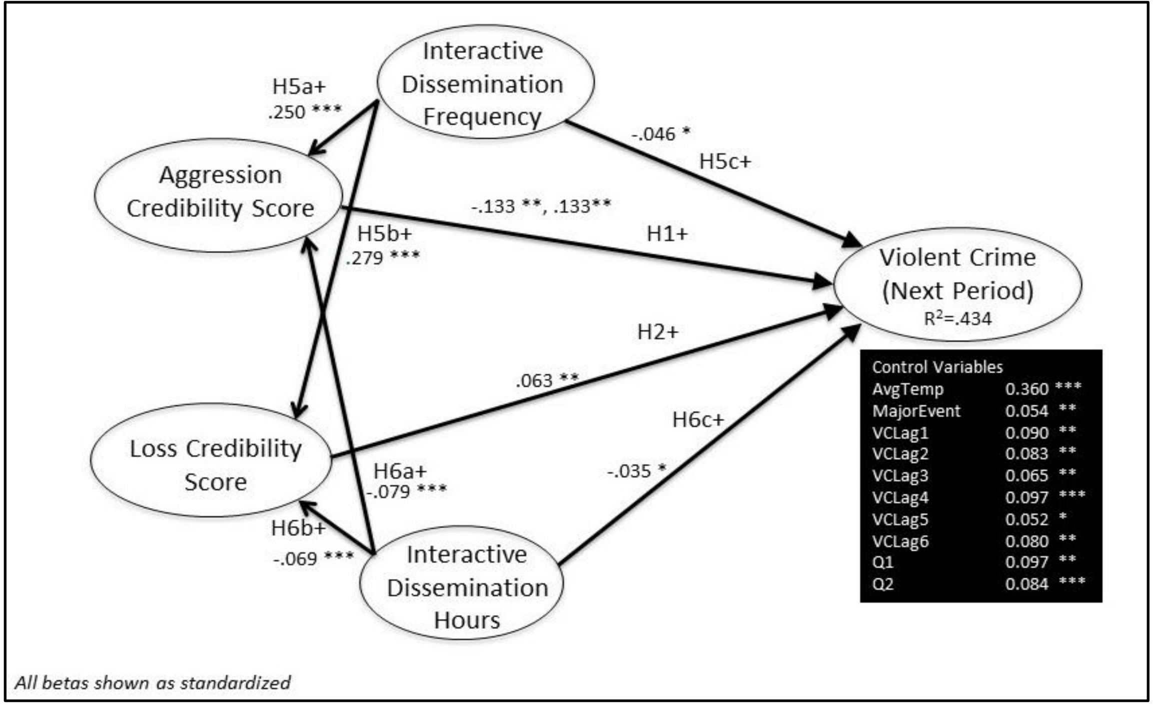
$f^2 = (R^2 \text{ partial} - R^2 \text{ full}) / (1 - R^2 \text{ partial})$.
Pseudo F statistic = $f^2 * (n-k-1)$, with 1, (n-k) degrees of freedom where n is the sample size (1311) and k is the number of constructs in the model. A statistically non-significant pseudo F indicates full mediation, while a significant pseudo F indicates partial mediation.

The study also employed another supplemental approach to the research model by incorporating a direct path from the daily sums of each of the nine credibility indicators to the outcome variable. This approach assessed whether all of the daily totals of the credibility indicators used in the Co-UGS Conceptual Model to produce the formative theoretical Credible Content Signaling construct are better as direct effects to violent crime rather than indirect effects. This model was compared with the prior competing model.

The first step in the test collected the output from the Stage 1 logistic regression for classified credible aggression. Similarly, the second step in the test collected the output from the Stage 1 logistic regression for classified credible loss. The third step in the test included aggregating the credible tweets into a credible aggression tweet frequency (per period) (TOTALCREDAGGRESSFREQ). Similarly, the count of credible loss tweets was aggregated into a credible loss tweet frequency (per period) (TOTALCREDLOSSFREQ).

The daily sums or averages of the other model variables were also aggregated. This resulted in 1,206 observations of days over the five years with aggregated frequencies of credible aggression or loss, as well as the other model variables.

**Figure 4: Structural Analysis Results for Research Model**



| Control Variables | |
|---|---|
| AvgTemp | 0.360 *** |
| MajorEvent | 0.054 ** |
| VCLag1 | 0.090 ** |
| VCLag2 | 0.083 ** |
| VCLag3 | 0.065 ** |
| VCLag4 | 0.097 *** |
| VCLag5 | 0.052 * |
| VCLag6 | 0.080 ** |
| Q1 | 0.097 ** |
| Q2 | 0.084 *** |

*All betas shown as standardized*

## 5.2.3 Supplemental Test for Direct Effects

### 5.2.3.1 Descriptive Statistics and Correlation Analysis

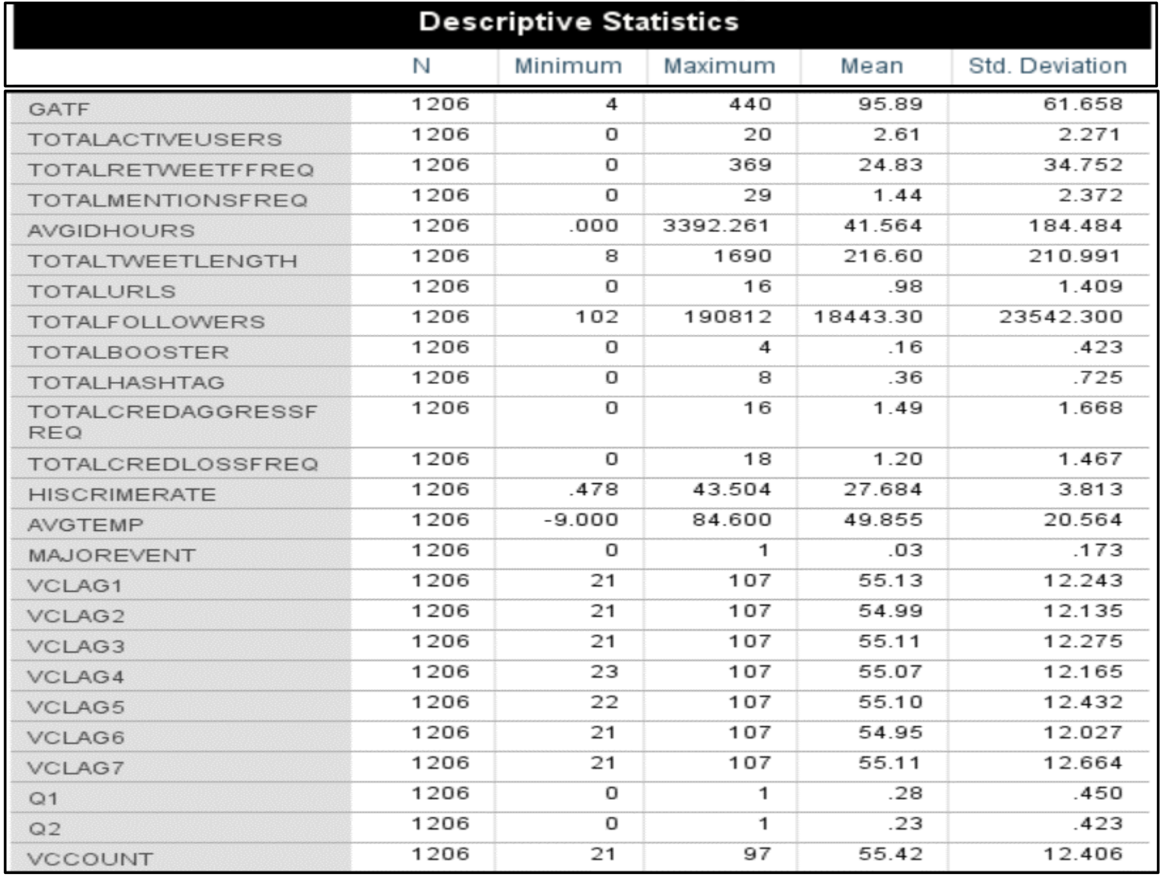
The dependent variable, violent crime count (next day), was automatically labeled with historical data sourced from the City of Chicago data portal. Table 21 shows the descriptive statistics of all variables used in the model. Violent crime counts were on average 55.42 instances (SD = 12.41). Total credible aggression frequency (per period) was 1.49 (SD = 1.67), while credible loss frequency (per period) was 1.20 (SD = 1.46). On average, the frequency of gang-affiliated tweets (per period) were (M = 95.89, SD = 61.66).

Total Retweet and Favorites Frequency (per period) was 24.83 (SD = 34.75), while Average Interactive Dissemination (AVGIDHOURS) was 41.56 (SD = 184.48).

The results of the Pearson correlation analysis revealed that there was a negative correlation between credible aggression frequency (per period) and violent crime count (per period) (r = -0.193, p < .001) and a slight positive correlation between credible loss frequency (per period) and violent crime count (per period) (r = 0.004). Gang-Affiliated Tweet Frequency is negatively associated with violent crime (r=-0.065, p<.05). Total Retweet and Favorites Frequency (TOTALRETWEETFFREQ) is negatively associated with violent crime count (r = -0.147, p < .001), while retweet time (AVGIDHOURS) is also negatively related to violent crime count (r = -0.086, p < .01) (Table 22).

All assumptions were assessed for linear regression in the research model. The model outcome variable is normally distributed, indicated by the variable histogram (Appendix E) and the Shapiro-Wilk Test of Normality. In the Shapiro-Wilk Test, one would reject the null hypothesis of normal distribution in the population if the p-value is <.05. However, because the p-value in the study is .045 (rounded up to .05), the data has not severely violated this rule. The correlation metrics for most variables were below .7 (or higher than -.7), none of the VIF values were above 10; thus, the assumption of no multicollinearity has been met with the study variables, with one exception. Total Active Users introduced multicollinearity into the model with Credible Aggression Frequency and with Total Tweet Length. Thus, Total Active Users was removed from the multiple regression model. Finally, the assumptions of linearity and homoscedasticity were met, as the scatterplots of standardized residual on standardized predicted value did not curve or exhibit an undesirable pattern in most variables (one exception noted below).

**Table 21: Direct Effects Model Descriptive Statistics**



Descriptive Statistics

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| GATF | 1206 | 4 | 440 | 95.89 | 61.658 |
| TOTALACTIVEUSERS | 1206 | 0 | 20 | 2.61 | 2.271 |
| TOTALRETWEETFFREQ | 1206 | 0 | 369 | 24.83 | 34.752 |
| TOTALMENTIONSFREQ | 1206 | 0 | 29 | 1.44 | 2.372 |
| AVGIDHOURS | 1206 | .000 | 3392.261 | 41.564 | 184.484 |
| TOTALTWEETLENGTH | 1206 | 8 | 1690 | 216.60 | 210.991 |
| TOTALURLS | 1206 | 0 | 16 | .98 | 1.409 |
| TOTALFOLLOWERS | 1206 | 102 | 190812 | 18443.30 | 23542.300 |
| TOTALBOOSTER | 1206 | 0 | 4 | .16 | .423 |
| TOTALHASHTAG | 1206 | 0 | 8 | .36 | .725 |
| TOTALCREDAGGRESSFREQ | 1206 | 0 | 16 | 1.49 | 1.668 |
| TOTALCREDLOSSFREQ | 1206 | 0 | 18 | 1.20 | 1.467 |
| HISCRIMERATE | 1206 | .478 | 43.504 | 27.684 | 3.813 |
| AVGTEMP | 1206 | -9.000 | 84.600 | 49.855 | 20.564 |
| MAJOREVENT | 1206 | 0 | 1 | .03 | .173 |
| VCLAG1 | 1206 | 21 | 107 | 55.13 | 12.243 |
| VCLAG2 | 1206 | 21 | 107 | 54.99 | 12.135 |
| VCLAG3 | 1206 | 21 | 107 | 55.11 | 12.275 |
| VCLAG4 | 1206 | 23 | 107 | 55.07 | 12.165 |
| VCLAG5 | 1206 | 22 | 107 | 55.10 | 12.432 |
| VCLAG6 | 1206 | 21 | 107 | 54.95 | 12.027 |
| VCLAG7 | 1206 | 21 | 107 | 55.11 | 12.664 |
| Q1 | 1206 | 0 | 1 | .28 | .450 |
| Q2 | 1206 | 0 | 1 | .23 | .423 |
| VCCOUNT | 1206 | 21 | 97 | 55.42 | 12.406 |

### 5.2.3.2 Simple Linear Regression Analyses

The fourth step in the test was executed via several Simple Linear Regression models that used violent crime count (per period) as the dependent variable. Each regression algorithm employed the daily total of one of the nine credibility variables (instead of a possible 18 contemplated in an original design of this study; that is, nine for aggression and nine for loss to remove the risk of introducing multicollinearity into the final model). All results are reported here with standardized beta coefficients and p-values.

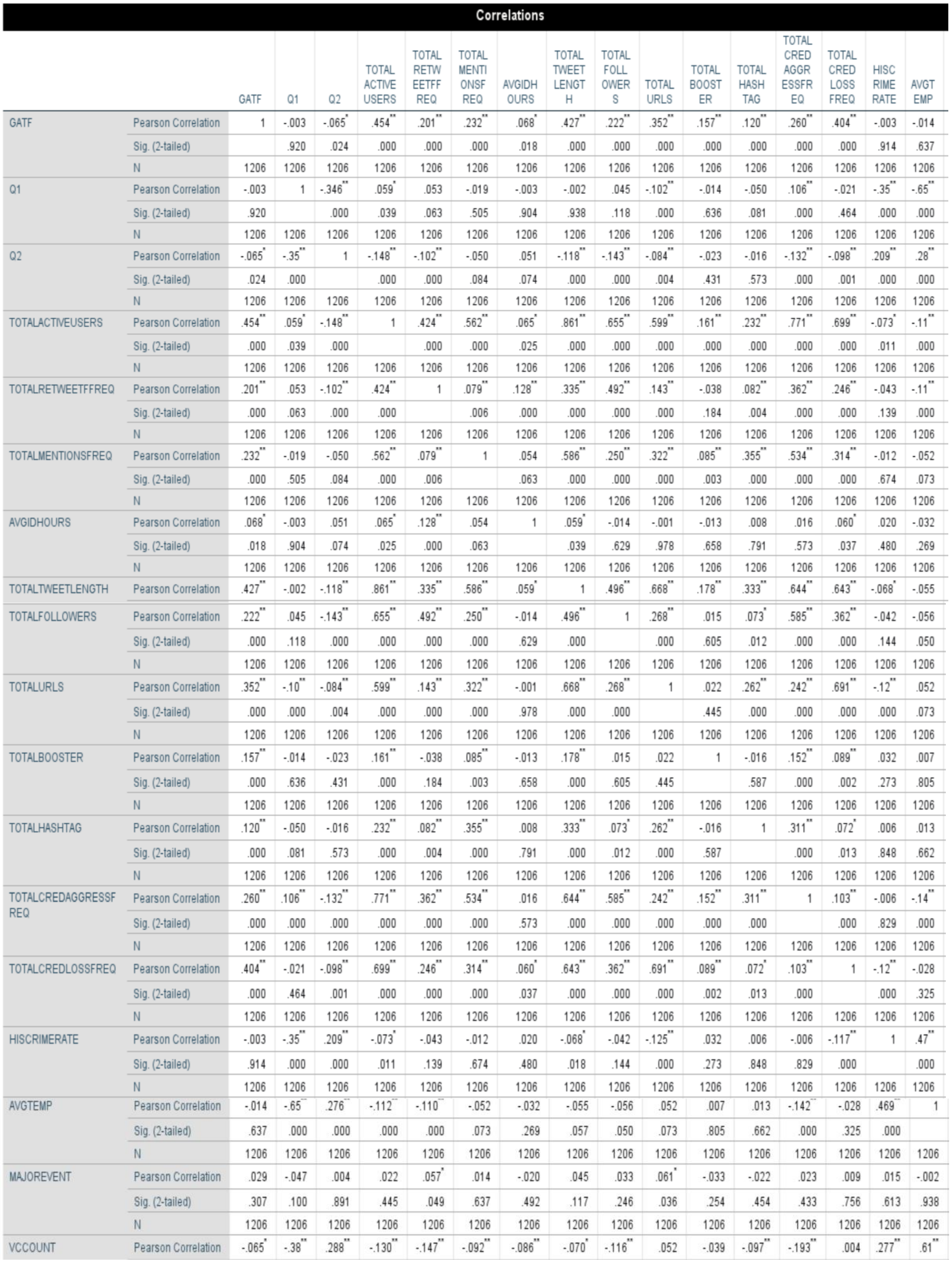
The results of the first simple linear regression analysis showed that Total Aggression Credibility Frequency (H1) has a significant link to Violent Crime Count (next day) (β = -0.193, t = -6.841, p < .001), which makes it a significant predictor of violent crime count; once total credible aggression count increases, violent crime count decreases.

This association explains variance ($R^2$) of 3.7% in the model. As a result, this model variable is confirmed, but in the opposite direction as indicated on the model. This result suggests a significant effect in the opposite direction as Violent Crime Count (next day).

The results of the second simple linear regression analysis showed that Total Loss Credibility Frequency (H2) does not have a significant link to violent crime count (next day) ($\beta = 0.003$, $t = .112$, $p = .878$), which does not make it a significant predictor of violent crime count. As a result, this variable is not confirmed, and this result suggests a non-significant effect for credible loss frequency in the same direction as violent crime count.

The variable Total Gang-Affiliated Tweet Frequency (H3) has a significant link to violent crime count (next day), which makes it a significant predictor of violent crime count. This relationship is also curvilinear, with an initial negative relationship between gang-affiliated tweet frequency and violent crime (next day) ($\beta = -0.266$, $t = -4.834$, $p < .001$). However, when the gang-affiliated tweet frequency exceeds approximately 225 per day, violent crime starts to shift in a positive direction ($\beta = 0.249$, $t = 5.332$, $p < .001$) the next day. This initial descent with a reversal in direction explains variance ($R^2$) of 2.4% in the model. As a result, this model variable is confirmed, and the result suggests a significant effect first in the opposite direction, and then in the same direction as violent crime count (next day). The variable Total Tweet Length (the sum of the daily length of credible tweets, in characters) (H4), has an initial significant negative link to violent crime count (next day) ($\beta = -.256$, $t = -3.779$, $p < .001$), which makes it a significant predictor of violent crime count. However, the association is curvilinear, such that when the sum of the daily tweet lengths exceeds approximately 900 per day in these data, violent crime (next day) starts to shift (curve) in a positive direction ($\beta = 0.205$, $t = 3.031$, $p < .05$). This initial descent with a later reversal in direction explains variance ($R^2$) of 1.2% in the model. As a result, this model variable is confirmed, and the result suggests a significant effect first in the opposite direction, and then in the same direction as violent crime count (next day). The variable Total Retweet and Favorites Frequency (the sum of retweet frequency and favorites count for all credible tweets) (H5) has a significant negative link to violent crime count (next day) ($\beta = -0.147$, $t = -5.163$, $p < .001$), which makes it a significant predictor of violent crime count. That is, as the total daily count of retweet and favorites frequency increases, violent crime count decreases. This association explains variance ($R^2$) of 2.2% in the model. As a result, this model variable is confirmed, suggesting that the total retweet and favorites frequency affects violent crime count (next day) but not in the theorized direction. The variable Average Interactive Dissemination Hours (H6), the average of daily retweet time, in hours), has a significant negative link to violent crime count (next day) ($\beta = -0.101$, $t = -3.509$, $p < .001$), which makes it a significant predictor of violent crime count. As a result, this model variable is confirmed. Translating to the model variable Average Interactive Dissemination Speed, this suggests that as retweet time decreases, speed increases. Thus, the significant result suggests that as average interactive dissemination speed increases, violent crime count (next day) also increases. It explains variance ($R^2$) of 1% in the model. As a result, this model variable is confirmed, as the result suggests a significant effect in the same direction as violent crime count (next day). The variable Total Tweets with URL (H7) shows a significant positive link to violent crime count (next day) ($\beta = 0.052$, $t = 1.799$, $p < .1$), which makes it a significant predictor of violent crime count. This association explains variance ($R^2$) of .3% in the model. As a result, this model variable is confirmed.

## Table 22: Direct Effects Model Correlation Matrix



Correlations

| | | GATF | Q1 | Q2 | TOTAL ACTIVE USERS | TOTAL RETWEETFFREQ | TOTAL MENTIONSFREQ | AVGIDHOURS | TOTAL TWEETLENGTH | TOTAL FOLLOWERS | TOTAL URLS | TOTAL BOOSTER | TOTAL HASHTAG | TOTAL CRED AGGRESSFREQ | TOTAL CRED LOSS FREQ | HISCRIMERATE | AVGTEMP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GATF | Pearson Correlation | 1 | -.003 | -.065* | .454** | .201** | .232** | .068* | .427** | .222** | .352** | .157** | .120** | .260** | .404** | -.003 | -.014 |
| | Sig. (2-tailed) | | .920 | .024 | .000 | .000 | .000 | .018 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .914 | .637 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| Q1 | Pearson Correlation | -.003 | 1 | -.346** | .059* | .053 | -.019 | -.003 | -.002 | .045 | -.102** | -.014 | -.050 | .106** | -.021 | -.35** | -.65** |
| | Sig. (2-tailed) | .920 | | .000 | .039 | .063 | .505 | .904 | .938 | .118 | .000 | .636 | .081 | .000 | .464 | .000 | .000 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| Q2 | Pearson Correlation | -.065* | -.35** | 1 | -.148** | -.102** | -.050 | .051 | -.118** | -.143** | -.084** | -.023 | -.016 | -.132** | -.098** | .209** | .28** |
| | Sig. (2-tailed) | .024 | .000 | | .000 | .000 | .084 | .074 | .000 | .000 | .004 | .431 | .573 | .000 | .001 | .000 | .000 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALACTIVEUSERS | Pearson Correlation | .454** | .059* | -.148** | 1 | .424** | .562** | .065* | .861** | .655** | .599** | .161** | .232** | .771** | .699** | -.073* | -.11** |
| | Sig. (2-tailed) | .000 | .039 | .000 | | .000 | .000 | .025 | .000 | .000 | .000 | .000 | .000 | .000 | .000 | .011 | .000 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALRETWEETFFREQ | Pearson Correlation | .201** | .053 | -.102** | .424** | 1 | .079** | .128** | .335** | .492** | .143** | -.038 | .082** | .362** | .246** | -.043 | -.11** |
| | Sig. (2-tailed) | .000 | .063 | .000 | .000 | | .006 | .000 | .000 | .000 | .000 | .184 | .004 | .000 | .000 | .139 | .000 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALMENTIONSFREQ | Pearson Correlation | .232** | -.019 | -.050 | .562** | .079** | 1 | .054 | .586** | .250** | .322** | .085** | .355** | .534** | .314** | -.012 | -.052 |
| | Sig. (2-tailed) | .000 | .505 | .084 | .000 | .006 | | .063 | .000 | .000 | .000 | .003 | .000 | .000 | .000 | .674 | .073 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| AVGIDHOURS | Pearson Correlation | .068* | -.003 | .051 | .065* | .128** | .054 | 1 | .059* | -.014 | -.001 | -.013 | .008 | .016 | .060* | .020 | -.032 |
| | Sig. (2-tailed) | .018 | .904 | .074 | .025 | .000 | .063 | | .039 | .629 | .978 | .658 | .791 | .573 | .037 | .480 | .269 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALTWEETLENGTH | Pearson Correlation | .427** | -.002 | -.118** | .861** | .335** | .586** | .059* | 1 | .496** | .668** | .178** | .333** | .644** | .643** | -.068* | -.055 |
| TOTALFOLLOWERS | Pearson Correlation | .222** | .045 | -.143** | .655** | .492** | .250** | -.014 | .496** | 1 | .268** | .015 | .073* | .585** | .362** | -.042 | -.056 |
| | Sig. (2-tailed) | .000 | .118 | .000 | .000 | .000 | .000 | .629 | .000 | | .000 | .605 | .012 | .000 | .000 | .144 | .050 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALURLS | Pearson Correlation | .352** | -.10** | -.084** | .599** | .143** | .322** | -.001 | .668** | .268** | 1 | .022 | .262** | .242** | .691** | -.12** | .052 |
| | Sig. (2-tailed) | .000 | .000 | .004 | .000 | .000 | .000 | .978 | .000 | .000 | | .445 | .000 | .000 | .000 | .000 | .073 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALBOOSTER | Pearson Correlation | .157** | -.014 | -.023 | .161** | -.038 | .085** | -.013 | .178** | .015 | .022 | 1 | -.016 | .152** | .089** | .032 | .007 |
| | Sig. (2-tailed) | .000 | .636 | .431 | .000 | .184 | .003 | .658 | .000 | .605 | .445 | | .587 | .000 | .002 | .273 | .805 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALHASHTAG | Pearson Correlation | .120** | -.050 | -.016 | .232** | .082** | .355** | .008 | .333** | .073* | .262** | -.016 | 1 | .311** | .072* | .006 | .013 |
| | Sig. (2-tailed) | .000 | .081 | .573 | .000 | .004 | .000 | .791 | .000 | .012 | .000 | .587 | | .000 | .013 | .848 | .662 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALCREDAGGRESSFREQ | Pearson Correlation | .260** | .106** | -.132** | .771** | .362** | .534** | .016 | .644** | .585** | .242** | .152** | .311** | 1 | .103** | -.006 | -.14** |
| | Sig. (2-tailed) | .000 | .000 | .000 | .000 | .000 | .000 | .573 | .000 | .000 | .000 | .000 | .000 | | .000 | .829 | .000 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| TOTALCREDLOSSFREQ | Pearson Correlation | .404** | -.021 | -.098** | .699** | .246** | .314** | .060* | .643** | .362** | .691** | .089** | .072* | .103** | 1 | -.12** | -.028 |
| | Sig. (2-tailed) | .000 | .464 | .001 | .000 | .000 | .000 | .037 | .000 | .000 | .000 | .002 | .013 | .000 | | .000 | .325 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| HISCRIMERATE | Pearson Correlation | -.003 | -.35** | .209** | -.073* | -.043 | -.012 | .020 | -.068* | -.042 | -.125** | .032 | .006 | -.006 | -.117** | 1 | .47** |
| | Sig. (2-tailed) | .914 | .000 | .000 | .011 | .139 | .674 | .480 | .018 | .144 | .000 | .273 | .848 | .829 | .000 | | .000 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| AVGTEMP | Pearson Correlation | -.014 | -.65** | .276** | -.112** | -.110** | -.052 | -.032 | -.055 | -.056 | .052 | .007 | .013 | -.142** | -.028 | .469** | 1 |
| | Sig. (2-tailed) | .637 | .000 | .000 | .000 | .000 | .073 | .269 | .057 | .050 | .073 | .805 | .662 | .000 | .325 | .000 | |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| MAJOREVENT | Pearson Correlation | .029 | -.047 | .004 | .022 | .057* | .014 | -.020 | .045 | .033 | .061* | -.033 | -.022 | .023 | .009 | .015 | -.002 |
| | Sig. (2-tailed) | .307 | .100 | .891 | .445 | .049 | .637 | .492 | .117 | .246 | .036 | .254 | .454 | .433 | .756 | .613 | .938 |
| | N | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 | 1206 |
| VCCOUNT | Pearson Correlation | -.065* | -.38** | .288** | -.130** | -.147** | -.092** | -.086** | -.070* | -.116** | .052 | -.039 | -.097** | -.193** | .004 | .277** | .61** |

The variable Total Tweets with Booster (H8) has a non-significant negative link to violent crime count (next day) (β = -0.039, t = -1.337, p = .181), which does not make it a significant predictor of violent crime count. As a result, this model variable is not confirmed, but the result suggests a non-significant effect in the opposite direction as violent crime count (next day). The variable Total Tweets with Hashtag (H9) has a significant negative link to violent crime count (next day) (β = -0.097, t = -3.39, $p < .001$), which makes it a significant predictor of violent crime count; that is, as the total daily count of tweets containing a hashtag increases, violent crime count decreases. This association explains variance ($R^2$) of 1% in the model. As a result, this model variable is confirmed; however, the result suggests a significant effect in the opposite direction as violent crime count (next day).

The variable Total User Mentions (the sum of daily tweets that contains a mention) (H10) has a significant negative link to violent crime count (next day) (β = -0.092, t = -3.222, $p < .01$), which makes it a significant predictor of violent crime count; that is, as the total daily count of tweets with mentions increases, violent crime count decreases. This association explains variance ($R^2$) of 0.9% in the model. As a result, this model variable is confirmed; however, the result suggests a significant effect in the opposite direction as violent crime count (next day). The variable Total Followers (H11) has a significant negative link to violent crime count (next day), which makes it a significant predictor of violent crime count (next day) (β = -0.116, t = -4.039, $p < .001$). This explains variance ($R^2$) of 1.3% in the model. As a result, this model variable is confirmed; however, the result suggests a significant effect in the opposite direction as violent crime count (next day). Interestingly, this empirical finding affirms the subjective statement by one of the study's domain experts that the more Twitter followers a gang-affiliated user has, the more the negative perception by other gang members, and thus, the less the gang-affiliated user tends to post on Twitter.

The variable Total Active Users (H12) has a significant link to violent crime count (next day) ($\beta$ =-.264, t =-6.116, p <.001). However, the association is curvilinear, such that when the total count of active Twitter gang-affiliated users exceeds approximately 10 per day in these data, violent crime (next day) starts to shift (curve) in a positive direction ($\beta$ = 0.18, t = 4.1647, p < .001). This initial descent with a later reversal in direction explains variance ($R^2$) of 3.1% in the model. As a result, this model variable is confirmed, and the result suggests a significant effect first in the opposite direction, and then in the same direction as violent crime count (next day).

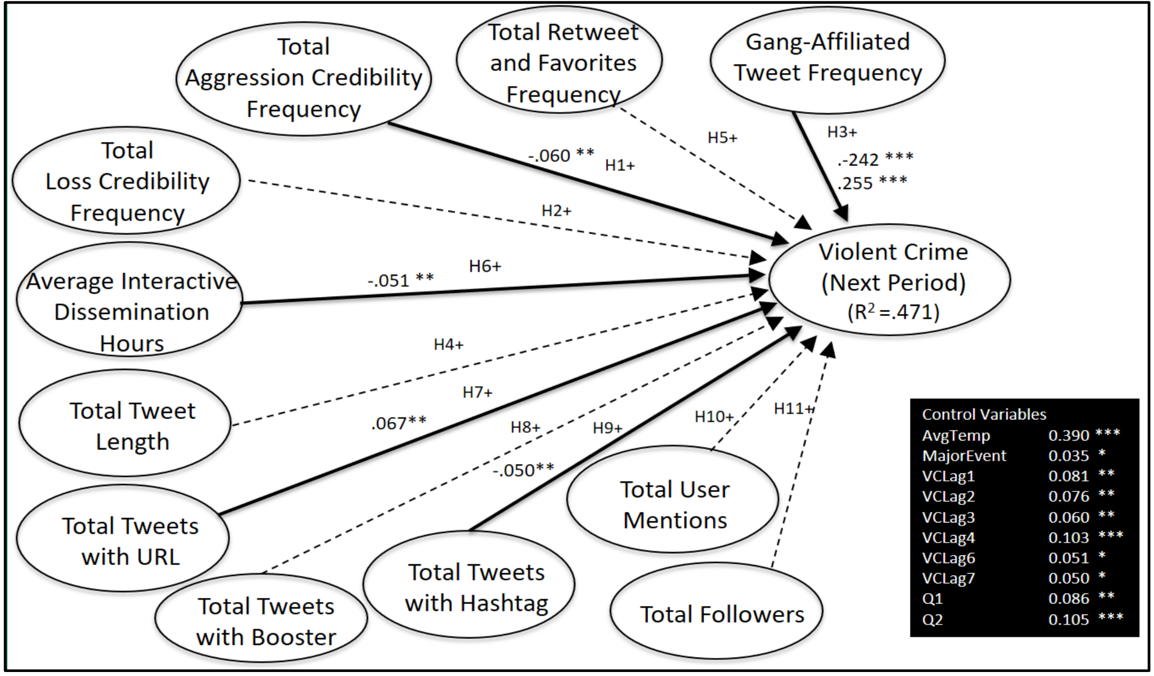
#### 5.2.3.3 Hierarchical Multiple Regression Analysis

In the Direct Effects Model, all but one of the variables used in the simple linear regressions (whether initially confirmed or not) were combined into a multiple regression to further assess them. The variable Total Active Users introduced multicollinearity into the model (with Total Tweet Length and Total Aggression Credibility Frequency), so it was removed. Further, the model controlled for several confounding variables and their impact on the constructs. This test was executed via a Hierarchical Multiple Regression that used Violent Crime Count (next period) as the dependent variable, and average temperature (F), the historical crime rate per 100,000, weekday, average daily tweet time (hour), a major event on tweet day, and prior seven periods' violent crime counts (including seasonality indicator variables) as the control variables. Figure 5 shows the path analysis results.

From the independent variables, Gang-Affiliated Tweet Frequency (GATF) (and its square), Total Aggression Credibility Frequency (TOTALCREDAGGESSFREQ), Total Tweets with Hashtag (TOTALHASHTAGS), Total Tweets with URL (TOTALURLS), and Average Interactive Dissemination Hours (AVGIDHOURS) were significant predictors of Violent Crime Count. Total Loss Credibility Frequency, Total Tweets with Booster, Total Followers, Total Tweet Length, the square of Total Tweet Length (similar to the simple linear regression), Total User Mentions, and Total Retweet and Favorites Frequency were not significant predictors of Violent Crime Count. The following control variables were not significant: Hour, Third Quarter Seasonality variable (Q3), Historical Crime Rate, Weekday, and one of the prior period violent crime count lag variables, VCLAG5.

Several useful independent variables are reported with standardized beta coefficients and significance. These include Total Tweets with URL ($\beta = 0.067$, $t = 2.782$, $p < .01$), Gang-Affiliated Tweet Frequency ($\beta = -0.242$, $t = -4.262$, $p < .001$) with its square term ($\beta = 0.255$, $t = 4.563$, $p < .001$), and Average Interactive Dissemination Hours ($\beta = -0.051$, $t = -2.405$, $p < .001$). This is interpreted as the violent crime being higher in the following situations: when the daily gang-affiliated tweet frequency is higher than 250, when there are more tweets with URLs, and when the average daily retweet speed is higher. Together, these independent variables account for 2.2% ($R^2$) of the variance in violent crime counts (next day).

**Figure 5: Regression Path Results for Direct Effects Model**

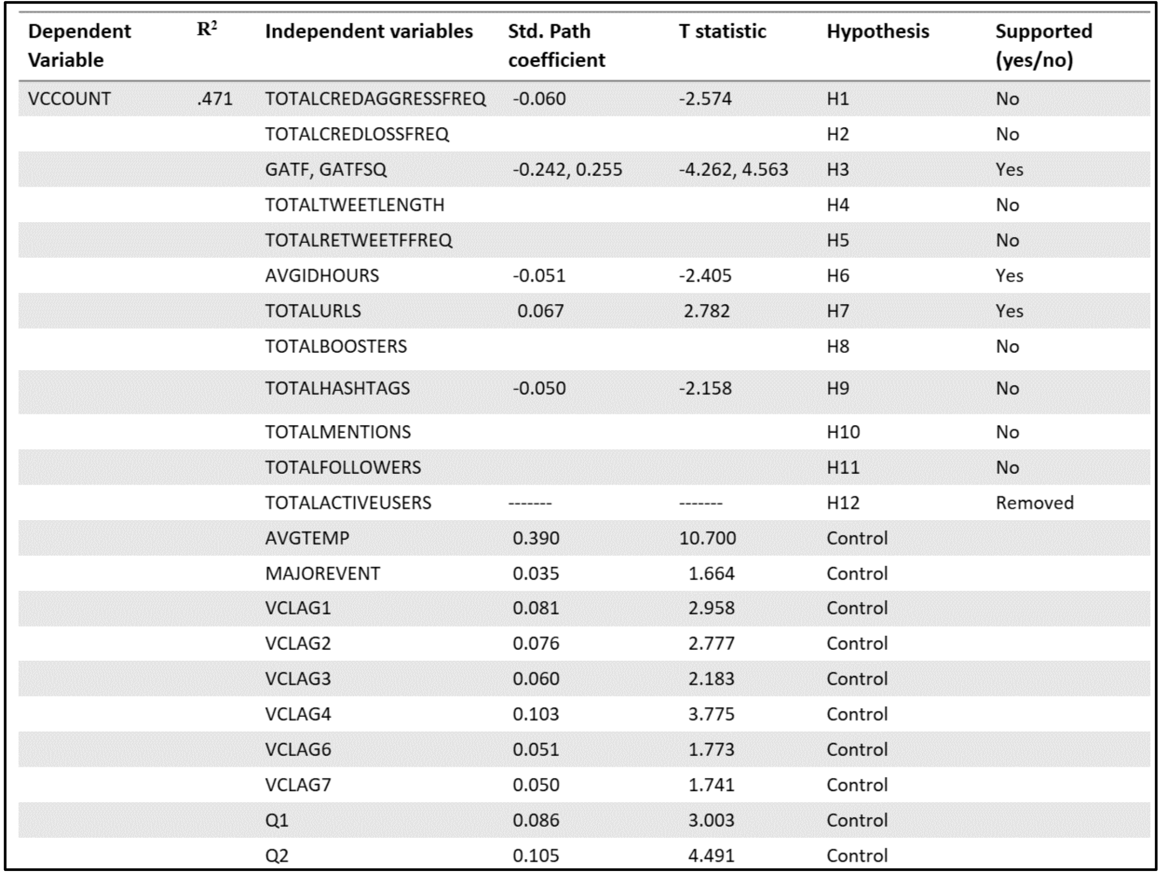


All significant control variables with their standardized beta coefficients and significance level are also reported. These include First Quarter Seasonality (Q1) (β = 0.086, t = 3.003, p < .01), Second Quarter Seasonality (Q2) (β = 0.105, t = 4.491, p = .001), Average Daily Temperature (β = .390, t = 10.700, p < .001), and Major Event (β = .035, t = 1.664, p < 0.10). Additional significant control variables include Violent Crime Count One Day Prior (VCLAG1) (β = .081, t = 2.958, p < .01), Violent Crime Count Two Days Prior (VCLAG2) (β = .076, t = 2.777, p < .01), Violent Crime Count Three Days Prior (VCLAG3) (β = .060, t = 2.183, p < .05), Violent Crime Count Four Days Prior (VCLAG4) (β = .103, t = 3.775, p < .001), Violent Crime Count Six Days Prior (VCLAG6) (β = .051, t = 1.773, p<0.10), and Violent Crime Count Seven Days Prior (VCLAG7) (β = .050, t = 1.741, p<0.10). This is interpreted as violent crime being higher in the following situations: when temperature is higher, if there is a major event on the day, in quarters one and two (with more occurrences in quarter two than quarter one), and in the fourth, first, second, third, sixth, and seventh days (ranked in order) leading up the tweet day. The following control variables were not significant: Tweet Hour, Third Quarter Seasonality variable (Q3), Historical Daily Crime Rate, Weekday, and VCLAG5.

Together, six of the variables have a significant link with violent crime count (per period), with three of the six (includes GATF's square) positively associated. Three of the six variables are negatively associated, such that as each of them increases, violent crime count decreases. Adding the independent variables to the hierarchical regression model with the significant control variables added to the efficacy of the model. The coefficient betas and significance for all variables are listed in Table 23 with all regression details shown in Appendix D. As a whole, the normalized regression model was significant (F(16,1189)=66.098, p<.000), with an $R^2$ of .471 explained by the independent and control variables. As a result, the Direct Effects Model is confirmed, and Hypotheses H3, H6, and H7 are confirmed, as this result suggests a significant effect in the same direction. Hypotheses H1, H2, H4, H5, H8, H9, H10, and H11 are not confirmed. Hypothesis 12 was removed due to multicollinearity.

***Feature Selection via Stepwise Regression.*** As an additional step, the study employed a feature analysis and stepwise regression on the predictors to determine their ranked importance in the model. From the significant control variables, the algorithm returned the following (ranked) results: average daily temperature ($R^2$ Change = .370), violent crime count lag 4 ($R^2$ Change = .028), violent crime count lag 1 ($R^2$ Change = .016), violent crime count lag 2 ($R^2$ Change = .011), violent crime count lag 7 ($R^2$ Change = .008), Q2 ($R^2$ Change = .006), violent crime count lag 3 ($R^2$ Change = .003), Q1 ($R^2$ Change = .003), violent crime count lag 6 ($R^2$ Change = .002), major event ($R^2$ Change = .002). From the significant predictor variables, the algorithm selected Credible Aggression Tweet Frequency ($R^2$ Change = .004), Total Tweets with URLs ($R^2$ Change = .004), Retweet Hour ($R^2$ Change = .003), and Total Tweets with Hashtags ($R^2$ Change = .003). Overall, this resulted in an $R^2$ = 46.3%. The beta values were similar to those reported earlier in the paper for the Direct Effects Model.

**Table 23: Direct Effects Model Regression Results**



| Dependent Variable | $R^2$ | Independent variables | Std. Path coefficient | T statistic | Hypothesis | Supported (yes/no) |
|---|---|---|---|---|---|---|
| VCCOUNT | .471 | TOTALCREDAGGRESSFREQ | -0.060 | -2.574 | H1 | No |
| | | TOTALCREDLOSSFREQ | | | H2 | No |
| | | GATF, GATFSQ | -0.242, 0.255 | -4.262, 4.563 | H3 | Yes |
| | | TOTALTWEETLENGTH | | | H4 | No |
| | | TOTALRETWEETFFREQ | | | H5 | No |
| | | AVGIDHOURS | -0.051 | -2.405 | H6 | Yes |
| | | TOTALURLS | 0.067 | 2.782 | H7 | Yes |
| | | TOTALBOOSTERS | | | H8 | No |
| | | TOTALHASHTAGS | -0.050 | -2.158 | H9 | No |
| | | TOTALMENTIONS | | | H10 | No |
| | | TOTALFOLLOWERS | | | H11 | No |
| | | TOTALACTIVEUSERS | ------- | ------- | H12 | Removed |
| | | AVGTEMP | 0.390 | 10.700 | Control | |
| | | MAJOREVENT | 0.035 | 1.664 | Control | |
| | | VCLAG1 | 0.081 | 2.958 | Control | |
| | | VCLAG2 | 0.076 | 2.777 | Control | |
| | | VCLAG3 | 0.060 | 2.183 | Control | |
| | | VCLAG4 | 0.103 | 3.775 | Control | |
| | | VCLAG6 | 0.051 | 1.773 | Control | |
| | | VCLAG7 | 0.050 | 1.741 | Control | |
| | | Q1 | 0.086 | 3.003 | Control | |
| | | Q2 | 0.105 | 4.491 | Control | |

The stepwise regression feature selection process removed the GATF variable (and its square, GATFSQ, which accounted for the curved relationship). Though significant, the feature selection algorithm did not consider the $R^2$ contribution from each variable high enough to add to the efficacy of the model. However, both variables were kept in the Direct Effects model due to their significance and the need to test the hypothesis to the dependent variable. The hypothesis was affirmed and resulted in an eventual positive association with violent crime count.

### 5.2.4 Model Comparison

The results of the Direct Effects Model were then compared and contrasted to the research model (with mediation) to determine the best model. In several ways, the models are similar. Both are significant overall. From a theoretical perspective, both models show a positive and significant connection between interactive dissemination speed and violent crime (agreeing with the conceptual model), though the beta weight is slightly higher in the Direct Effects Model.

Though interactive dissemination frequency is positively and significantly associated with tweet credibility, both models show the surprisingly negative relationship between interactive dissemination frequency and violent crime. Perhaps this result is best explained in the fact that more interactive dissemination (engagement) of perceived credible social media posts of aggression or loss does not necessarily equate to a violent crime increase due to the possibility that this may be the result of a concerted effort to provoke collective cognition, awareness, or understanding (Mitra, Wright, and Gilbert 2017). This hypothesis was initially based on Network Embeddedness Theory, which explains the sharing of content (interactions) because the desired increase of gang embeddedness drives content dissemination via interactions, which, in turn, increases the cost of not following through with the threat stated in the post. Past research has shown that retweet frequency is a good indicator for information sharing (Bild et al. 2015), for content sentiment going viral (Stieglitz and Dang-Xuan 2013), and for gang network psychological embeddedness and social capital due to gang member affirmation. The study only confirmed the relationship between interactive dissemination frequency and tweet credibility; it did not demonstrate that dissemination frequency ultimately affects an increase in violent crime in the theorized direction, though it was significant in the opposite direction.

However, there are notable differences between the two models used in the study. Firstly, the research model empirically aligns better with the theoretical assertions in several ways. Its results show a positive and significant association of credible loss with violent crime, whereas, in the Direct Effects model, the association between credible loss and violent crime was not significant. Though the direction affirms the theoretical assertion, the data points were too few in the sample, reflecting the low distribution in real life.

Secondly, the Direct Effects Model's curvilinear relationship between credible aggression frequency and violent crime was not significant, and, thus, was not observed or evidenced in this model. That is, the preprocessing algorithms for the Direct Effects Model do not use a scale for credibility. As a result, the model does not discriminate enough with its lack of granularity in the credible frequency metrics, created by the classification algorithm's (1/0) approach. Thirdly, in the research model, the Gang-Affiliated Tweet Frequency construct introduced multicollinearity into the model; thus, it was removed. The Direct Effects Model suggests an initial negative association between GATF and Violent Crime Count, curving positive at a specific point, and then trending upward.

Fourthly, the research model affirms the theoretical assertion that all nine credibility indicators (without outlier removal) are significantly and positively associated with both credible aggression and credible loss, whereas the Direct Effects Model only affirmed seven of the nine for both aggression and loss.

Fifthly, another distinction between the two models lies in their efficacy metrics. The Direct Effects Model exhibits a smaller MSE (an indication of the amount of error in the model), while the research model demonstrates a slighter higher F Statistic. The Direct Effects Model showed slight superior predictive efficacy with its higher $R^2$ (Table 24), though this is perhaps due to the number of additional variables in the model.

**Table 24: Comparison of Research Models**

| **Algorithm Type** | **$R^2$** | **F-Score** | **Sig.** | **MSE** |
|---|---|---|---|---|
| Research Model | .434 | 66.212 | .000 | 5791.85 |
| Direct Effects Model | .471 | 66.113 | .000 | 5457.56 |

The research model has a foundation build on all nine credibility indicators (affirming the Co-UGS model). It also empirically affirms the connection between credible aggression and credible loss to violent crime, shows the partial mediation between the model constructs, and exhibits a slightly higher F Statistic. The difference in the $R^2$ between the models is not significant enough to warrant the adoption of the alternative model. Thus, for these reasons, the adoption of the research model (with partial mediation) is suggested.

# CHAPTER 6: DISCUSSION, FUTURE CONSIDERATIONS, AND CONCLUSION

This study examined the task of extracting text- and emoji-based features representing loss and aggression from Twitter tweets of street gang affiliates in a large urban U.S. city to determine whether the content of these tweets are credible expressions of aggression and loss that have an impact on violent crime. Secondly, it determined whether the characteristics of interactive dissemination behavior on social media, including the frequency and speed of retweets, are effective predictors of violent crime in a city. In addition to answering these two main research questions, other theoretical goals of the research were to determine the combination of predictors of social media content credibility from individuals in a gang-affiliated community and to explore credible predictors of violent crime from social media unstructured data, time-series data, and other structured sources. The practical goal of the research was to assist current manual interventions of gang-related violent crime and to understand how to mitigate gang-related urban violent crime going forward.

## 6.1 Discussion of Results

In this section, the meaning of the study findings are discussed. The study asked two research questions. The first question asked whether credible taunts and threats (aggression) and expressions of loss in social media posts by street gang members were effective predictors of violent crime. A correlation was found between credible gang-affiliated tweet content expressions of aggression and retaliatory loss and violent crime count (per period). The Direct Effects Model results confirmed an eventual positive association between credible aggression and loss tweet frequency by gang-affiliated individuals and violent crime and credible aggression tweet frequency on upcoming violent crime count.

The second research question in the study asked whether the characteristics of interactive dissemination behavior of street gangs on social media (e.g., frequency and speed of retweets) were effective predictors of violent crime. In short, the results from the research model demonstrated an affirmative answer to the question of interactive dissemination speed. However, the results also suggested a significant negative association between interactive dissemination frequency and violent crime. Additionally, though not primary in the study's investigation, the data confirmed other control factors that positively correlate to the violent crime. These include average daily temperature, when a major event occurred on the day, the violent crime counts of the prior six days, and the first and second quarters of the year.

### 6.1.1 Interpretation of Findings on Credibility

In this section, the findings in the study are interpreted. The empirical results from Stage 1 of the study significantly confirmed all but two of the theoretical assertions made in the Co-UGS model. The results strengthened the claim of (Cha et al. 2010; O'Donovan et al. 2012; Yang and Counts 2010) that larger counts of retweets (a measure of the value of the tweet content) and user mentions (the value of the tweeter's name) lend credibility to the tweet content. In the study, user mentions had a positive impact on the credibility of gang-affiliated tweets, such that, with each additional user mention, the credibility of the tweet increased by a factor of 1.7 standard deviation times in aggressive tweets and 1.9 in loss-filled tweets. The combined frequency of retweets and favorites also had a positive impact on the credibility of tweets, such that, with each additional retweet or favorite (like) of the tweet, the tweet’s credibility increased by a factor of 17.1 standard deviation times in aggressive tweets. These results affirm the theoretical findings of (Lee and Ma 2012), suggesting that the expanded reach of information sharing, via interactive dissemination, increases the reputational cost to the Twitter user and thus positively affects the tweet’s credibility.

The study’s theoretical and empirical results built on existing evidence that aggressive tweets with booster words (Mitra, Wright, and Gilbert 2017) and a URL (Gupta, Lamba, and Kumaraguru 2013; O’Donovan et al. 2012) are considered more credible. For tweets with aggressive content, the presence of a booster word increases the credibility by a factor of 1.5 standard deviations, while increasing the credibility by 1.3 in loss-filled tweets. For tweets with aggressive content, the presence of a URL increases the credibility by a factor of 1.2 standard deviations, while increasing the credibility by 1.6 (standard deviation) in loss-filled tweets. For tweets with aggressive content, the presence of a hashtag increased the credibility by a factor of 1.72 standard deviations, while it increased the credibility by 1.96 standard deviations in loss-filled tweets.

The study empirically confirmed the Co-UGS conceptual model’s assertion that characteristics of Twitter users themselves lend credibility to tweets. Users with more followers author more credible aggressive tweets (by a factor of 3 standard deviations) and loss-filled tweets (by a factor of 2.1), affirming the research of (Cha et al. 2010; Gupta, Lamba, and Kumaraguru 2013). The credibility score regressions confirmed the Co-UGS conceptual model assertion, and the research of (Castillo, Mendoza, and Poblete 2011), that active users post more believable content; however, this variable was not significant in the credibility frequency model.

The study empirically affirmed the positive direction in the association between retweet speed and credibility in gang-affiliated tweets and found that credibility in tweets is affected (by a factor of 2.1 standard deviations in aggressive tweets and 1.1 in loss-filled tweets) by retweet time. Lastly, the study provided somewhat mixed results on the claims of (Mitra, Wright, and Gilbert 2017) and the Co-UGS conceptual model that tweets with longer message lengths affect social media content credibility. These results were not significant in the credibility frequency test for loss-filled tweets, though that test did affirm the claim that aggressive tweets with longer message lengths are more credible by a standard deviation factor of 1.2. In contrast, the credibility score model affirmed the claims for both aggression and loss.

### 6.1.2 Interpretation of Findings on Violent Crime

The empirical results in Stage 2 provide insights into the relationships between various factors and violent crime. It was discovered that the weekday and the average tweet hour of the day had no significant influence on violent crime. Further, the study found that quarterly seasonality in quarter three, the total number of followers of gang-affiliated Twitter users, the total tweet length, and the total user mentions and booster words had little impact or were not significant in predicting violent crime in the study. Though the inclusion of a hashtag in a tweet is positively associated with tweet credibility, the total daily hashtag frequency is negatively related to violent crime by a small amount. It was discovered that total interactive dissemination frequency, though significant, has a negative influence on violent crime and is partially mediated through credible aggression and credible loss.

The research in Stage 2 furthered the understanding of other factors that significantly and positively contribute to violent crime in gang-affiliated communities. The correlation matrix revealed that violent crime count is higher on days where the crime rate one year earlier was higher. The regression results revealed that total interactive dissemination (retweet) speed (measured in hours) is a significant positive predictor of violent crime and is partially mediated through Credible Aggression and Credible Loss. Higher violent crime is associated with days that have a higher average temperature and days that include a special event. Higher violent crime is also associated with higher violent crime counts in the prior six days and in the first two quarters of the year, relative to the fourth quarter. The Direct Effects Model results revealed that gang-affiliated tweet frequency is positively correlated with violent crime count after certain levels. Lastly, and perhaps most importantly, the study's research model affirmed that aggressive and loss-filled gang-affiliated social media content can indeed be positive signals to imminent violent crime.

## 6.2 Contributions to Research

The study made several contributions to the research literature. It confirmed seven of the ten study hypotheses. It expanded the theoretical understanding (via the novel Co-UGS model) that posting or sharing of credible content related to loss or aggression, combined with dissemination factors and other effective predictors, convey credible underlying intentions and indicates the individual is seeking gratification, status, or acceptance. It created, tested, and confirmed a new credibility construct, adding or affirming eight determinants of social media post credibility in gang-affiliated communities.

It empirically affirmed the theoretical assertions of (DeWitt 2018; Pyrooz and Densley 2016; Lee and Ma 2012) that credible aggressive and loss-filled gang-affiliated social media content is indeed a signal to imminent violent crime. It demonstrated a positive and significant association between retweet speed and violent crime when partially mediated through credible aggression and credible loss. It also showed a somewhat surprising negative association between interactive dissemination frequency and violent crime when partially mediated through credible aggression and credible loss.

Other insights from the study revealed that Gang-Affiliated Tweet frequency is an eventual positive predictor of violent crime, aligning theoretically with (DeWitt 2018) and empirically with (Aghababaei 2017) and (Gerber 2014). It was discovered that tweets with URLs not only affect tweet credibility positively but also influence violent crime positively.

Tweets with hashtags positively affect tweet credibility (affirming the Co-UGS model) but negatively impact violent crime. Finally, it was found that booster words and user mentions positively affect tweet credibility (affirming the Co-UGS model) but have a non-significant impact on violent crime. The number of Twitter followers predicts violent crime, but not significantly and not in the direction the model theorized. Surprisingly, the number of followers affects violent crime count negatively. It is interesting to compare this finding to the recent subjective statement from a domain expert that having too many followers can tarnish the reputation of a gang affiliate on social media.

The study revealed the positive association of control predictors to violent crime. These included the six prior period violent crime counts and the quarterly seasonality (Q1 and Q2) in the data. The work expanded the work of (Aghababaei 2017) by exposing the positive association of the major event predictor (including both local and exogenous events) to violent crime. This study furthered the research of (Aghababaei 2017; Anderson 1987; Chen, Cho, and Jang 2015) by substituting more granular weather daily averages, instead of aggregated monthly ones, and by empirically affirming their strong positive association with violent crime. It added new determinants of violent crime in the research model by revealing the positive and significant association of six prior period violent crime counts and seasonality significance in quarters one and two of the year.

## 6.3 Implications for Practice

A goal of this study was to not only assist academic researchers but practitioners as well. Practitioners realize they can leverage social media analytics to help combat crime, as they understand the potential of social media as a valuable information source. While this study used historical data, ultimately, it is a goal to move it towards a practitioner-based proof of concept. If an automated process can identify Twitter messages (tweets) that credibly taunt, threaten, or exhibit retaliatory loss to others in the virtual space, it may quickly alert law enforcement officers or community workers to stop a future crime in the physical space. The researcher should also consider how to determine the current day's crime count if in a live system. One solution is to implement a connection to live aggregated crime data by the end of the day (e.g., midnight). Other such practitioner-based implementation ideas and decisions should be addressed if the research moves toward a proof of concept.

## 6.4 Limitations and Future Research

While this study revealed worthy information about the determinants of social media content credibility and violent crime related to gangs, it does have limitations. Firstly, the research did not include in its scope the initial (and continuing) puzzling question of why violent crime in U.S. cities, like Chicago, IL reversed direction and began to increase, starting in 2015. Future studies could collect data from years before 2015, compare it to data in 2015 (and later), and study this issue. Collecting more data would also result in another benefit; it could improve previous violent crime models. If the new research used more labeled data, this could lead to better word embedding models, which would eventually improve the accuracy of the final classification and prediction models.

Secondly, this study on gang-related social media employed a pre-built set of tweet ids and Twitter ids of known gang members from 2014. However, because gang membership can change frequently, a potentially better approach is to automate the process of identifying and curating a set of verifiable and specific gang-affiliated Twitter profiles. Automatically finding online gang member profiles to use in developing a training dataset is challenging but very useful. Features could be extracted from tweet text, images, shared Twitter videos and images, shared YouTube video titles, descriptions, posted comments, and top emojis (via the YouTube API16). Researchers could implement this process by creating a dataset of gang members from a substantial set of unbiased Twitter gang member profiles (not specific to any particular neighborhood), comparing against non-gang member profiles, and finding contrasting features via the use of word embeddings and trained Machine Learning classifiers. Because descriptions of profile images may improve a classifier, researchers could extract and tag each Twitter profile and cover image and translate them into features using web services and APIs to tag images with a set of scored keywords. Further, to collect tweets, researchers could also automatically search Twitter profile descriptions by keywords, by gang name and gang slang names (from crowd-sourced knowledge bases, like HipWiki), and by those recently murdered. To discover unbiased Twitter profiles, researchers could follow the methodology used by (Balasuriya et al. 2016) and look for others via retweets, followers, followees, and hashtags commonly used by gang members across the U.S. These improvements would result in a more efficacious framework resulting in automatic retraining of the algorithms over time and location.

Thirdly, there are other ways to improve the gang-affiliated crime prediction architecture by including additional specific features built not only from tweet text and emoji usage but also created from profile images and YouTube video links showcasing gang-related rap music. For example, researchers could add more location-agnostic keywords (Balasuriya et al. 2016) and also introduce custom image tagging models that detect common gang-related items or signals (e.g., gang hand signs or pointed guns) in gang members' profile images as opposed to ones which tend to tag images with generic keywords such as ‘people’ or ‘hands’. Past research used tagging models that did not recognize gang-related objects and thus mischaracterized those (Balasuriya et al. 2016).

Fourthly, future research could add the *count of tweet replies* as an additional measure of credibility for the credibility algorithms in Stage 1 of this study. The research could not include this number for older tweets due to the Twitter Search API limitation of retrieving only recent replies of 6-9 days from the date of the original tweet, and even then, a replies list that is not exhaustive. Though the replies count is available via the Twitter interface, to look up each tweet and enter this count manually would be very time-consuming and most likely disproportionate to the predictive value it would provide. Instead, an option for future researchers is to use the streaming API to capture and store the replies count on a daily or weekly basis for recent timeframes. This was not a possibility for this study, as the timeframe was one to five years ago. However, future researchers could collect additional tweets over the timeframe and include the number of followees and the person to whom the aggression was directed as other measures of credibility. Further, future scholars could establish a credibility construct that can be extended to other social media platforms (not just Twitter) to look at social media credibility using the construct this study developed as the source.

Fifthly, the study confirmed that the total loss credibility score positively associates with the credible signaling of violent crime. However, this construct was not a strong predictor in the research, reflecting its low distribution in real life. Thus, a future researcher or practitioner could design a study or industry-led project that combines credible aggression and credible loss tweet instances and test the improved efficacy of such a full effect model.

Sixthly, future research could employ more novel approaches to text mining and data storage. For example, scholars could use other word embedding techniques like Google’s Bidirectional Encoder Representations from Transformers (BERT), a technique for NLP pre-training, to see if the model performance improves or other neural network models like Recurrent Neural Network (RNN) models and Long Short-Term Memory (LSTM) models. Researchers could further investigate the use of a SQL-based or newer, non-relational data management techniques like Not Only Structured Query Language

(NoSQL) for a more scalable solution. For the loss classifier, where there were not many labeled instances in the corpus, future researchers could experiment with active learning. Active learning is a method where the algorithm can choose the data from which it wants to learn and perform well with much less training data. Active learning systems ask queries via unlabeled instances to be annotated by a human, and thus, attempts to gain high accuracy with few labeled instances to minimize the cost associated with obtaining labeled data (Settles 2009).

Seventhly, future researchers could enrich the corpus word embeddings (vectors) with automatically-induced lexicon knowledge and semantic relations from the unlabeled corpus. Traditionally, researchers have introduced lexicon-based features involving sentiment, emotion, and activity (Zhang et al. 2015). In the gang-affiliated domain, appropriate *manual* features could include sentiment (negative including taunting and threatening language or language including grief), emotion (anger or sadness), weapon (gun), behavior/activity (shoot), cursing, drugs, and emojis (gun, bomb, sad face, etc.). Researchers could also use a derived glossary to locate customary English terms corresponding to slang to access the correct word in a lexicon for each Twitter word. These derived words may appear both in the feature set from the corpus and the seed set as final training features. One potentially efficacious framework could be used to induce an automatic lexicon, SENTPROP. This approach *learns* accurate sentiment lexicons from small sets of gang-affiliated seed words using word embeddings derived from a domain-specific corpus and combines label propagation with advances in word embeddings. It is designed to be accurate, even when using a smaller-sized corpus. It not only provides a learned lexicon but also displays confidence scores (Hamilton et al. 2017), which allows researchers to ethically quantify uncertainty. The design accomplishes this task by building a lexical graph from the unlabeled corpus and using cosine similarity to connect each word (via its meaning) with its nearest k neighbors, and then spreading the sentiment labels over this graph. Thus, a future study could use this method to run random walks from gang-affiliated seed words (Chang et al. 2018) and assign polarity (probability) scores based on the frequency of random walk visits, mapping the words to their association with aggression and loss. This will form the lexicon, scale the class probabilities to a zero mean and a variance of one and combine the lexicon results to get the weighted average of the embeddings and any lexicon scores.

Eighthly, future research could compare the models used in this study to another model that employed a Vector Autoregressive (VAR) technique. In a VAR model, every variable depends on every other variable, so the notation changes because every variable is a y-variant. Further, each row may written as a separate equation, such that a general autoregressive model of order 2, VAR(2), with one lag is written as:

$$y_{1t} = a_{11}y_{1t-1} + a_{12}y_{2t-1} + \epsilon_{1t}$$

$$y_{2t} = a_{21}y_{2t-1} + a_{22}y_{1t-1} + \epsilon_{2t}$$

To deal with the premise of stationarity in a VAR model, the research could use differencing, while also considering the use of the log to transform the variable (required if there is a trend). For seasonality, one could use another transformation technique. Once all of the conditions for a VAR were met, the researcher could use the following series of equations for a one-day lag for *Violent Crime Count* and the other model variables:

$$VC_t = \alpha_{11}VC_{1t-1} + a_{12}CA_{t-1} + a_{13}CL_{t-1} + a_{14}GATF_{t-1} + a_{15}CALF_{t-1} + a_{16}IDF_{t-1} + a_{17}IDS_{t-1} + e_{1t}$$

$$CA_t = \alpha_{21}CA_{1t-1} + a_{22}VC_{t-1} + a_{23}CL_{t-1} + a_{24}GATF_{t-1} + a_{25}CALF_{t-1} + a_{26}5IDF_{t-1} + a_{27}IDS_{t-1} + e_{2t}$$

$$CL_t = \alpha_{31}CL_{1t-1} + a_{32}VC_{t-1} + a_{33}CA_{t-1} + a_{34}GATF_{t-1} + a_{35}CALF_{t-1} + a_{36}IDF_{t-1} + a_{37}IDS_{t-1} + e_{3t}$$

$GATF_t = \alpha_{41}GATF_{1t-1} + a_{42}VC_{t-1} + a_{43}CA_{t-1} + a_{44}CL_{t-1} + a_{45}CALF_{t-1} + a_{46}IDF_{t-1} + a_{47}IDS_{t-1} + e_{4t}$

$CALF_t = \alpha_{51}CALF_{1t-1} + a_{52}VC_{t-1} + a_{53}CA_{t-1} + a_{54}CL_{t-1} + a_{55}GATF_{t-1} + a_{56}IDF_{t-1} + a_{57}IDS_{t-1} + e_{5t}$

$IDF_t = \alpha_{61}IDF_{1t-1} + a_{62}VC_{t-1} + a_{63}CA_{t-1} + a_{64}CL_{t-1} + a_{65}GATF_{t-1} + a_{66}CALF_{t-1} + a_{67}IDS_{t-1} + e_{6t}$

$IDS_t = \alpha_{71}IDS_{1t-1} + a_{72}VC_{t-1} + a_{73}CA_{t-1} + a_{74}CL_{t-1} + a_{75}GATF_{t-1} + a_{76}CALF_{t-1} + a_{77}IDF_{t-1} + e_{7t}$

In these equations, VC is Violent Crime Count, CA is Credible Aggression (Sum of scores), CL is Credible Loss (Sum of Scores), GATF is Gang-Affiliated Tweet Frequency, CALF is Credible Aggression and Loss Tweet Frequency, IDF is Interactive Dissemination Frequency, IDS is Average Interactive Dissemination Speed, t is the time unit (e.g., day), $\alpha$ is the parameter coefficient, and e is the error term. The researcher could also experiment with whether to include (and how many) lags in the equations by performing a test to determine the optimal number or order (p) of lags (up t-365). Measuring the partial autocorrelation function (PACF), one would only keep the lags in the model that are high in magnitude (positive or negative) and are direct effects to the variable of interest. PACF provides the partial correlation of a stationary time series with its own lagged values. These are regressed on the time series values at shorter lags. The following equation could be used for multiple lags: $X_t = \alpha + \Phi_1 X_{t-1} + ... + \Phi_p X_{t-p} + \varepsilon_t$. After training the model, the researcher must roll back the transformations and evaluate the model using the test dataset.

To evaluate and compare the model results for the VAR model, the researcher could use the fit criterion Akaike's Information Criteria (AIC) due to its favorable small sample forecasting features. However, because this metric can choose large numbers of lags, it should be used with caution (Hyndman and Athanasopoulos 2018) and compared to another fit assessment metric, the Bayesian Information Criteria (BIC). BIC is an estimate of a function of the posterior probability of a model being true. After selecting the best

VAR model, the researcher could attempt to interpret its estimated parameter values. However, these provide only limited information on the reaction of a system to a shock since all VAR model variables are dependent on one another. Thus, the researcher may need to use impulse responses to understand the model's dynamic behavior.

## 6.5 Conclusion

In conclusion, though this interdisciplinary study focused on using IS-based predictive techniques in the domain of Criminology, the social-media-based approach is generalizable (with some changes), and researchers could employ it on problems in other domains. Thus, it is the researcher's desire that IS and Criminology scholars, as well as those in other disciplines, continue to use and improve the approach revealed in this research for the good of society, especially for the benefit of individuals who live and work in dangerous urban areas.

# REFERENCES

Aghababaei, S. 2017. 'A Temporal Topic Model for Social Trend Prediction', University of Ontario Institute of Technology

Aghababaei, S., and M. Makrehchi. 2018. 'Mining social media content for crime prediction', *Intelligent Data Analysis: IDA*, 22: 117-41.

AJ+Docs. 2017. 'Chicago's Struggle with Gun Violence '. https://www.youtube.com/watch?v=QbkA1e_uj_E.

AnalyticsVidyha. 2017. 'An Intuitive Understanding of Word Embeddings: From Count Vectors to Word2Vec'. https://www.analyticsvidhya.com/blog/2017/06/word-embeddings-count-word2veec/.

Anderson, C. A. 1987. 'Temperature and Aggression: Effects on Quarterly, Yearly, and City Rates of Violent and Nonviolent Crime', *Journal of Personality and Social Psychology*, 52: 1161.

Ang, R. 2015. 'Adolescent cyberbullying: A review of characteristics, prevention and intervention strategies', *Aggression and Violent Behavior*.

Aral, S., and D. Walker. 2014. 'Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment ', *Management Science* 60: 1352-70.

Austen, B. 2013. "Public Enemies: Social Media is Fueling Gang Wars in Chicago." In *Wired*, 1.

Balasuriya, L., S. Wijeratne, D. Doran, and A. Sheth. 2016. "Finding Street Gang Member Profiles on Twitter." In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 685-92.

Bandura, A. 1977. *Social learning theory* ( Prentice-Hall: Englewood Cliffs, NJ).

Beckett, K. 1997. *Making crime pay: The politics of law and order in the contemporary United States* (Oxford University Press: New York, NY).

Bild, D. R., L. Yue, R. P. Dick, A. M. Mao, and D. S. Wallach. 2015. 'Aggregate Characterization of User Behavior in Twitter and Analysis of the Retweet Graph', *ACM Transactions on Internet Technology*, 15: 1-24.

Blevins, T., R. Kwiatkowski, J. Macbeth, K. McKeown, and D. Patton. 2016. "Automatically Processing Tweets from Gang-Involved Youth: Towards Detecting Loss and Aggression Technical Papers." In *COLING 2016, the 26th International Conference on Computational Linguistics*, 2196–206. Osaka, Japan.

Boateng, G. O., T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quiñonez, and S. L. Young. 2018. 'Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer', Frontiers in Public Health, Accessed 3/17/2020.

Bowen, D. A., L. M. Kollar, D. T. Wu, D. A. Fraser, C. E. Flood, J. C. Moore, E. W. Mays, and S. A. Sumner. 2018. 'The ability of crime, demographic and business data to forecast areas of increased violence', *International Journal of Injury Control and Safety Promotion*, 25: 443-48.

Braga, A. A., and B. Bond. 2008. 'Policing crime and disorder hot spots: A randomized controlled trial', *Criminology* 46: 577–607.

Braga, A. A., A. V. Papachristos, and D. M. Hureau. 2012. 'The Effects of Hot Spots Policing on Crime: An Updated Systematic Review and Meta-Analysis', *Justice Quarterly*, 31: 633-63.

Braga, A., and D. Weisburd. 2010. *Policing Problem Places: Crime Hot Spots and Effective Prevention.* (University Press, Inc. : Oxford).

Bray, P. 2012. 'When Is My Tweet's Prime of Life? (A brief statistical interlude)'. https://moz.com/blog/when-is-my-tweets-prime-of-life.

Brownlee, J. 2017. 'How to Develop Word Embeddings in Python with Gensim'. https://machinelearningmastery.com/develop-word-embeddings-python-gensim/.

Bruinius, H. 2018. "Why New York crime has plunged to record lows." In *The Christian Science Monitor*.

Bumgarner, B. A. 2007. 'You have been poked: Exploring the uses and gratifications of Facebook among emerging adults', *First Monday*, 12.

Burke, M., C. Marlow, and T. Lento. 2009. "Feed me: motivating newcomer contribution in social network sites." In *27th international conference on human factors in computing systems*, 945–54. Boston, MA, USA.

Burnap, P., and M. Williams. 2015. 'Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making', *Policy and Internet*, 7: 223-42.

Butler, B., L. Sproull, S. Kiesler, and R. Kraut. 2002. 'Community effort in online groups: Who does the work and why.' in S. P. Weisband (ed.), *Leadership at a Distance: Research in Technologically-Supported Work* (Lawrence Erlbaum Associates: New York, NY, USA).

Byers, C. 2014. "Crime up after Ferguson and more police needed, top St. Louis area chiefs say." In *St. Louis Post-Dispatch*.

Byrne, B. 1994. *Coping with Bullying in Schools* (Columba Press: Dublin. Cassell, London, England).

Byrne, C. 2015. 'Drugs, Guns, and Selfies: Gangs on Social Media'. https://igarape.org.br/en/drugs-guns-and-selfies-gangs-on-social-media/.

Carlie, M. 2002. 'Into the Abyss: A Personal Journey into the World of Street Gangs, Part 7: The Structure of Gangs'.

Castillo, C., M. Mendoza, and B. Poblete. 2011. "Information credibility on Twitter." In *20th International Conference on World Wide Web*, 675-84. Hyderabad, India.

Cawley, G. C.; Talbot, N. L. C. 2010. "On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation" (PDF). 11. *Journal of Machine Learning Research*: 2079–2107.

Cha, M., H. Haddadi, F. Benevenuto, and Gummad. K. P. 2010. "Measuring User Influence on Twitter: The Million Follower Fallacy." In *Fourth International AAAI Conference on Weblogs and Social Media*, 10–17. Palo Alto, CA,.

Chainey, S., L. Tompson, and S. Uhlig. 2008. 'The utility of hotspot mapping for predicting spatial patterns of crime', *Security Journal*, 21: 4-28.

Chalfin, A., and J. McCrary. 2013. 'Are U.S. Cities Underpoliced?: Theory and Evidence', *Review of Economics and Statistics*.

Chan, J., A. Ghose, and R. Seamans. 2016. 'The Internet and Racial Hate Crime: Offline Spillovers from Online Access', *MIS Quarterly*, 40: 381-403.

Chang, S., R. Zhong, E. Adams, F. Lee, S. Varia, D. Patton, W. Frey, C. Kedzie, and K. McKeown. 2018. "Detecting Gang-Involved Escalation on Social Media Using Context." In *Conference on Empirical Methods in Natural Language Processing*, 46-56. Brussels, Belgium.

Chen, G. 2011. 'Tweet this: A Uses and Gratifications Perspective on How Active Twitter Use Gratifies a Need to Connect with Others', *Computers in Human Behavior*, 27: 755–62.

Chen, X., Y. Cho, and S. Y. Jang. 2015. "Crime prediction using twitter sentiment and weather." In *Systems and Information Engineering Design Symposium (SIEDS)*, 63-68.

Chicago Police Department. 2018. 'Chicago Police Department End of Year Crime Statistics'. https://home.chicagopolice.org/cpd-end-of-year-crime-statistics-2018/

Chung, D. S., and S. Kim. 2008. 'Blogging activity among cancer patients and their companions: Uses, gratifications, and predictors of outcomes', *Journal of the American Society for Information Science and Technology*, 59: 297-306.

Cohen, L. E., and M. Felson. 1979. 'Social change and crime rate trends: A routine activity approach', *American Sociological Review*: 588–608.

Collobert, R., J. Weston, L. Bottou, M. Karlen, K. Kavukcuglu, and P. Kuksa. 2011. 'Natural Language Processing (Almost) from Scratch', *Journal of Machine Learning Research*, 12: 93–2537.

Conquergood, D. 1997. 'Street literacy.' in J. Flood, S. Brice Heath and D. Lapp (eds.), *Handbook of research on teaching literacy through the communicative and visual arts* (Simon and Schuster: New York, NY).

Dalgleish, D., and A. Myhill. 2004. "Reassuring the Public: A Review of International Policing Interventions." In, edited by Home Office. London.

Deci, E. L., and R. M. Ryan. 1995. 'Human autonomy: The basis for true self-esteem.' in M. Kemis (ed.), *Efficacy, Agency, and Self-esteem* (Plenum: New York, NY, USA).

Decker, S. H. 1996. 'Collective and normative features of gang violence', *Justice Quarterly*, 13: 243-64.

DeWitt, S. 2018. 'On the Prospects of Signaling Theory for Criminological Research: A Comment on Potential Avenues for Future Research '. https://ssrn.com/abstract=3108026.

Dong, W., S. Liao, and Z. Zhang. 2018. 'Leveraging Financial Social Media Data for Corporate Fraud Detection', *Journal of Management IS* 35: 461-87.

Easley, D., and J. Kleinberg. 2010. *Networks, Crowds, and Markets* (Cambridge University Press: New York, NY).

Ehrenfreund, M., and D. Lu. 2016. ' More people were killed last year than in 2014, and no one’s sure why'.

Esbensen, F. A., L. T. Winfree Jr, N. He, and T. J. Taylor. 2001. 'Youth gangs and definitional issues: When is a gang a gang, and why does it matter?', *Crime & delinquency*, 47: 105-30.

FBI. 2015. 'National Gang Report', National Gang Intelligence Center, Federal Bureau of Investigation. http://www.fbi.gov/

———. 2018. "Chicago Police Department – Illinois Uniform Crime Reporting (IUCR) Codes " In, edited by FBI Uniform Crime Reporting (UCR) Program.

Featherstone, R., and M. Deflem. 2003. 'Anomie and Strain: Context and Consequences of Merton’s Two Theories', *Sociological Inquiry*, 73: 471–89.

Firth, J. 1962. 'A synopsis of linguistic theory 1930-1955.' in, *Studies in Linguistic Analysis*.

Fisher, A. B. 1978. *Perspectives on human communication* (Macmillan: New York, NY).

Fisher, D. 2012. "How Washington D.C. Got Off The Most Dangerous Cities List." In *Forbes*.

Foltz, C. 2004. 'Cyberterrorism, computer crime, and reality', *Information Management & Computer Security*, 12: 154-66.

Franco, D., C. Romero, and E. Saffell. 2018. *The gang book: a detailed overview of street gangs in the Chicago metropolitan area* (Chicago Crime Commission: Chicago, IL).

Frey, W. R., D. U. Patton, M. B. Gaskell, and K. A. McGregor. 2018. 'Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured twitter data', *Social Science Computer Review*.

Friedrich, M. 2017. 'The Mean Tweets of New York'. https://www.citylab.com/equity/2017/03/the-mean-tweets-of-new-york/520077/.

Gerber, M. S. 2014. 'Predicting Crime using Twitter and Kernel Density Estimation', *Decision Support Systems* 61: 115–25.

Gerrell, M. 2016. 'Hot Spot Policing With Actively Monitored CCTV Cameras: Does it Reduce Assaults in Public Places?', *International CJ Review*, 26: 187-201.

Gimpel, K., N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith. 2011. "Part-of-speech tagging for Twitter: Annotation, features, and experiments." In *ACL*.

Gottfredson, M., and T. Hirschi. 1990. *A General Theory of Crime* (Stanford University Press: Stanford).

Goudie, C. 2015. 'Despite 'Chiraq' label, Data Show Chicago not Even Close to Iraq', ABC Chicago. https://abc7chicago.com/news/despite-chiraq-label-data-show-chicago-not-even-close-to-iraq/886958/.

Granovetter, M. 1985. 'Economic Action and Social Structure: The Problem of Embeddedness', *American Journal of Sociology*, 91: 481-510.

Gupta, A., H. Lamba, and P. Kumaraguru. 2013. "$1.00 per RT #BostonMarathon #PrayForBoston: Analyzing Fake Content on Twitter." In *eCrime Researchers Summit*, 1-12. eCrime.

Hagan, J. 1993. 'The Social Embeddedness of Crime and Unemployment', *Criminology* 31: 65–91.

Hamilton, W. L., K. Clark, J. Leskovec, and D. Jurafsky. 2017. 'Inducing domain-specific sentiment lexicons from unlabeled corpora'. https://arxiv.org/abs/1606.02820.

Han, B., and T. Baldwin. 2011. "Lexical Normalisation of Short Text Messages: Makn sens a #twitter." In *49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT2011*, 368–78. Portland, Oregon, USA.

Helsley, R., and Y. Zenou. 2014. 'Social Networks and Interactions in Cities', *Journal of Economic Theory*, 150: 426-66.

Hirschfield, P. J., and D. Simon. 2010. 'Legitimating police violence newspaper narratives of deadly force', *Theoretical Criminology*, 14: 155–82.

Hoang, T. A., and E. P. Lim. 2011. "Virality and Susceptibility in Information Diffusions." In *Sixth International AAAI Conference on Weblogs and Social Media.*, edited by N. B. Ellison, J. G. Shanahan and Z. Tufekci, 146–53. Palo Alto, CA.

Hollenbaugh, E. E. 2010. 'Personal journal bloggers: Profiles in disclosiveness', *Computers in Human Behavior* 26: 1657-66.

Hootsuite. 2020. '25 Twitter Stats All Marketers Need to Know in 2020', Hootsuite.com. https://blog.hootsuite.com/twitter-statistics/.

Horowitz, R. 1983. *Honor and the American dream* (Rutgers University Press: New Brunswick, NJ).

Howell, J. C. 2010. "Gang Prevention: An Overview of Research and Programs " In, edited by Office of Juvenile Justice and Delinquency Prevention, 1-24. Washington, DC.

Hu, Y., K. Talamadupula, and S. Kambhampati. 2013. "Dude, srsly?: The surprisingly formal nature of twitter's language." In *7th International Conference on Weblogs and Social Media ICWSM*, 244-53.

Huebner, B., K. Martin, R. Jr. Moule, D. Pyrooz, and S. H. Decker. 2016. 'Dangerous places: Gang members and neighborhood levels of gun assault', *Justice Quarterly*: 836-62.

Hughes, L., and J. Short. 2005. 'Disputes involving youth street gang members: Micro social contexts', *Criminology*: 43-76.

Hyland, K. 2002. 'Authority and invisibility: Authorial identity in academic writing', *Journal of Pragmatics*, 34: 1091-112.

Hyndman, R. J., and G. Athanasopoulos. 2018. *Forecasting: Principles and Practice* (OTexts: Melbourne, Australia.).

Jakobsen, J.C., Gluud, C., Wetterslev, J. 2017. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Med Res Methodol* 17, 162. https://doi.org/10.1186/s12874-017-0442-1

Jamieson, R., P. Land, D. Winchester, G. Stephens, A. Steel, A. Maurushat, and R. Sarre. 2012. 'Addressing Identity Crime in Crime Management IS: Definitions, Classification, and Empirics', *Computer Law and Security Review*, 28: 381–95.

Johnson, P. R., and S. Yang. 2009. "Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use." In *Communication Technology Division of the annual convention of the Association for Education in Journalism and Mass Communication*. Boston, MA.

Katz, C., and S. Schnebly. 2011. 'Neighborhood variation in gang member concentrations', *Crime & delinquency*: 377-407.

Katz, E., J. G. Blumler, and M. Gurevitch (ed.)^(eds.). 1974. *Utilization of mass communication by the individual* (SAGE: Beverly Hills, CA).

Khanapur, N., and A. Patro. 2015. 'Design and Implementation of Enhanced Version of MRC6 Algorithm for Data Security', *International Journal of Advanced Computer Research*, 5.

Kim, Y. 2014. "Convolutional Neural Networks for Sentence Classification." In *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–51. Doha, Qatar: Association for Computational Linguistics.

King, J. E., C. E. Walpole, and K. Lamon. 2007. 'Surf and Turf Wars Online—Growing Implications of Internet Gang Violence', *Journal of Adolescent Health* 41: 566-68.

Ko, H. 2000. "Internet uses and gratifications: Understanding motivations for using the Internet." In *83rd Annual Meeting of the Association for Education in Journalism and Mass Communication*. Phoenix, AZ.

Krueger, R. F., K. E. Markon, C. J. Patrick, S. D. Benning, and M. D. Kramer. 2007. 'Linking Antisocial Behavior, Substance Use, and Personality: An Integrative

Quantitative Model of the Adult Externalizing Spectrum', *Journal of Abnormal Psychology*, 116: 645-66.

Kwak, H., C. Lee, H. Park, and S. Moon. 2010. "What is Twitter, a Social Network or a News Media?" In *International World Wide Web Conference Committee (IW3C2)*. Raleigh, North Carolina, USA.

Latos, A. 2017. '9 Investigates: An inside look at how feds took down MS-13 in Charlotte'. https://www.wsoctv.com/news/9-investigates/thursday-inside-look-at-how-federal-agents-took-down-ms-13-in-charlotte/489221757.

Lauger, T. R., and J. A. Densley. 2018. 'Broadcasting Badness: Violence, Identity, and Performance in the Online Gang Rap Scene', *Justice Quarterly*, 35: 816–41.

Lazzati, N., and A. A. Menichini. 2016. 'Hot Spot Policing: A Study of Place-Based Strategies for Crime Prevention', *Southern Economic Journal*, 82: 893-913.

Lee, C. S., and L. Ma. 2012. 'News Sharing in Social Media: The Effect of Gratifications and Prior Experience', *Computers in Human Behavior*, 28: 331–39.

Leetaru, K. H., S. Wang, G. Cao, A. Padmanabhan, and E. Shook. 2013. 'Mapping the global Twitter heartbeat: The geography of Twitter', *First Monday*, 18.

Leovy, J. 2015. *Ghettoside* (Spiegel & Grau: New York, NY).

Lilleberg, J., Y. Zhu, and Y. Zhang. 2015. "Support vector machines and word2vec for text classification with semantic features." In *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)*, 136–40.

Lowry, P. B., J. Gaskin, N. W. Twyman, B. Hammer, and T. L. Roberts. 2013. 'Taking fun and games seriously: Proposing the hedonic-motivation system adoption model (HMSAM)', *Journal of the Association for IS*, 14: 617–71.

Mac Donald, H. 2016. *The War on Cops: How the New Attack on Law and Order Makes Everyone Less Safe* (Encounter Books: United Kingdom).

Malleson, N., and M. A. Andresen. 2015. 'Spatio-temporal crime hotspots and the ambient population', *Crime Science*, 4: 4-8.

Maslow, A. 1987. *In Motivation and personality* (Harper and Row: New York, NY).

Maslow, A. H. 1943. 'A theory of human motivation', *Psychological Review*, 50: 370-96.

McClure, S., J. Scambray, and G. Kurtz. 2009. *Hacking Exposed 6: Network Security Secrets & Solutions* (McGraw-Hill Osborne Media: New York, NY:).

McHugh M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3), 276–282.

Mikolov, T, K. Chen, G. S. Corrado, and J. Dean. 2013. 'Efficient estimation of word representations in vector space', 1301.3781.

Mikolov, T. 2013. "de-obfuscated Python + question." In *word2vec-toolkit*.

Miller, W. B. 1992. "Crime by youth gangs and groups in the United States." In, edited by Office of Justice Programs US Department of Justice, Office of Juvenile Justice and Delinquency Prevention Washington, DC.

Mitra, T., G. P. Wright, and E. Gilbert. 2017. "A Parsimonious Language Model of Social Media Credibility Across Disparate Events." In *20th ACM Conference on Computer-Supported Cooperative Work and Social Computing*, 126-45. Portland, Oregon.

Morselliand, C., and D. Décary-Hétu. 2013. 'Crime facilitation purposes of social networking sites: A review and analysis of the 'cyberbanging' phenomenon', *Small Wars and Insurgencies*, 24: 152–70.

Murray, H. A. 1953. *Explorations in personality* (Oxford Hill: New York, NY).

NBC. 2018. 'A Look at Today’s Gangs and How They’ve Changed'. https://www.nbcchicago.com/news/local/chicago-crime-commission-gang-book-485166811.html.

Neyfakh, L. 2016. 'How Did Chicago Get So Violent?', Slate. https://slate.com/news-and-politics/2016/09/is-chicagos-ghastly-murder-rate-the-result-of-its-1990s-anti-gang-policies.html

NoSlang. 2018a. 'Noslang drug slang translator'. http://www.noslang.com/drugs/dictionary.php.

———. 2018b. 'Noslang slang translator'. http://www.noslang.com/.

NY Crime Commission. 2017. 'NYC Program Addresses Online Violence Epidemic, Gets Real Results ', Citizens Crime Commission of New York City. http://www.nycrimecommission.org/pdfs/Release-E-Responder-Evaluation-Interruption-Toolkit.pdf.

NY Times. 2019. 'MS-13 Gang Members Charged With Brutal Murders in Los Angeles'. https://www.nytimes.com/2019/07/16/us/ms13-la-murders-indictment.html.

O’Donovan, J., B. Kang, G. Meyer, T. Hollerer, and Adalii. S. 2012. "Credibility in context: An analysis of feature distributions in twitter." In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, 293-301 ASE/I. Amsterdam, Netherlands.

Ormrod, J. E. 2012. *Human Learning* (Pearson).

Owoputi, O., B. O’Connor, C. Dyer, K. Gimpel, N. Schneider, and N. A. Smith. 2013. "Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters." In *Conference of the North American Chapter of the Association for Computational Linguistics*, 380–90. Atlanta, Georgia: Human Language Technologies.

Papachristos, A., and L. Hughes. 2015. 'Neighborhoods and Street Gangs.' in S. H. Decker and D. C. Pyrooz (eds.), *The handbook of gangs* (Wiley: Hoboken, NJ).

Park, M. 2018. 'Chicago police count fewer murders in 2017, but still 650 people were killed', CNN. https://www.cnn.com/2018/01/01/us/chicago-murders-2017-statistics/index.html.

Patton, D. U., R. D. Eschmann, and D. A. Butler. 2013. 'Internet banging: New trends in social media, gang violence, masculinity and hip hop', *Computers in Human Behavior*, 29: A54–A59.

Patton, D. U., J. S. Hong, M. Ranney, S. Patel, C. Kelley, R. Eschmann, and T. Washington. 2014. 'Social media as a vector for youth violence: A review of the literature', *Computers in Human Behavior*, 35: 548-53.

Pavlou, P., and D. Gefen. 2005. 'Contract violation in online marketplaces: Antecedents, Consequences, and Moderating Role', *IS Research*, 16: 372–99.

Peng, J., A. Agarwal, K. Hosanagar, and R. Iyengar. 2016. 'Network Embeddedness and Content Sharing on Social Media Platforms', 55: 571-85.

PER Forum. 2013. "Social media and tactical considerations for law enforcement." In, edited by the Department of Justice. United States Office of Community Oriented Policing Services and the United States.

Piergallini, M., A. S. Doğruöz, P. Gadde, D. Adamson, and C. Rose. 2014. "Modeling the use of graffiti style features to signal social relations within a multi-domain learning paradigm." In *14th Conference of the European Chapter of the Association for Computational Linguistics*, 107–15.

Pyrooz, D. C. 2014. 'From Your First Cigarette to Your Last Dyin' Day': The Patterning of Gang Membership in the Life-Course', *Journal of Quantitative Criminology*: 349–72.

Pyrooz, D. C., S. H. Decker, and R. K. Moule Jr. 2015. 'Criminal and Routine Activities in Online Settings: Gangs, Offenders, and the Internet', *Justice Quarterly*, 32: 471-99.

Pyrooz, D. C., and J. A. Densley. 2016. 'Selection into Street Gangs: Signaling Theory, Gang Membership, and Criminal Offending', *Journal of Research in Crime and Delinquency*, 53: 447–81.

Raacke, J., and J. Bonds-Raacke. 2008. 'Myspace and Facebook: Applying the uses and gratifications theory to exploring friend-networking sites', *CyberPsychology and Behavior*, 11: 169-74.

Radil, S. M., C. Flint, and G. E. Tita. 2010. 'Spatializing social networks: Using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in Los Angeles', *Annals of the Association of American Geographers*, 100: 307–26.

Raphael, S. and Winter-Ebmer, R. 2001. 'Identifying the effect of unemployment on crime', *Journal of Law and Economics*, 44: 259–83.

Ratcliffe, J., T. Taniguchi, E. Groff, and J. Wood. 2011. 'The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots', *Criminology*, 49: 795–831.

Redmond, M. V. 2015. 'Social Exchange Theory.' in, *English Technical Reports and White Papers*.

Rogers, E. M. 2003. *Diffusion of innovations* (Free Press: New York, NY).

Rosenfeld, R. 2016. "Documenting and Explaining the 2015 Homicide Rise: Research Directions." In, 1-26. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Rosenfeld, R., T. M. Bray, and A. Egley. 1999. 'Facilitating violence: A comparison of gang-motivated, gang-affiliated, and nongang youth homicides', *Journal of Quantitative Criminology*, 15: 495-516.

Roth, R. 2009. *American Homicide* (Harvard University Press: Cambridge, MA).

Rubin, A. M. 2009a. 'Uses-and-gratifications perspective on media effect.' in J. Bryant and M. B. Oliver (eds.), *Media effects: Advances in Theory and Research* (New York: Routledge).

———. 2009b. 'Uses and gratifications: An evolving perspective on media effects.' in J. Bryant and M. B. Oliver (eds.), *The SAGE handbook of media processes and effects* (Washington, D.C.).

Sampson, R. J., and D. J. Bartusch. 1998. 'Legal cynicism and (subcultural?) tolerance of deviance: The neighborhood context of racial differences', *Law & Society Review*, 32: 777-804.

Sampson, R. J., and W. B. Groves. 1989. 'Community structure and crime: Testing social-disorganization theory', *American Journal of Sociology* 94: 774-802.

Sampson, R. J., and J. L. Lauritsen. 1990. 'Deviant lifestyles, proximity to crime, and the offender-victim link in personal violence', *Journal of Research in Crime and Delinquency*, 27: 110-39

Scrivens, R., and R. Frank. 2016. "Sentiment-based Classification of Radical Text on the Web." In *European Intelligence and Security Informatics Conference*, 104-07. Uppsala, Sweden.

Sela-Shayovitz, R. 2012. 'Gangs and the Web: Gang Members' Online Behavior', *Journal of Contemporary CJ* 28: 389-405.

Settles, B. 2009. "Active Learning Literature Survey." In *Computer Sciences Technical Report 1648*. University of Wisconsin–Madison.

Sharkey, P. 2006. 'Navigating dangerous streets: The sources and consequences of street efficacy', *American Sociological Review*: 826-46.

Sheth, A. P., S. Perera, S. Wijeratne, and K. Thirunarayan. 2017. "Knowledge will propel machine understanding of content: extrapolating from current examples " In *International Conference on Web Intelligence*, 1-9. Leipzig, Germany.

Simpson, D., T. J. Gradel, and M. R. Rossi. 2019. 'Corruption in Chicago and Illinois: Anti-Corruption Report #11'. https://www.foxnews.com/us/chicago-is-most-corrupt-big-city-illinois-third-most-corrupt-state-in-country-study-finds.

Singer, J. B. 1998. 'Online Journalists: Foundations for Research into their Changing Roles', *Journal of Computer-Mediated Communication*, 14: 1-19.

Siponen, M., A. Vance, and R. Willison. 2012. 'New Insights into the Problem of Software Piracy: The Effects of Neutralization, Shame, and Moral Beliefs', *Information & Management*, 49: 334–41.

Slutkin, G., C. L. Ransford, and R.B. Decker (ed.)^(eds.). 2015. *Cure Violence—Treating Violent Behavior as a Contagious Disease* (Springer).

Socher, R., A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. 2013. "Recursive deep models for semantic compositionality over a sentiment treebank." In *2013 Conference on Empirical Methods in Natural Language Processing*, 1631–42. Seattle, WA, USA: Association for Computation Linguistics.

Stets, J. E., and P. J. Burke. 2000. 'Identity Theory and Social Identity Theory', *Social Psychology Quarterly*, 63: 224-37.

Stieglitz, S., and L. Dang-Xuan. 2013. 'Emotions and Information Diffusion in social media', *Journal of Management IS*, 29: 217–47.

Storrod, M. L., and J. A. Densley. 2017. "Going viral' and 'going country': the expressive and instrumental activities of street gangs on social media', *Journal of Youth Studies*, 20: 677-96.

Stuart, F. 2019. 'Code of the Tweet: Urban Gang Violence in the Social Media Age', *Social Problems*: 1–17.

Stylianou, A., C. Subramaniam, and Y. Niu. 2019. 'The Role of Knowledge Management in the Relationship between IT Capability and Interorganizational Performance: An Empirical Investigation', *Communications of the Association for IS*, 45: 65-94.

Subramani, M. 2004. 'How do suppliers benefit from information technology use in supply chain relationships', *MIS Quarterly*, 28: 45-73.

Suttles, G. 1972. *The social construction of communities* (University of Chicago Press: Chicago, IL).

Swanson, D. L. 1979. 'Political communication research and the uses and gratifications model: A critique', *Communication Research*, 6: 37-53.

Sysomos. 2010. 'Replies and Retweets on Twitter', Accessed June 4. https://sysomos.com/inside-twitter/twitter-retweet-stats/.

Tan, A. S. 1985. *Mass communications theories and research* (Macmillan: New York, NY).

Tarm, M. 2018. 'Social media altering Chicago street-gang culture fueling violence', *Chicago Sun Times*.

Taylor, B., C. Koper, and D. Woods. 2011. 'A Randomized Controlled Trial of Different Policing Strategies at Hot Spots of Violent Crime', *Journal of Experimental Criminology*, 7: 149-81.

Thornberry, T. P., M. D. Krohn, A. J. Lizotte, K. Tobin, and C. A. Smith. 2003. *Gangs and delinquency in developmental perspective* (Cambridge University Press: New York, NY).

Tian, G., J. Huang, M. Peng, J. Zhu, and Y. Zhang. 2017. 'Dynamic sampling of text streams and its application in text analysis', *Knowledge IS* 507–31.

Tita, G., J. Cohen, and J. Engberg. 2005. 'An ecological study of the location of gang “Set Space”', *Social Problems*: 272-99.

Travis, J., B. Western, and S. Redburn. 2014. *The Growth of Incarceration in the United States: Exploring Causes and Consequences* (National Academies Press: Washington, DC).

Tyler, T. R. 2006. *Why People Obey the Law* (Princeton University Press: Princeton, NJ).

Van Osch, W., and C. K. Coursaris. 2015. "A Meta-analysis of Theories and Topics in Social Media Research." In *48th Hawaii International Conference on System Sciences*, 1668-75. Kauai, Hawaii.

Venkatesh, V. 2000. 'Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model', *IS Research*, 11: 342–65.

Wang, P., B. Xu, J. Xu, G. Tian, C. Liu, and H. Hao. 2016. 'Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification', *Neurocomputing*, 174, Part B: 806 – 14.

Wang, X., and D. E. Brown. 2012. 'The spatio-temporal modeling for criminal incidents', *Security Informatics*, 1: 1-17.

Wang, X., D. E. Brown, and M. S. Gerber. 2012. "Spatiotemporal modeling of criminal incidents using geographic, demographic, and twitter-derived information." In *IEEE International Conference*, 36–41. Washington, DC.

Watson, D. 2016. '‘Hotspot policing’: a comparative analysis of sanctioned acts of policing versus media representations of policing in a stigmatized community in Trinidad', *Police Practice and Research* 17: 520-30.

Whissell, C. 2009. 'Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language', *Psychological Reports*: 509–21.

Whitlock, T. 2020. 'apps.timwhitlock.info'. https://apps.timwhitlock.info/emoji/tables/unicode.

Wijeratne, S., L. Balasuriya, Doran D., and A. Sheth. 2016. "Word Embeddings to Enhance Twitter Gang Member Profile Identification." In *IJCAI Workshop on Semantic Machine Learning (SML 2016)*, 18-24. New York City, NY.

Wu, J., S. Wang, and H. Tsai. 2010. 'Falling in love with online games: The uses and gratifications perspective', *Computers in Human Behavior*, 26: 1862-71.

Yadav, C. S., A. Sharan, and M. L. Joshi. 2014. "Semantic graph based approach for text mining." In *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 596-601. Ghaziabad.

Yang, J., and S. Counts. 2010. "Predicting the speed, scale, and range of information diffusion in Twitter." In *Fourth International AAAI Conference on Weblogs and Social Media*, edited by W. Cohen and S. Gosling, 355–58. Palo Alto, CA.

Yang, J., M. Yu, H. Qin, M. Lu, and C. Yang. 2019. 'Twitter Data Credibility Framework—Hurricane Harvey as a Use Case', *ISPRS International Journal of Geo-Information.*, 8: 1-21.

Ye, Y., and K. Kishida. 2003. "Toward an understanding of the motivation of open source software developers." In *International Conference on Software Engineering (ICSE)*, 419-29.

Zhang, D., H. Xu, Z. Su, and Y Xu. 2015. 'Chinese classification based on word2vec and SVM perf', *Expert Systems with Applications*, 42: 1857-63.

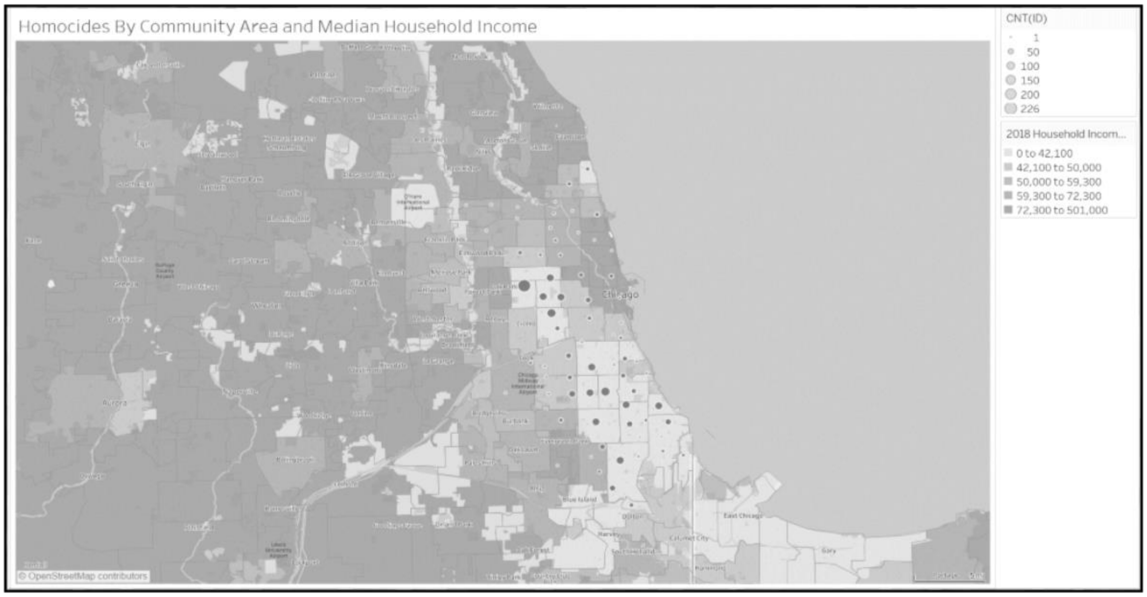
Zhao, X., and J. Tang. 2018. 'Crime in Urban Areas: A Data Mining Perspective', *ACM SIGKDD Explorations Newsletter*, 20: 1-12.

## APPENDIX A: CPD IUCR CODES COMPRISING VIOLENT CRIME

| Description: | IUCR Codes: |
|---|---|
| Homicide | 110, 130, 141, 142 |
| Assault with a Deadly Weapon | 051A, 051B, 520, 530, 545, 550, 551-560 |
| Criminal Sexual Assault | 261-266, 271-275, 281, 291 |
| Robbery of Handgun or other Firearm | 031A, 031B, 033A, 033B |
| Battery of Handgun or other Firearm | 041A, 041B, 450, 451, 480, 481, 488, 489 |
| Ritualism of Handgun or other Firearm | 490, 491 |

https://data.cityofchicago.org/widgets/c7ck-438e

### Homicides (IUCR Code = 110, 130, 141, and 142) in Chicago, IL (2016 – 2018)

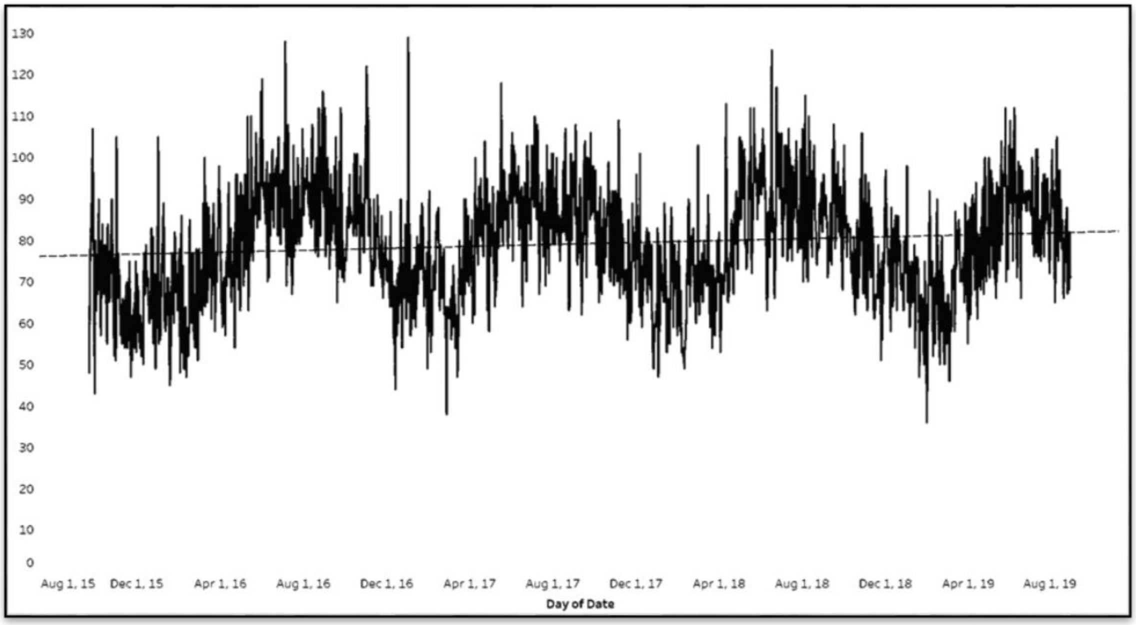


# APPENDIX B: OTHER RELEVANT DATA SOURCES

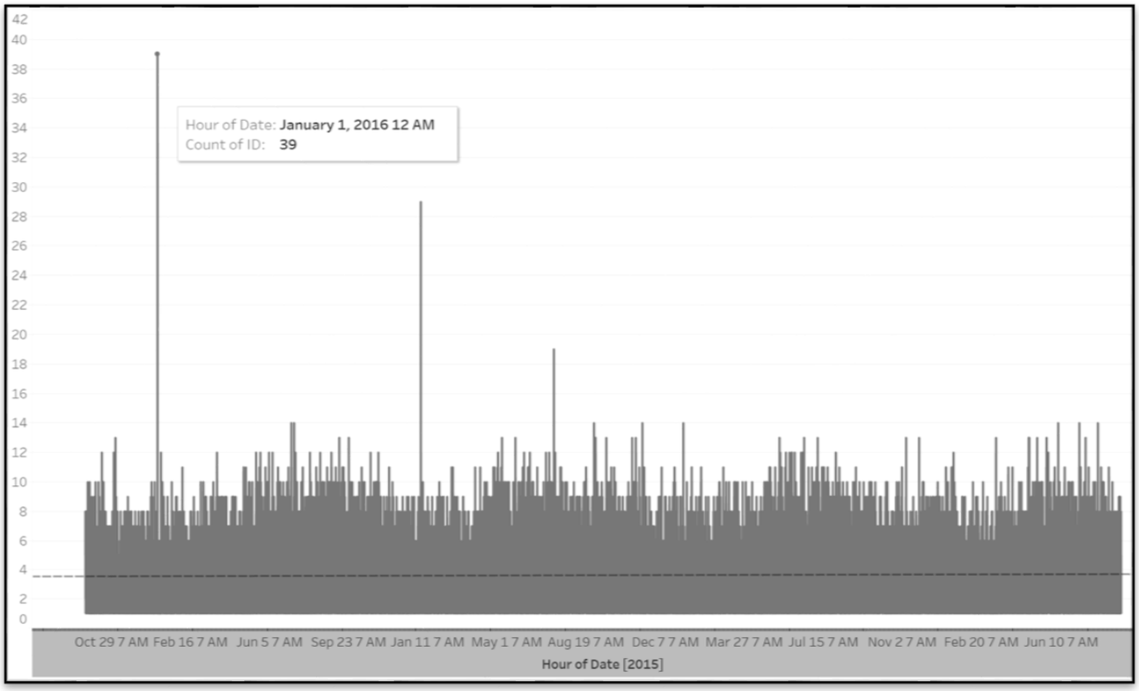
| | |
|---|---|
| Arrest data. These data include District, Beat, Month, Year, Race Code, FBI Code, Statute, Statute Description, Charge Class Code (X, Z, A, B), and Charge Type Code. | https://home.chicagopolice.org/statistics-data/public-arrest-data/ |
| Socioeconomic data. | https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2/data |
| FBI data. The N-DEx system and other well-known FBI systems, such as the National Crime Information Center (NCIC), Interstate Identification Index (III), and Next Generation Identification (NGI) provide critical information to the CJ community. | https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017<br>https://www.fbi.gov/services/cjis/ucr |

# APPENDIX C: CHICAGO, IL VIOLENT CRIME COUNTS

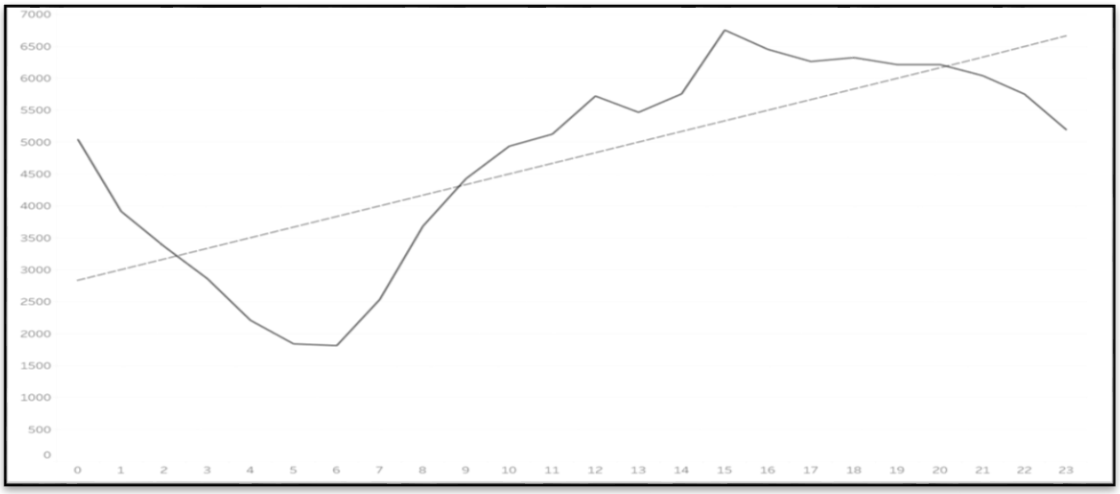
## By Day (9/2015 – 8/2019)



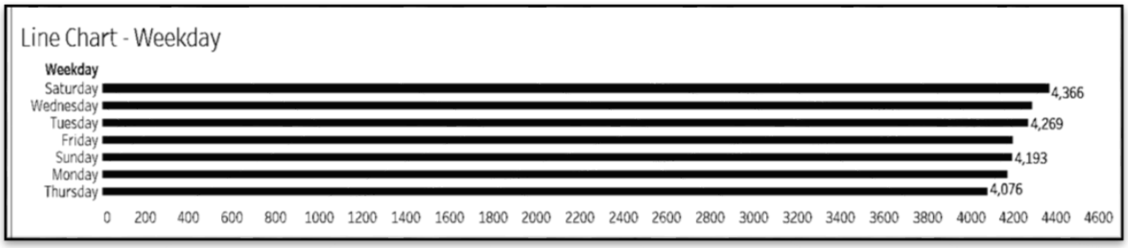
## By Hour within Day:



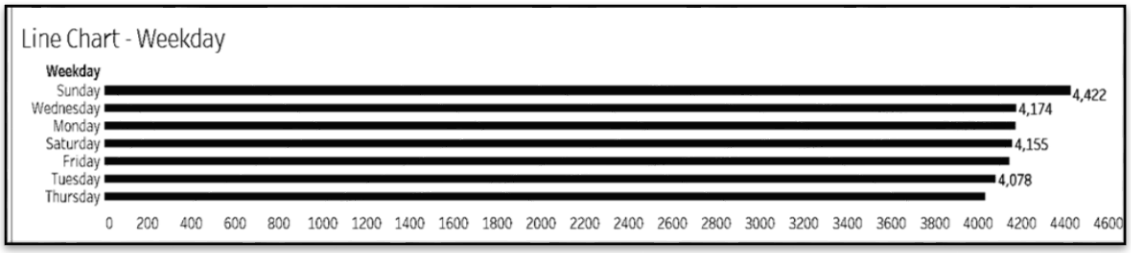
**By Hour:**



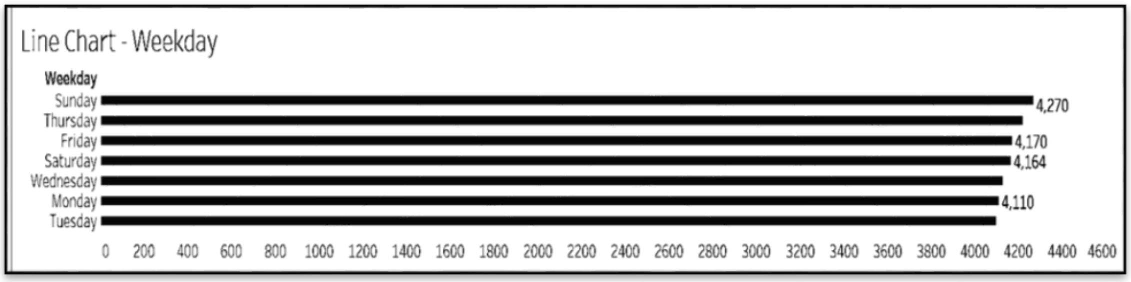
**By Weekday (2016):**



**By Weekday (2017):**

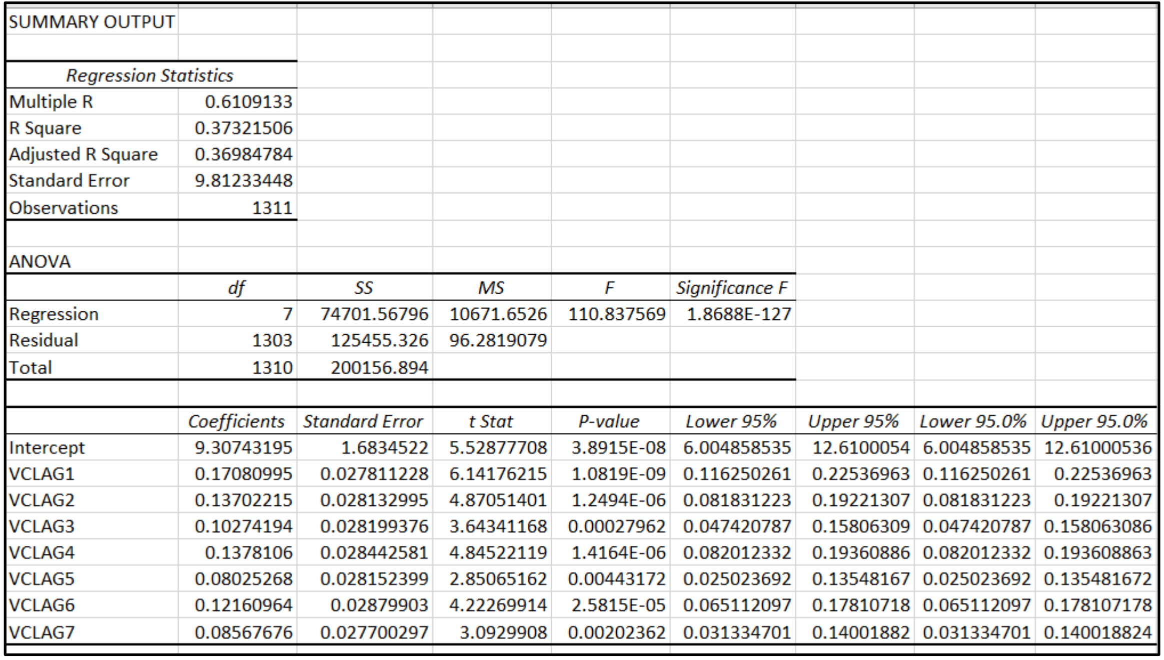


**By Weekday (2018):**



# APPENDIX D: REGRESSION ANALYSIS RESULTS

## Autoregressive Test Results for Violent Crime Counts



SUMMARY OUTPUT

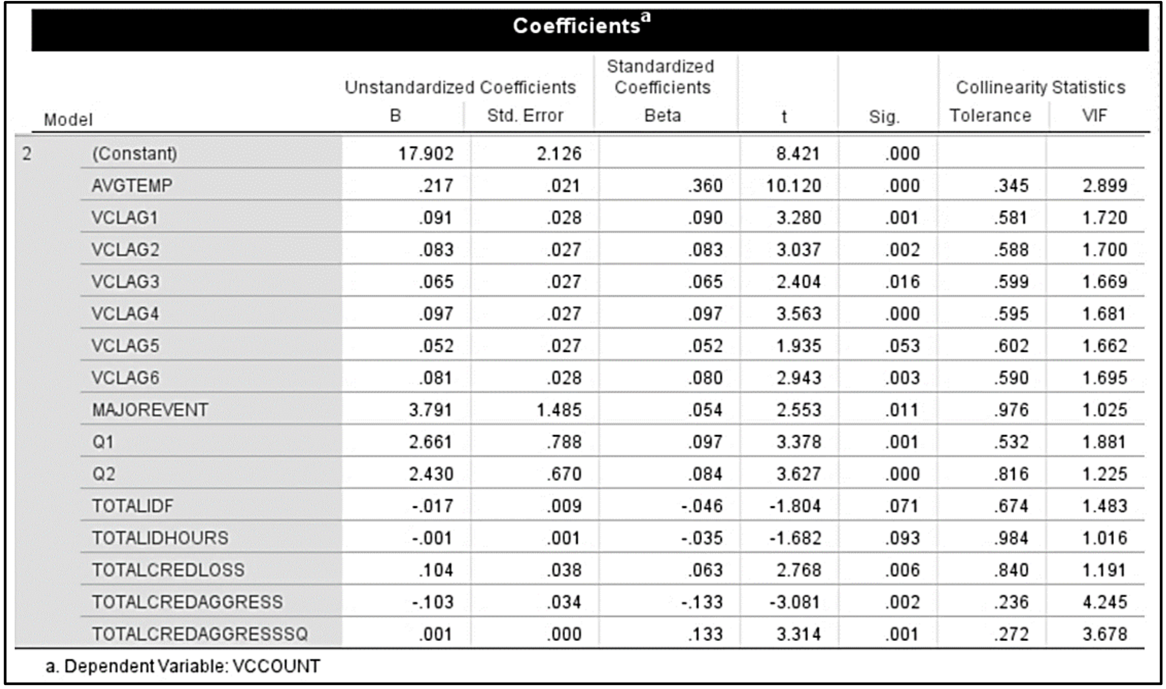
| *Regression Statistics* | |
|---|---|
| Multiple R | 0.6109133 |
| R Square | 0.37321506 |
| Adjusted R Square | 0.36984784 |
| Standard Error | 9.81233448 |
| Observations | 1311 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 7 | 74701.56796 | 10671.6526 | 110.837569 | 1.8688E-127 |
| Residual | 1303 | 125455.326 | 96.2819079 | | |
| Total | 1310 | 200156.894 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 9.30743195 | 1.6834522 | 5.52877708 | 3.8915E-08 | 6.004858535 | 12.6100054 | 6.004858535 | 12.61000536 |
| VCLAG1 | 0.17080995 | 0.027811228 | 6.14176215 | 1.0819E-09 | 0.116250261 | 0.22536963 | 0.116250261 | 0.22536963 |
| VCLAG2 | 0.13702215 | 0.028132995 | 4.87051401 | 1.2494E-06 | 0.081831223 | 0.19221307 | 0.081831223 | 0.19221307 |
| VCLAG3 | 0.10274194 | 0.028199376 | 3.64341168 | 0.00027962 | 0.047420787 | 0.15806309 | 0.047420787 | 0.158063086 |
| VCLAG4 | 0.1378106 | 0.028442581 | 4.84522119 | 1.4164E-06 | 0.082012332 | 0.19360886 | 0.082012332 | 0.193608863 |
| VCLAG5 | 0.08025268 | 0.028152399 | 2.85065162 | 0.00443172 | 0.025023692 | 0.13548167 | 0.025023692 | 0.135481672 |
| VCLAG6 | 0.12160964 | 0.02879903 | 4.22269914 | 2.5815E-05 | 0.065112097 | 0.17810718 | 0.065112097 | 0.178107178 |
| VCLAG7 | 0.08567676 | 0.027700297 | 3.0929908 | 0.00202362 | 0.031334701 | 0.14001882 | 0.031334701 | 0.140018824 |

## Regression Results for Research Model (Significant variables, Standardized)

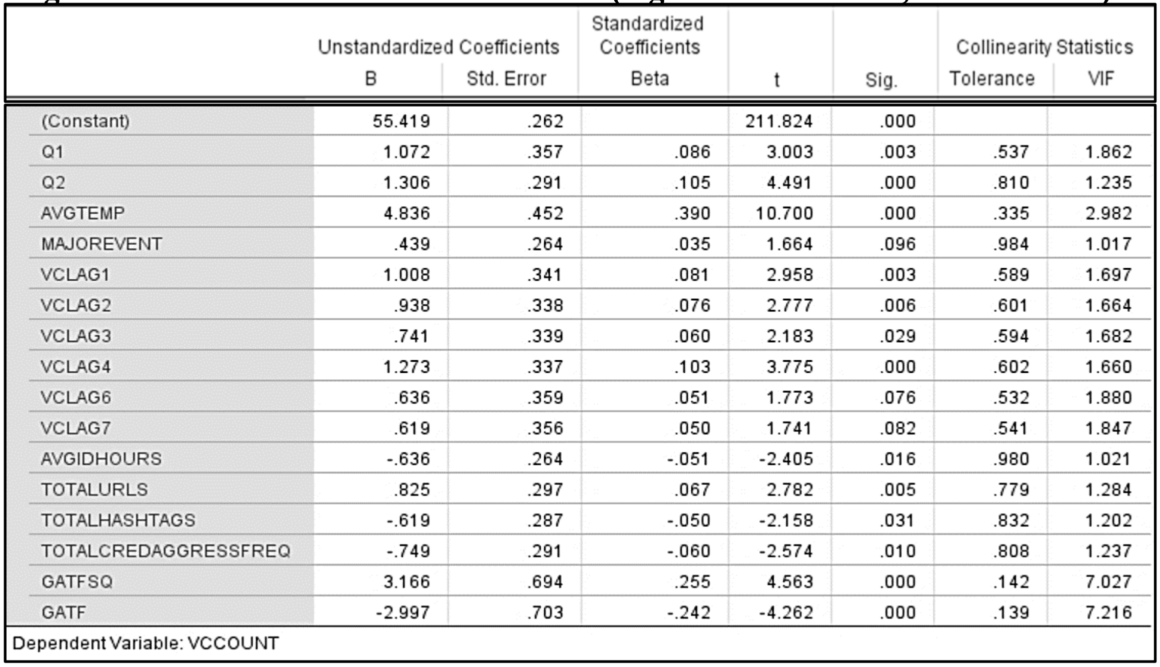


**Coefficients**[a]

| | | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| Model | | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| 2 | (Constant) | 17.902 | 2.126 | | 8.421 | .000 | | |
| | AVGTEMP | .217 | .021 | .360 | 10.120 | .000 | .345 | 2.899 |
| | VCLAG1 | .091 | .028 | .090 | 3.280 | .001 | .581 | 1.720 |
| | VCLAG2 | .083 | .027 | .083 | 3.037 | .002 | .588 | 1.700 |
| | VCLAG3 | .065 | .027 | .065 | 2.404 | .016 | .599 | 1.669 |
| | VCLAG4 | .097 | .027 | .097 | 3.563 | .000 | .595 | 1.681 |
| | VCLAG5 | .052 | .027 | .052 | 1.935 | .053 | .602 | 1.662 |
| | VCLAG6 | .081 | .028 | .080 | 2.943 | .003 | .590 | 1.695 |
| | MAJOREVENT | 3.791 | 1.485 | .054 | 2.553 | .011 | .976 | 1.025 |
| | Q1 | 2.661 | .788 | .097 | 3.378 | .001 | .532 | 1.881 |
| | Q2 | 2.430 | .670 | .084 | 3.627 | .000 | .816 | 1.225 |
| | TOTALIDF | -.017 | .009 | -.046 | -1.804 | .071 | .674 | 1.483 |
| | TOTALIDHOURS | -.001 | .001 | -.035 | -1.682 | .093 | .984 | 1.016 |
| | TOTALCREDLOSS | .104 | .038 | .063 | 2.768 | .006 | .840 | 1.191 |
| | TOTALCREDAGGRESS | -.103 | .034 | -.133 | -3.081 | .002 | .236 | 4.245 |
| | TOTALCREDAGGRESSSQ | .001 | .000 | .133 | 3.314 | .001 | .272 | 3.678 |

a. Dependent Variable: VCCOUNT

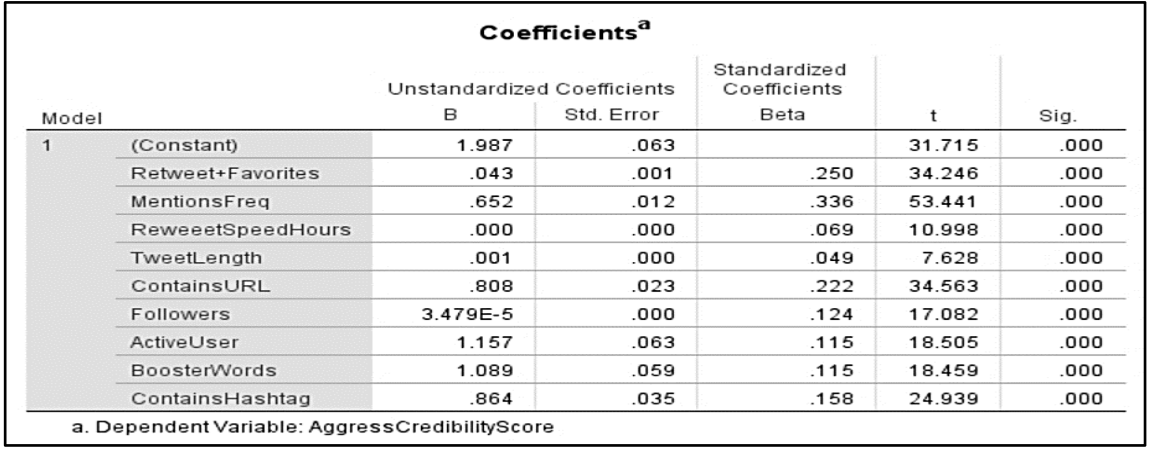
## Regression Results for Direct Effects Model (Significant variables, Standardized)



| | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|
| | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| (Constant) | 55.419 | .262 | | 211.824 | .000 | | |
| Q1 | 1.072 | .357 | .086 | 3.003 | .003 | .537 | 1.862 |
| Q2 | 1.306 | .291 | .105 | 4.491 | .000 | .810 | 1.235 |
| AVGTEMP | 4.836 | .452 | .390 | 10.700 | .000 | .335 | 2.982 |
| MAJOREVENT | .439 | .264 | .035 | 1.664 | .096 | .984 | 1.017 |
| VCLAG1 | 1.008 | .341 | .081 | 2.958 | .003 | .589 | 1.697 |
| VCLAG2 | .938 | .338 | .076 | 2.777 | .006 | .601 | 1.664 |
| VCLAG3 | .741 | .339 | .060 | 2.183 | .029 | .594 | 1.682 |
| VCLAG4 | 1.273 | .337 | .103 | 3.775 | .000 | .602 | 1.660 |
| VCLAG6 | .636 | .359 | .051 | 1.773 | .076 | .532 | 1.880 |
| VCLAG7 | .619 | .356 | .050 | 1.741 | .082 | .541 | 1.847 |
| AVGIDHOURS | -.636 | .264 | -.051 | -2.405 | .016 | .980 | 1.021 |
| TOTALURLS | .825 | .297 | .067 | 2.782 | .005 | .779 | 1.284 |
| TOTALHASHTAGS | -.619 | .287 | -.050 | -2.158 | .031 | .832 | 1.202 |
| TOTALCREDAGGRESSFREQ | -.749 | .291 | -.060 | -2.574 | .010 | .808 | 1.237 |
| GATFSQ | 3.166 | .694 | .255 | 4.563 | .000 | .142 | 7.027 |
| GATF | -2.997 | .703 | -.242 | -4.262 | .000 | .139 | 7.216 |

Dependent Variable: VCCOUNT

## Aggression Credibility Score Train Results

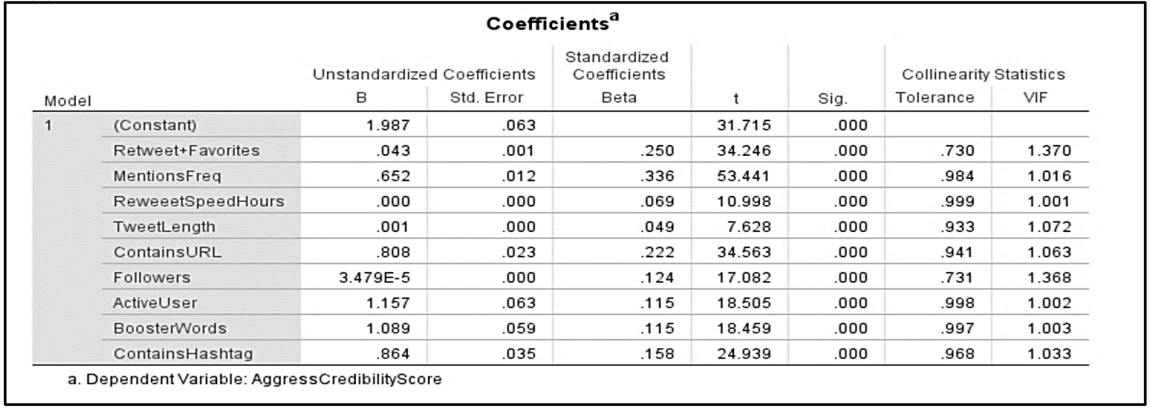


**Coefficients**[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. |
| 1 | (Constant) | 1.987 | .063 | | 31.715 | .000 |
| | Retweet+Favorites | .043 | .001 | .250 | 34.246 | .000 |
| | MentionsFreq | .652 | .012 | .336 | 53.441 | .000 |
| | ReweeetSpeedHours | .000 | .000 | .069 | 10.998 | .000 |
| | TweetLength | .001 | .000 | .049 | 7.628 | .000 |
| | ContainsURL | .808 | .023 | .222 | 34.563 | .000 |
| | Followers | 3.479E-5 | .000 | .124 | 17.082 | .000 |
| | ActiveUser | 1.157 | .063 | .115 | 18.505 | .000 |
| | BoosterWords | 1.089 | .059 | .115 | 18.459 | .000 |
| | ContainsHashtag | .864 | .035 | .158 | 24.939 | .000 |

a. Dependent Variable: AggressCredibilityScore

## Aggression Credibility Score Train and Test Results

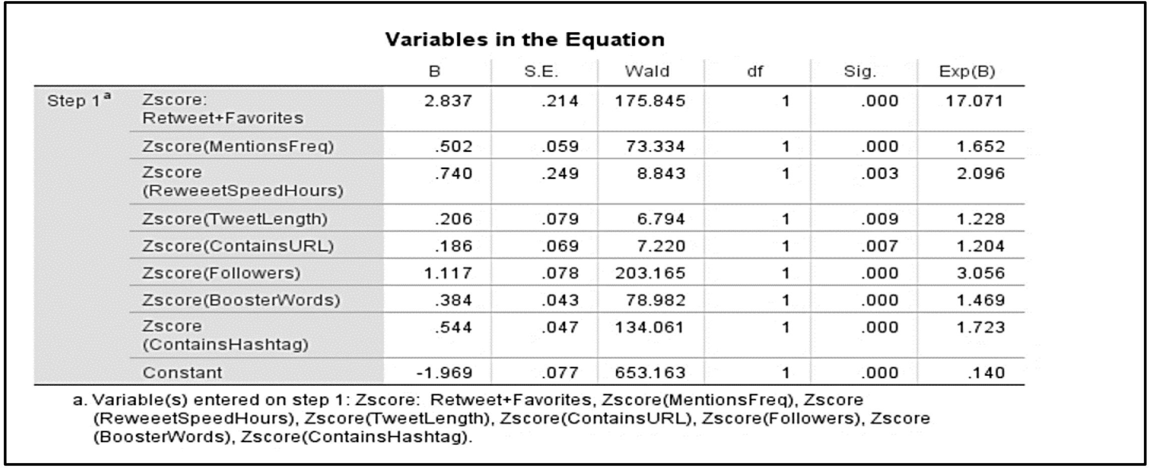


**Coefficients**[a]

| Model | | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| 1 | (Constant) | 1.987 | .063 | | 31.715 | .000 | | |
| | Retweet+Favorites | .043 | .001 | .250 | 34.246 | .000 | .730 | 1.370 |
| | MentionsFreq | .652 | .012 | .336 | 53.441 | .000 | .984 | 1.016 |
| | ReweeetSpeedHours | .000 | .000 | .069 | 10.998 | .000 | .999 | 1.001 |
| | TweetLength | .001 | .000 | .049 | 7.628 | .000 | .933 | 1.072 |
| | ContainsURL | .808 | .023 | .222 | 34.563 | .000 | .941 | 1.063 |
| | Followers | 3.479E-5 | .000 | .124 | 17.082 | .000 | .731 | 1.368 |
| | ActiveUser | 1.157 | .063 | .115 | 18.505 | .000 | .998 | 1.002 |
| | BoosterWords | 1.089 | .059 | .115 | 18.459 | .000 | .997 | 1.003 |
| | ContainsHashtag | .864 | .035 | .158 | 24.939 | .000 | .968 | 1.033 |

a. Dependent Variable: AggressCredibilityScore

## Aggression Credibility Frequency Classification Results - Final

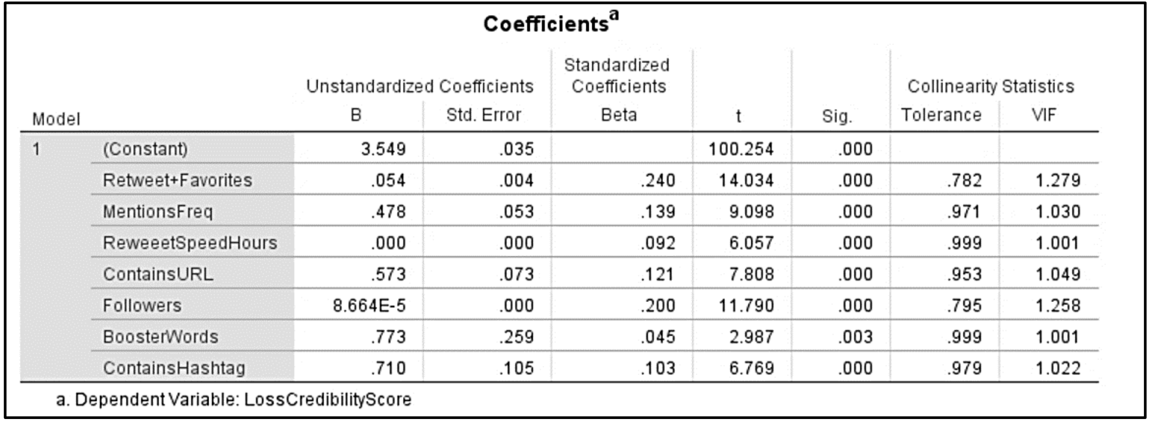


**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | Zscore: Retweet+Favorites | 2.837 | .214 | 175.845 | 1 | .000 | 17.071 |
| | Zscore(MentionsFreq) | .502 | .059 | 73.334 | 1 | .000 | 1.652 |
| | Zscore (ReweeetSpeedHours) | .740 | .249 | 8.843 | 1 | .003 | 2.096 |
| | Zscore(TweetLength) | .206 | .079 | 6.794 | 1 | .009 | 1.228 |
| | Zscore(ContainsURL) | .186 | .069 | 7.220 | 1 | .007 | 1.204 |
| | Zscore(Followers) | 1.117 | .078 | 203.165 | 1 | .000 | 3.056 |
| | Zscore(BoosterWords) | .384 | .043 | 78.982 | 1 | .000 | 1.469 |
| | Zscore (ContainsHashtag) | .544 | .047 | 134.061 | 1 | .000 | 1.723 |
| | Constant | -1.969 | .077 | 653.163 | 1 | .000 | .140 |

a. Variable(s) entered on step 1: Zscore: Retweet+Favorites, Zscore(MentionsFreq), Zscore (ReweeetSpeedHours), Zscore(TweetLength), Zscore(ContainsURL), Zscore(Followers), Zscore (BoosterWords), Zscore(ContainsHashtag).

## Loss Credibility Score Train Results

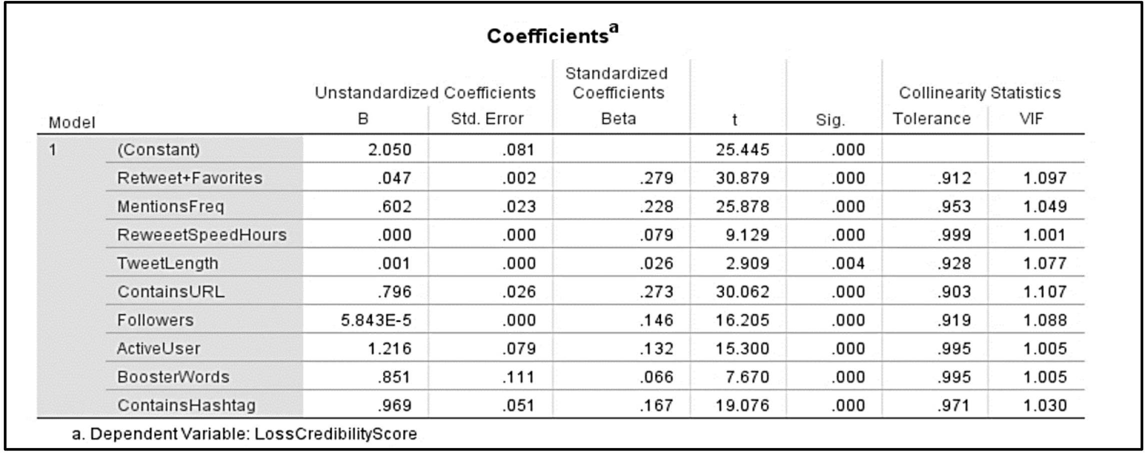


**Coefficients[a]**

| | | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| Model | | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| 1 | (Constant) | 3.549 | .035 | | 100.254 | .000 | | |
| | Retweet+Favorites | .054 | .004 | .240 | 14.034 | .000 | .782 | 1.279 |
| | MentionsFreq | .478 | .053 | .139 | 9.098 | .000 | .971 | 1.030 |
| | ReweeetSpeedHours | .000 | .000 | .092 | 6.057 | .000 | .999 | 1.001 |
| | ContainsURL | .573 | .073 | .121 | 7.808 | .000 | .953 | 1.049 |
| | Followers | 8.664E-5 | .000 | .200 | 11.790 | .000 | .795 | 1.258 |
| | BoosterWords | .773 | .259 | .045 | 2.987 | .003 | .999 | 1.001 |
| | ContainsHashtag | .710 | .105 | .103 | 6.769 | .000 | .979 | 1.022 |

a. Dependent Variable: LossCredibilityScore

## Loss Credibility Score Train and Test Results

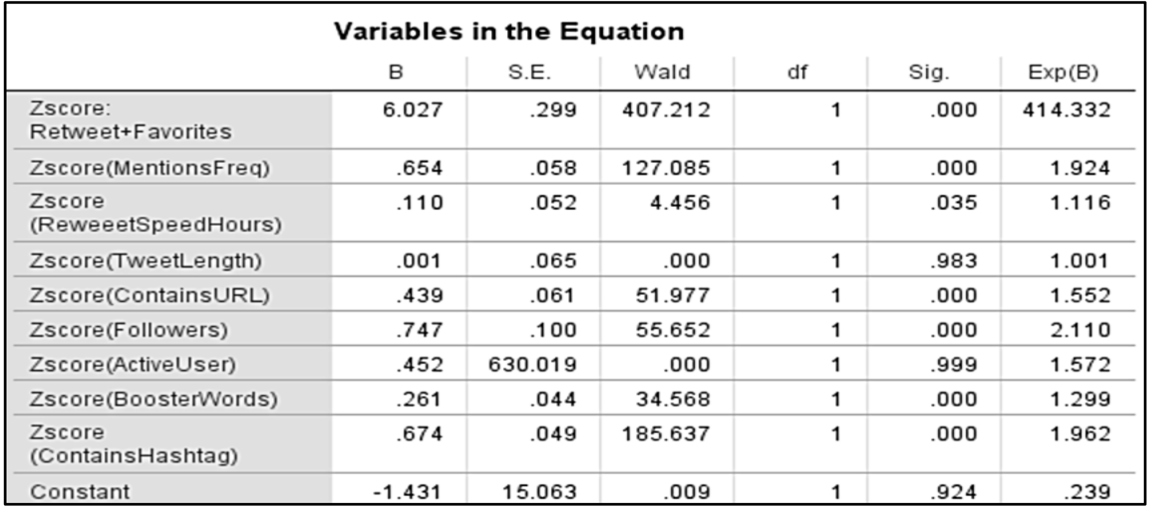


**Coefficients[a]**

| | | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| Model | | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| 1 | (Constant) | 2.050 | .081 | | 25.445 | .000 | | |
| | Retweet+Favorites | .047 | .002 | .279 | 30.879 | .000 | .912 | 1.097 |
| | MentionsFreq | .602 | .023 | .228 | 25.878 | .000 | .953 | 1.049 |
| | ReweeetSpeedHours | .000 | .000 | .079 | 9.129 | .000 | .999 | 1.001 |
| | TweetLength | .001 | .000 | .026 | 2.909 | .004 | .928 | 1.077 |
| | ContainsURL | .796 | .026 | .273 | 30.062 | .000 | .903 | 1.107 |
| | Followers | 5.843E-5 | .000 | .146 | 16.205 | .000 | .919 | 1.088 |
| | ActiveUser | 1.216 | .079 | .132 | 15.300 | .000 | .995 | 1.005 |
| | BoosterWords | .851 | .111 | .066 | 7.670 | .000 | .995 | 1.005 |
| | ContainsHashtag | .969 | .051 | .167 | 19.076 | .000 | .971 | 1.030 |

a. Dependent Variable: LossCredibilityScore

## Loss Credibility Frequency Classification Train and Test Results

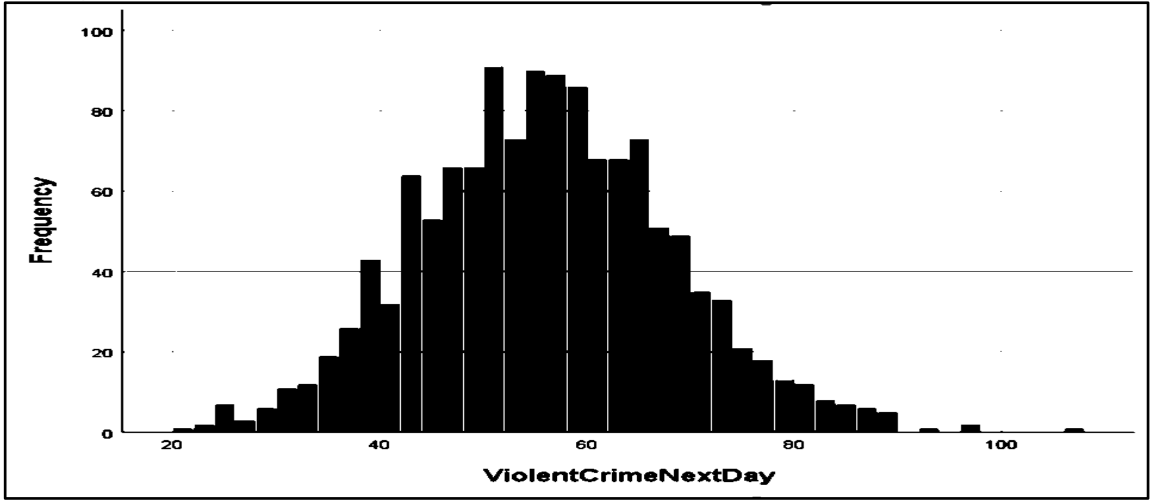


Variables in the Equation

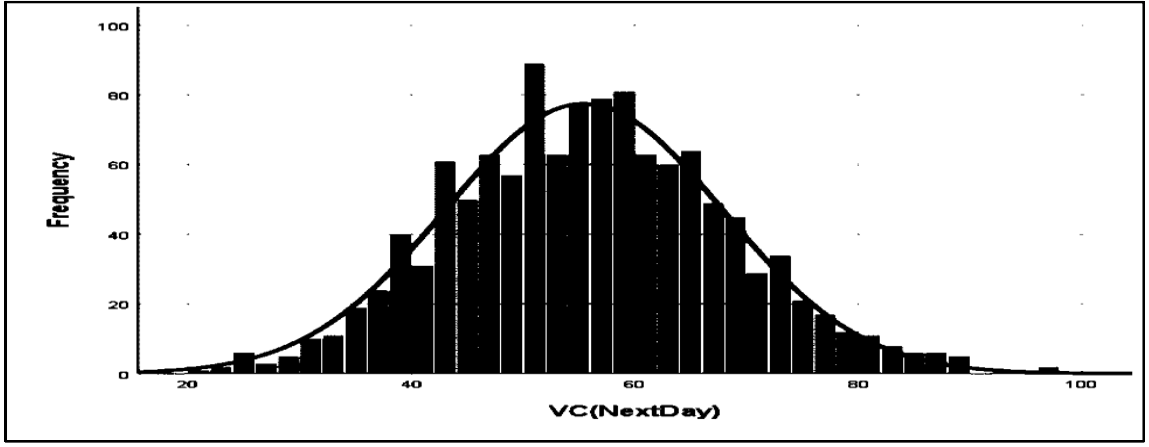
| | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|
| Zscore: Retweet+Favorites | 6.027 | .299 | 407.212 | 1 | .000 | 414.332 |
| Zscore(MentionsFreq) | .654 | .058 | 127.085 | 1 | .000 | 1.924 |
| Zscore (ReweeetSpeedHours) | .110 | .052 | 4.456 | 1 | .035 | 1.116 |
| Zscore(TweetLength) | .001 | .065 | .000 | 1 | .983 | 1.001 |
| Zscore(ContainsURL) | .439 | .061 | 51.977 | 1 | .000 | 1.552 |
| Zscore(Followers) | .747 | .100 | 55.652 | 1 | .000 | 2.110 |
| Zscore(ActiveUser) | .452 | 630.019 | .000 | 1 | .999 | 1.572 |
| Zscore(BoosterWords) | .261 | .044 | 34.568 | 1 | .000 | 1.299 |
| Zscore (ContainsHashtag) | .674 | .049 | 185.637 | 1 | .000 | 1.962 |
| Constant | -1.431 | 15.063 | .009 | 1 | .924 | .239 |

# APPENDIX E: STUDY HISTOGRAMS

## Histogram for Research Model Outcome Variable, Violent Crime (Next Day)



## Histogram for Direct Effects Outcome Variable, Violent Crime (Next Day)


Frequency
VC(NextDay)