

11-6-2019

Twitter Activity Of Urban And Rural Colleges: A Sentiment Analysis Using The Dialogic Loop

Eugene H. Pons

Florida International University, epere222@fiu.edu

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>



Part of the [Communication Technology and New Media Commons](#), [Educational Technology Commons](#), [Mass Communication Commons](#), and the [Social Media Commons](#)

Recommended Citation

Pons, Eugene H., "Twitter Activity Of Urban And Rural Colleges: A Sentiment Analysis Using The Dialogic Loop" (2019). *FIU Electronic Theses and Dissertations*. 4342.
<https://digitalcommons.fiu.edu/etd/4342>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

TWITTER ACTIVITY OF URBAN AND RURAL COLLEGES: A SENTIMENT
ANALYSIS USING THE DIALOGIC LOOP

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CURRICULUM AND INSTRUCTION

by

Eugene H. Pons

2019

To: Dean Michael R. Heithaus
College of Arts, Science and Education

This dissertation, written by Eugene H. Pons and entitled Twitter Activity of Urban and Rural Colleges: A Sentiment Analysis Using the Dialogic Loop, having been approved in respect to style and intellectual content, is referred to you for judgement.

We have read this dissertation and recommend that it be approved.

Leonardo Ferreira

Maria Lovett

Thomas Reio

M. O. Thirunarayanan, Major Professor

Date of Defense: November 6, 2019

The dissertation of Eugene H. Pons is approved.

Dean Michael R. Heithaus
College of Arts, Sciences and Education

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2019

© Copyright 2019 by Eugene H. Pons

All rights reserved.

DEDICATION

This dissertation is dedicated to my mother Angela D. Pons and my aunt Angela M. Cruz, who encouraged me to advance my studies to the point of obtaining a doctoral degree. It is also dedicated to my brother, Alexander Pons, my wife, Jennifer Coccaro-Pons, and my children Sabrina and Nicholas, whose support and motivation pushed me to complete this journey.

ACKNOWLEDGMENTS

I would like to thank my brother, Alexander Pons, for his support through the difficult points of this challenging endeavor. Also my wife, Jennifer Coccaro-Pons, who convinced me to continue my studies and pursue a doctoral degree. They both provided emotional support throughout the entire process, sat through countless hours of discussion, provided extensive feedback, and helped me through the meticulous editing process. Without their invaluable backing and guidance, the completion of this endeavor would have been impossible.

Also, my sincerest appreciation goes to Dr. M. O. Thirunarayanan for always providing valuable input and guidance, Dr. Leonardo Ferriera, Dr. Maria Lovett, and Dr. Thomas Reio for the time spent in reading this dissertation and providing valuable feedback.

ABSTRACT OF THE DISSERTATION
TWITTER ACTIVITY OF URBAN AND RURAL COLLEGES: A SENTIMENT
ANALYSIS USING THE DIALOGIC LOOP

by

Eugene H. Pons

Florida International University, 2019

Miami, Florida

Professor M. O. Thirunarayanan, Major Professor

The purpose of the present study is to ascertain if colleges are achieving their ultimate communication goals of maintaining and attracting students through their microblogging activity, which according to Dialogic Loop Theory, is directly correlated to the use of positive and negative sentiment. The study focused on a cross-section of urban and rural community colleges within the United States to identify the sentiment score of their microblogging activity. The study included a content analysis on the Twitter activity of these colleges. A data-mining process was employed to collect a copious of the tweets associated with these colleges. Further processing was then applied using data linguistic software that removed all irrelevant text, word abbreviations, emoticons, and other Twitter specific classifiers. The resulting data set was then processed through a Multinomial Naive Bayes Classifier, which refers to a probability of word counts in a text. The classifier was trained using a data source of 1.5 million tweets, called Sentiment140 that analyzed the corpus of these tweets, labeling them as positive and negative sentiment. The Multinomial Naive Bayes Classifier distinguished specific wording and phrases from the corpus, comparing the data to a specific database of

sentiment word identifiers. The sentiment analysis process categorized the text as being positive or negative. Finally, statistical analysis was conducted on the outcome of the sentiment analysis.

A significant contribution of the current work was extending Kent and Taylor's (1998) Dialogic Loop Theory, which was designed specifically for identifying the relationship building capabilities of a Web site, to encompass the microblogging concept used in Twitter. Specifically, Dialogic Loop Theory is applied and enhanced to develop a model for social media communication to augment relationship building capabilities, which the current study established as a new form for evaluating Twitter tweets, labeled in the current body of work as Microblog Dialogic Communication. The implication is that by using Microblog Dialogic Communication, a college can address and correct their microblogging sentiment.

The results of the data collected found that rural colleges tweeted more positive sentiment tweets and less negative sentiment tweets when compared to the urban colleges' tweets.

TABLE OF CONTENTS

CHAPTER	PAGE
CHAPTER I: INTRODUCTION.....	2
Pilot Study – Tweeter Sentiment Analysis.....	9
Build Classification Model to perform Sentiment Analysis	11
Analyze the tweets from a college	12
Purpose Statement	14
Background	16
Significance of the Study	20
Research Questions	29
Hypotheses	31
Assumptions	32
Delimitations	32
Summary	33
Definition of Terms.....	34
CHAPTER II. Literature Review.....	37
Dialogic Loop Theory	38
Content Analysis	47
Sentiment Analysis.....	53
Machine Learning Classifiers.....	58
Naive Bayes.....	59
Multinomial Naive Bayes.....	61
Maximum Entropy	61
Support Vector Machines.....	62
Social Media and Microblogging.....	64
Summary	70
CHAPTER III. METHODOLOGY	73
Data Source	75
Data Set - Twitter API - Twitter Data Collection	76
Register Data Extraction Application with Twitter - SpeechBusComm.....	77
SpeechBusComm - Data Extraction Application.....	80
SpeechBusComm – Load CSV Files into Excel	84
Data Preprocessing	84
Feature Extraction	87
Sentiment Classifier	89
Training Data.....	89
Statistical Analysis	91
CHAPTER IV: RESULTS.....	93
Participants	93
Data Set Analysis	96
Model Building Process	97
Unigram Classifier	101

Bigram Classifier.....	103
Combogram Classifier.....	106
Research Analysis	118
Results for Research Question # 1	119
Results for Research Question # 2	121
Results for Research Question #3	123
Results for Research Question #4	127
Results for Research Question # 5	132
URL Component	135
Hashtag Component	137
At Sign Component.....	140
Research Question # 5 Analysis.....	142
Hypothesis.....	143
Summary	144
CHAPTER V: DISCUSSION.....	147
Restatement of the Problem	147
Summary of the Study.....	148
Discussion of the Results	150
Summary and Discussion of Results for Research Question # 1	150
Summary and Discussion of Results for Research Question # 2	150
Summary and Discussion of Results for Research Question # 4	152
Summary and Discussion of Results for Research Question # 5	152
Practical Implications.....	153
Limitations	156
Recommendations for Future Research	157
Conclusion.....	159
REFERENCES	162
VITA.....	175

LIST OF TABLES

TABLE	PAGE
Table 1 Top Eight Universities Using Social Media	9
Table 2 Student Enrollment Totals by Gender	94
Table 3 Ethnic Demographic on Urban Colleges	95
Table 4 Ethnic Demographic on Rural Colleges	96
Table 5 Raw Tweet Acquisition Values	96
Table 6 Confusion Matrix Accuracy Chart.....	99
Table 7 Data Confusion Matrix	100
Table 8 Unigram Sentiment Classifier Data	102
Table 9 Bigram Sentiment Classifier Data	105
Table 10 Combogram Sentiment Classifier Data	107
Table 11 Sentiment Analysis Results by Classifier	109
Table 12 Sentiment Analysis Percentage Results by Classifier.....	110
Table 13 Total Tweet Social Media Components.....	111
Table 14 Total Social Media Component Frequencies.....	112
Table 15 Percentage of Tweet Social Media Components for Unigram Classifier	114
Table 16 Percentage of Tweet Social Media Components for Bigram Classifier	115
Table 17 Percentage of Tweet Social Media Components for Combogram Classifier ..	116
Table 18 Total Twitter Social Media Interaction.....	117
Table 19 Urban Colleges Student Enrollment Totals/Averages by Gender	119
Table 20 Student Demographic Data by Urban College	119
Table 21 Urban Colleges Combogram Classifier Tweet Sentiment Totals/Averages....	121

Table 22 Rural Colleges Student Enrollment Totals/Averages by Gender	121
Table 23 Student Demographic Data by Rural College.....	122
Table 24 Rural Colleges Combogram Classifier Tweet Sentiment Totals/Averages.....	123
Table 25 Urban Rural College Combogram Classifier Average Sentiment Analysis	124
Table 26 Sentiment Type and College Type Variables Cross Tabulation.....	125
Table 27 Urban Retweet Count.....	128
Table 28 Rural Retweet Count.....	129
Table 29 Retweet Count Sentiment and College Type Variables Cross Tabulation	130
Table 30 Aggregate Social Media Components by College Type.....	133
Table 31 Frequency Totals for Component: Likes by College Type.....	134
Table 32 Frequency Totals for Component: URLs by College Type.....	136
Table 33 Frequency Totals for Component: Hashtags (#) by College Type	138
Table 34 Frequency Totals for Component: At Signs (@) by College Type	140

LIST OF FIGURES

FIGURE	PAGE
Figure 1. Sentiment Classification Process Summary Overview.	11
Figure 2. Sentiment Analysis Classification Model Design.	12
Figure 3. Tweet Analysis Process.....	14
Figure 4. Sentiment Classification Process.....	75
Figure 5. Twitter Developer Website.....	79
Figure 6. Twitter Application Creation.....	79
Figure 7. Application Customer Key and Token.....	80
Figure 8. Sample extracted tweets	81
Figure 9. Python Program Code.....	83
Figure 10. Excel Tweet Layout and Composition Sample	84
Figure 11. Sentiment Analysis per College Unigram	103
Figure 12. Sentiment Analysis per College Bigram	105
Figure 13. Sentiment Analysis per College Combogram	108
Figure 14. Demographic Data by Urban College	120
Figure 15. Demographic Data by Rural College	122
Figure 16. Comparison of Urban vs Rural Tweet Sentiment.....	126
Figure 17. Comparison of Urban vs Rural Retweet Sentiment	131
Figure 18. Comparison of Social Media Likes by Aggregate College Types	134
Figure 19. Comparison of Social Media Likes by College.....	135
Figure 20. Comparison of Social Media URLs by Aggregate College Types.....	136
Figure 21. Comparison of Social Media URLs by College	137

Figure 22. Comparison of Social Media Hashtags (#) by Aggregate College Types.....	138
Figure 23. Comparison of Social Media Hashtags (#) by College	139
Figure 24. Comparison of Social Media At Signs (@) by Aggregate College Types....	141
Figure 25. Comparison of Social Media At Signs (@) by College	141
Figure 26. Microblog Dialogic Communication Framework	154
Figure 27. Microblog Dialogic Communication Tweet Example	155

CHAPTER I: INTRODUCTION

The Dialogic Loop Theory attests that an organization must continuously interact effectively and ethically with its public to form long lasting relationships that will benefit the long-term goals of the organization (Kent & Taylor, 2002). The Dialogic Loop concept was furthered by Agozzino (2015) who considered dialogic communication as a means to establish strategic relationships by counting the number of followers a social media microblogging page contained as their interactions. Further research considered Dialogic Loop Theory as a way to establish relationships by conducting two-way communication in the form of comments, likes, following, and reposting (McAllister, 2013; Muckensturm, 2013; Watkins, 2017). Kent and Taylor (1989) designed a dialogic approach, incorporating the Dialogic Loop Theory, to building relationships that was applied to the World Wide Web (WWW), identifying five principles that needed to be integrated into a website to promote the ever-so-desired relationship building between an organization and its public. Several academic studies have been conducted that utilize Kent and Taylor's dialogic approach as the foundation for the research (Agozzino, 2015; Bortree & Seltzer, 2009; Lim & Lee-Won, 2017; Lovejoy, 2012; McAllister-Spooner 2009; McAllister, 2013; Muckensturm, 2013; Rybalko & Seltzer, 2010; Waters & Jamal, 2011; Watkins, 2017). Much has changed in the WWW and the Internet since the late 1990s when Kent and Taylor developed their theory-formulated framework. Primarily, the social media revolution of Facebook and Twitter from 2005 to 2010 dynamically changed the perspective of online interaction and relationship building. Given the change of online interaction and relationship building, there exists a gap in the literature to address to use of Dialogic Loop Theory within the framework of today's social media

environment of microblogging. Specifically, Kent and Taylor's theory-formulated framework using a dialogic approach to building relationships and the five principles of dialogic communication for designing websites have not been accurately tested or modified for the microblogging social media interaction of the current Internet environment. The purpose of the present study is to extend Kent and Taylor's Dialogic Loop Theory to the microblogging concept used in Twitter. The study develops specific standards that examine the relationship between Twitter and Microblog Dialogic Communication, establishing precise principles for relationship building within the microblogging concept of social media.

Although Dialogic Loop Theory has been applied to the relationship building capabilities of websites, there is no significant understanding of dialogic loop applied to microblogging. The current research provides a perspective on the Dialogic Loop Theory and its use to effectively generate relationship building capabilities through microblogging. Relationship building capabilities through microblogging are accomplished through the principles developed for the Microblog Dialogic Communication framework.

Furthermore, the research provides an appropriate way for universities to use Twitter as a means to generate relationship building strategies that further their interactions with potential students, current students, alumni, the community and stakeholders. For example, a university can structure microblog messages that use gender inclusive words, that consider and promote some event occurring at the university, that call for some form of feedback, and that provides a visual intriguing image or video which can trigger sharing and/or tagging of the message. The above mentioned

relationship building strategies originate from the Microblog Dialogic Communication framework and are applied to all communication created in every microblogging activity.

The current study considered if colleges are applying proper Microblog Dialogic Communication to maintain and attract students through their use of Twitter. The current research analyzed a cross-section of urban and rural community colleges within the United States, identifying a sentiment score using the positive and negative sentiment of their microblogging activity. Depending on the sentiment score obtained, a college can determine if their microblogging activity is favorable or not in relationship building aspects. The sentiment score allows colleges to modify their microblogging communication by incorporating the Microblog Dialogic Communication framework presented in the present work to obtain attainable constructive communication goals.

A significant contribution of the present work is the modification of the original Kent and Taylor's (1998) Dialogic Loop Theory to integrate the aspects of today's social media communication as a means to establish positive relationship building components. The expansion of Kent and Taylor's theory to include microblogging activity is labeled in the current body of work as Microblog Dialogic Communication, which allows a college to understand and correct their microblogging sentiment.

The use of microblogging has established a significant way for people, businesses and academic institutions to interact and communicate with their audiences in a direct and efficient manner. The microblogging social media medium, such as Twitter, presents the possibility to engage with an audience by allowing real-time feedback and dialog. The microblogging environment is one of self-expression with the need to connect with other individuals (O'Neil, 2009). Connecting with individuals through microblogging can lead

to relationship building aspects through daily communication interactions. A concern of social media interaction is the use of ethical language communication. The content contained in a college's microblogging activity is intended to reach and interact with its student population, including all individuals pertaining to stakeholders, such as alumni, the community, and future potential students. Specifically, academic institutions must be concerned with ethical issues such as honesty, telling the truth, and the fair treatment of individuals (Austin and Jin, 2015) when engaged in microblogging activities. With regard to ethical communication, microblogging activity for any college can be impacted by various ethical theories.

According to the ancient Greek philosopher Epicurus, the theory of Ethical Egoists consists of the idea that all communication that is expressed should serve to promote one's own interest and perspectives. The theories' concept is that "everyone should promote his or her own self-interest" (Mitsis, 1988, p. 458) and stating untruths would not be of concern if there is no risk of being detected. In the Ethical Egoists aspect, anything said or done, regardless of the truth, is permissible as long as it promotes a self-interest that results in a pleasurable or beneficial outcome.

On the other hand, British philosopher and economist John Stuart Mill presented Utilitarianism as an extension to the Epicurean view. Mill's perspective consists of taking some type of action or expressing something that is bad or good that is dependent on the consequences that will result from the action or statement (Jacobson, 2003). In other words, one must seek the greatest happiness for the greatest number as determined by the potential outcome. In the Utilitarianism aspect, a truth or lie is dependent on the greater good or the consequence of what is said or done.

Finally, German philosopher Immanuel Kant discussed the ethical theory of the Categorical Imperative. The theory considers truth to use the demands of reason and moral law, in determining if an act is good or bad, right or wrong, regardless of the outcome (Bordum, 2002). In the Categorical Imperative perspective, any type of untruth is not accepted.

The ethical implication of what a college tweets or posts, is a complex and difficult issue that most likely displays a combination of the above-mentioned theories. The ethical implication of a college's communication activities are particularly relevant with social media in general since it dominates the way in which people in the new century exchange opinions, ideas, and thoughts. It has transformed communication, the way individuals use verbal, nonverbal, and written messages to express musing across continents. Today's culture expects and demands communication interaction to be immediate, brief, and continuous. It allows users to receive a message, comment and provide feedback or input, and propagate the message by forwarding or linking to it, thus extending the interaction among end users and audiences. Although traditional blogging web sites, which are rooted in the web logs of the mid-1990s where people established online journals or diaries that contained extensive thought, commentary and experiences (Paulus, 2008), provided viable outlets for communication, it lacked the immediacy and spontaneity of microblogging. Microblogging allows individuals to communicate their thoughts in brief blasts of information to friends and followers (Devoe, 2009). A common microblog is Twitter, which allows individuals to share or "tweet" their ideas in no more than 280-characters. Twitter was built on the standard text messaging service concept known primarily as a short message service (SMS), where the standard character limit is

160-characters. Twitter originally allowed a 140-character limit that followed the SMS limit, leaving 20 characters for the username, therefore tweets could stay within the constraints of one text message (Gartner, 2012). As of November 2017, Twitter increased the character limit to 280-characters "so that every person around the world could express themselves easily in a Tweet" (Rosen, 2017, p. 1). Tweets appear as a continuous newsfeed combined with comments from other subscribers. Individuals can also directly respond to the message sender or retweet the message in their own microblog newsfeed to further propagate the message (Mills & Chandra, 2011). Hence, microblogging is part blog and part instant messaging, allowing an individual, business, or academic institution to build or further relationships with its audiences while attracting new audiences. Its appeal stems from two main attributes: it allows easy communication among individuals in succinct manner through the use of brief thoughts intended to keep the reader and audiences engaged, and it allows simple filtering of content most likely to interest the follower (Grove, 2009).

Of particular interest when considering Twitter and its 280-character message, is the sentiment analysis of the message. Specifically, Twitter sentiment analysis is a research process of data-mining words, language, and text analysis using computational linguistics (Hennessy, 2014) to determine if a message is positive, negative, or neutral in regard to a particular topic (Jhaveri et al., 2015). The message attitude or sentiment is of concern, since organizations, such as academic institutions, use microblogging to continuously interact with its audiences or students.

The communication topics promoted by a college establishes an image of what the academic institution represents and promotes. Branding, reputation management, and

creating an academic identity is crucial for recruiting and increasing student enrollment, improving student retention, and fostering an integrative communication platform for a college's stakeholders (Laws, 2014). According to the leading international higher education directory and search engine uniRank, there are 6,415 out of 13,146 colleges and universities or 48.8% that have adopted an official Twitter page (uniRank, 2018) to market higher education and help connect them with their students, staff, alumni, and stakeholders. Colleges and universities use social media platforms to interact with students far beyond the standard of sharing campus photos or school events. The top social media colleges and universities implement the power of branding and digital marketing to attract students while keeping them engaged with informative content throughout their academic journey and beyond (BestColleges, 2018). The top colleges and universities that use social media effectively are Harvard University with 853,000 Twitter followers and over five million Facebook followers, Stanford University with 609,000 Twitter followers and 1.2 million Facebook followers, and Yale University with 404,000 Twitter followers and 1.3 million Facebook followers (See Table 1). These colleges social media activity highlight issues from faculty achievements, pictures of prominent individuals at campus, general campus events, to information on artificial intelligence, healthy eating suggestions, and self-help videos (BestColleges, 2018).

Table 1
Top Eight Universities Using Social Media

Rank	University/College	Twitter Followers	Facebook Followers
1	Harvard University	853,000	Over 5 million
2	Stanford University	609,000	1.2 million
3	Yale University	404,000	1.3 million
4	Massachusetts Institute of Technology	945,000	1.0 million
5	University of Michigan - Ann Arbor	203,000	750,000
6	Ohio State University	307,000	700,000
7	Texas A&M University	258,000	611,808
8	University of Florida	181,000	673,439

Note: Retrieved from <https://www.bestcolleges.com/features/best-college-social-media/>

The messages communicated through microblogging media are an important factor that can shape the enrollment and economic future of colleges. Understanding the sentiment analysis of messages can provide vital measurements and insights to continue or change the communication topics of an institution (Abdelrazeq et al., 2015). Furthermore, student engagement through microblogging can serve as a key performance indicator for a college or university.

Pilot Study – Tweeter Sentiment Analysis

Since the nature of social media, specifically for colleges, is to propagate a message that engages its audience in continued discourse, a plethora of activity or tweets are generated on a regular basis by these institutions. To evaluate expeditiously the large volume of corpus that is obtained through microblogging, it is possible to apply a classification model that has the capability to analyze the sentiment of a college's microblogging activity. A key component of the sentiment analysis process involves

training a machine learning system with previously evaluated tweets to analyze the sentiment of the targeted study, in the present case, the college tweets corpus. Figure 1 illustrates a summary overview of the sentiment classification process, including the machine learning stages used in building the model to applying it to the targeted data for tweet sentiment classification.

As a means to verify the accuracy and functionality of the sentiment analysis process, a pilot study was conducted as part of the research prior to commencing the formal study to evaluate the sentiment analysis process. The pilot study incorporated an existing corpus of tweets that have already been categorized as positive or negative called Sentiment140 to train the classification model. Sentiment140 is a Twitter sentiment analysis tool created by computer science graduate students at Stanford University that contains 1.5 million tweets which have been analyzed for positive and negative sentiment (Kumar, 2014; Sentiment140, 2018). It is a robust sentiment analysis tool used to train a sentiment classifier in identifying or tagging words in a tweet as positive or negative words. Examples on the sentiment lexicon for Sentiment140 of positive words include words such as great, beautiful, nice, good, and honest. Examples for Sentiment140 of negative words include words such as terrible, shame, bad, ugly, and negative (Kiritchenko, et. al. 2014).

The sentiment analysis process consists of two stages 1) build the classification model to perform Sentiment Analysis, and 2) Analyze the tweets from a college.

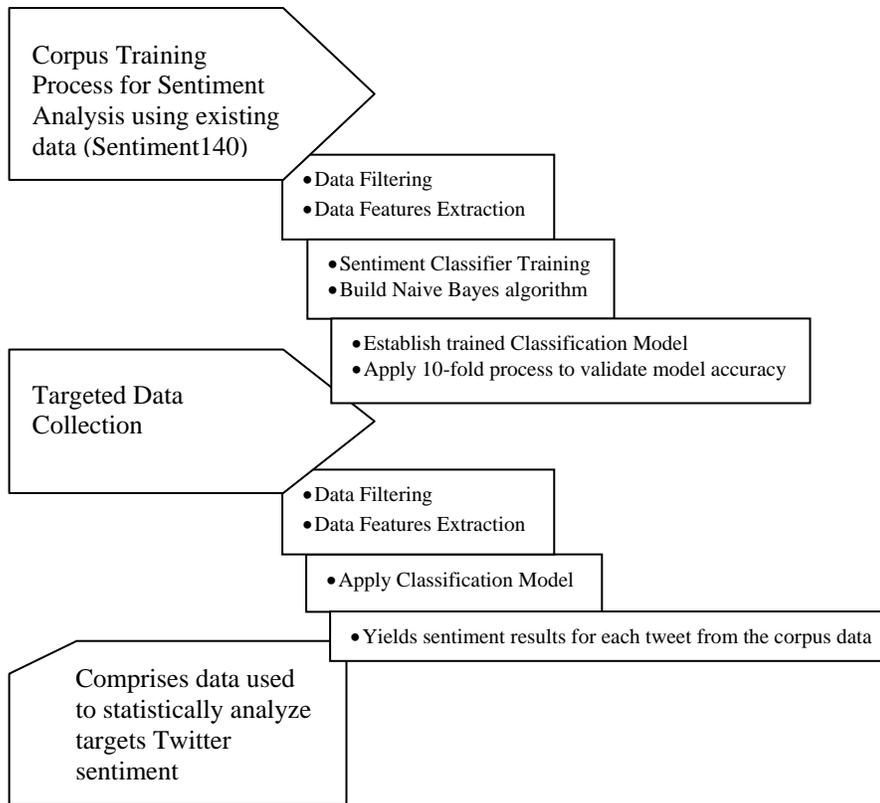


Figure 1. Sentiment Classification Process Summary Overview.

Build Classification Model to perform Sentiment Analysis

The classification model is trained using the Sentiment140 data source, which consists of over 1.5 million tweets that have been tagged as positive or negative. In other words, 1.5 million tweets have been labeled or tagged for sentiment that is used to train the classification models. I have selected the Naïve Bayes classifier as the classification model to build since the academic literature mentions it as the most effective model in performing text classification, which includes tweets. Before using the tweets from Sentiment140, some preprocessing is performed (which was also done for the tweets from the college) that cleans up the tweets.

The data cleansing includes various components that are not necessary for the sentiment analysis process, such as the Twitter tags, markup language, retweets, links, punctuations, abbreviations, stop-words, stemming, and repeated letter spellings. Once the set of tweets has been cleaned, the classification model is then built using the open source scikit-learn Machine Learning library in Python. The result is a classifier that has been trained with over 1.5 million tweets using Naïve Bayes that can be used to determine the sentiment of a set of tweets from any source, but in the present study, tweets from the respective colleges. Figure 2 illustrates the classification model design process, indicating the training and building aspects used to establish the sentiment classification model.

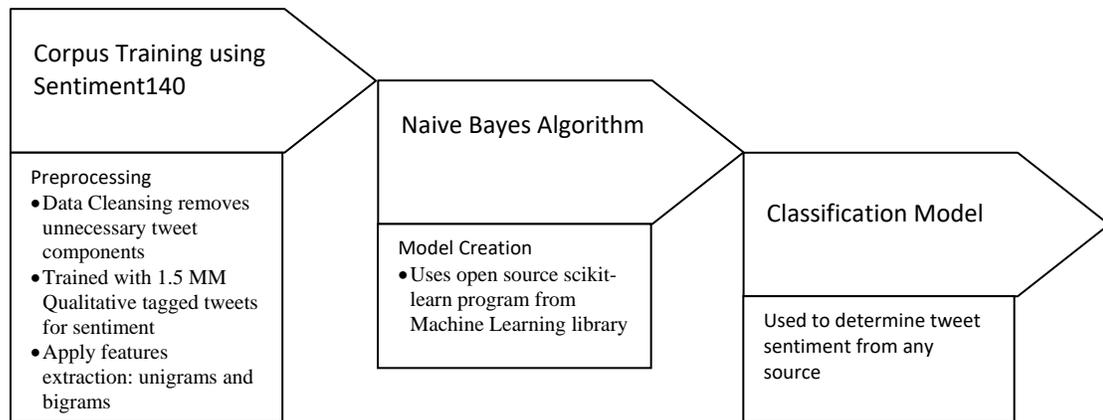


Figure 2. Sentiment Analysis Classification Model Design.

Analyze the tweets from a college

A random college was selected from the search list of colleges presented by the web site Campus Explorer. No specific criterion was used in the selection other than that the college selected had to have an active Twitter page. For the purpose of the pilot study, the college selected was called SampleTest-A. The next step in the process was to capture the tweets that have been posted by SampleTest-A on Twitter.

Twitter provides access to tweets associated with any entity, like a college through its Twitter API (Application Programming Interface). An open source computer software known as Tweepy was used to capture the tweets from Twitter for the specific identifier SampleTest-A. Tweepy extracts all related data source tweets into an Excel spreadsheet. The data-mining process is employed to collect a copious of the tweets associated with the college.

The next step was to prepare the corpus data by cleaning or removing all irrelevant information. The data cleansing included removing various components that were not necessary for the sentiment analysis process, such as the Twitter tags, markup language, retweets, links, punctuations, abbreviations, stop-words, stemming, and repeated letter spellings (the same process was followed during the classification model training). The SampleTest-A corpus was essentially reduced to a set of words that contain the main significance of the tweets.

With the trained model and the set of cleaned tweets from the college, called SampleTest-A, the classification model can then receive each tweet from SampleTest-A and obtain a sentiment for that tweet. As the complete SampleTest-A corpus was submitted to the classification model, corresponding sentiments were collected, which comprised the sentiment classification process that yielded the results used to evaluate the data from the college. Figure 3 illustrates the tweet analysis process, indicating the data processing and applying the trained classification model to obtain the necessary sentiment analysis results.

The SampleTest-A consisted of 6340 tweets. Once these tweets were evaluated through the Naïve Bayes classification model, it was determined that there were 2372

positive, 1945 negative and 2023 neutral tweets. These results comprise the data that were used to statistically analyze tweets across various colleges.

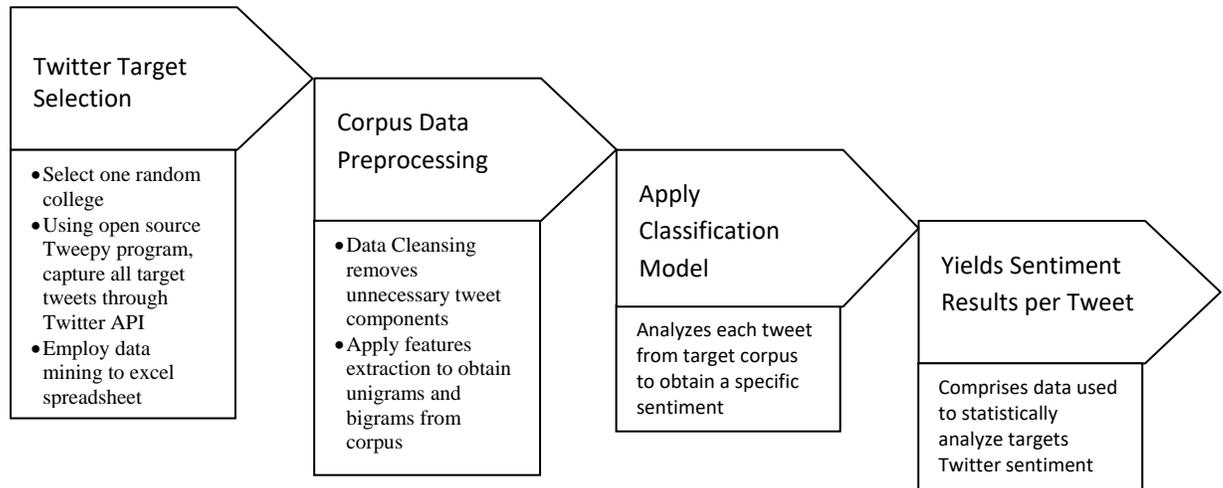


Figure 3. Tweet Analysis Process.

Purpose Statement

A college's microblogging is significant in maintaining student's academic commitments, increased student involvement in the institution's activities, and establishing a sense of student pride and satisfaction for belonging to an institution that furthers their professional values and beliefs after graduation, which enhances alumni involvement and contributions. Since the aforementioned key performance indicators are impacted through the positive or negative perception of its constituents and stakeholders, the assessment of social media sentiment is a key factor that needs to be addressed.

The present study focuses on how a college uses social media, in particular the microblogging application known as Twitter, to communicate messages to its students and stakeholders in a way that fosters relationship building. In particular, the current research considers two principle factors: (a) the microblogging messages of a college are viewed through the lens of sentiment analysis applying data-mining techniques and

linguistic analysis, and (b) Dialogic Loop Theory is considered given the outcome of the sentiment analysis to obtain insight on the relationship building capabilities of messages when interacting with the college audience such as students, parents, and stakeholders.

Colleges and universities communicate messages through the use of Twitter to students regarding a plethora of issues. The microblogging interaction and the sentiment expressed in the messages could have a serious impact on the student's academic perspective in how they perceive the institution, what type of information is communicated, and the relevancy toward their academic future. Colleges and universities need to incorporate special care in using appropriate wording that is well received, that includes supportive language, and that does not connote an unconstructive sentiment. Furthermore, colleges and universities must be aware of the potential negative aspects that can arise from a college's microblogging activity because of the nature of social media. A college or university that utilizes social media can experience negative feedback, derogatory comments, anonymous insults or wording unbecoming of an academic institution, and public ridicule through the sharing, re-tweeting or re-posting, hashtagism, comments, and the propagating of messages, that could minimize the intended purpose of their social media activity.

Regardless, communication has evolved from the spoken word in dyadic interaction to the use of succinct messages broadcast globally via microblogging. The sentiment analysis of these microblog messages are founded on the initial beliefs and concepts of the forms of rhetoric and dialog, and the efficacy of words and messages (McCornack, 2016; McNally, 1970).

Background

The forms of rhetoric and dialog date back to 2200 B.C.E. with possibly the oldest book ever written, the Precepts of Ptah-hotep. The Egyptian sage Ptah-hotep, is known as the earliest-known author and the oldest-known teacher in the world. Ptah-hotep was the teacher of a young prince, the son of King Assa of the Fifth Dynasty, who provided his writings and instruction to explain his philosophy of life (Horne, 1917). McCornack (2016) explains that the Precepts of Ptah-hotep postulated that people should be truthful and kind in their communication, urging active listening skills and most importantly emphasize mindfulness in their choice of words. The art of effective communication was furthered with the notion of rhetoric. Socrates, 469-399 B.C., viewed rhetoric as a practical skill that people could develop through experience, while his student Plato, 427-347 B.C., argued that communication skills could be used in both ethical and unethical ways to manipulate its audiences (McCornack, 2016). Rhetoric henceforth has been defined in various ways, primarily it is the ability to influence others, or at least an attempt at persuading, whether successful or not, through the use of language and visual ability. It is seen as a distinctive communication mode, which can be admirable or deplorable, and is present as soon as one person addresses another (Wardy, 1998). Henceforth rhetoric can be considered the art of adapting discourse, written or spoken communication, to its end (McNally, 1970). The perspective on rhetoric as the ability to influence others is of significance to Dialogic Loop Theory and the aspects of sentiment analysis, since it establishes the foundation for persuasive communication within the microblogging messages used on Twitter by colleges and businesses. Beneficial and ethical communication should be the central focus for all academic

institutions incorporating social media deployment as a means to interact with its audiences or students, parents, alumni, and stakeholders.

The development of the Internet, web 2.0 technologies such as Facebook, Twitter, Second Life, and YouTube, and the popularity and relevance of microblogging have taken the form of traditional rhetoric and communication to new heights and levels of concern. The information shared through these microblogging environments are transient, collaborative, and free-flowing (Mackey, 2011). The concept of the social network within the context of microblogging is defined as a web-located service that allows an individual or organization to interact among a group of individuals that have chosen to follow or accept to receive communication messages attributed to a shared sense of interest. Boyd (2007) defines social network sites as web-based services that allow individuals to (a) construct a public or semi-public profile within a bounded system, (b) communicate with a group of other users with whom they share a connection, and (c) view and traverse their list of connections and those made by others within the system. A 2016 Pew Research Center study found that a majority of Americans receive their news from social media outlets such as Facebook, Twitter, and Reddit (Gottfried & Shearer, 2016), with approximately 71% of Internet users using social networking sites regularly (Number of social media users worldwide from 2010 to 2021 (in billions), 2017). It is estimated that in 2017, there are 2.46 billion social media users worldwide, with 596 million users in China, 208 million users in the United States, and 196 million users in India (Number of social network users in selected countries in 2017 and 2022 (in millions), 2017). The microblog social media sites such as Twitter, rely on rapid and limited communication as the main form of interaction. Microblogs are considered part blog and part instant

messaging, allowing users to share small elements of information in the form of short sentences or phrases, individual images, and video links (Grove, 2009). The main difference of a microblog from a traditional blog is the size of the post, which cannot exceed 280 characters for most microblogging services (Ntaflos, 2013). Some common microblog sites are Twitter, Plurk, Tumblr, Sin Weibo, and Soup io. Of particular interest is Twitter, which accounts for a substantial amount of activity with 357 million users worldwide. Twitter is a free, private microblogging service started in 2007 that allows users to post information or tweet initially within a 140-character limit, currently 280 characters, and allows users to follow other users without reciprocation. In other words, a user can follow any other user, unless they are blocked, and the user being followed does not require a connection or relation to that user. A key component of Twitter is its spontaneous nature that empowers individuals to share and/or augment discussions or threads as fast as they can think about a thought. There are many factors that need to be considered in making a post of such a limited size, such as what will the message communicate, what tone is used in the message, what types of limited words can best communicate the complete message, and how will the message be received. It is imperative that messages are formulated and structured effectively, so that they will be received as they were intended, clearly and with no ambiguity. Clearly structured messages minimize the possibility of a miss-interpretation and increases the understanding and outreach of the communication.

Posted messages have a specific lexicon that includes: retweets, mentions or replies which require a @ (at sign) followed by a user identifier address, that allows users to reply to someone else's tweets and/or allows the message to be sent to any user

regardless if they are followers or being followed; and a categorization method that requires a # (hashtag) followed by a relevant keyword or phrase to make those tweets searchable and to propagate those messages under a specific category or tag (Kwak et al., 2010). The Twitter hashtag element is a public reference that is searchable to include all messages with the similar tag. Twitter users use hashtags to target specific topics within their messages, target specific audiences that are interested in such keywords, to augment the possibility of being retweeted by joining popular Twitter conversations, and to boost follower counts and popularity (Dugan, 2011). Hashtags place the tweets in front of potential followers, resulting in an increase in acceptability and recognition. The categorization of tweets is a relevant attribute for the intent or effectiveness of the message.

The significance of a categorized tweet extends beyond the simple classification of a message. Rather, individuals, businesses, and organizations consider the success or acceptance of a Twitter message as a sought out commodity. Followers view messages that are of interest to their values, beliefs, and general knowledge. These messages need to be relevant, thought provoking and expanding, and convey the proper attitude for maximum effectiveness. The attitude or sentiment of a Twitter message can also influence the perception of existing followers as well as would-be followers. Twitter sentiment is defined as a method that analyzes how opinions, reactions, impressions, emotions and perspectives are expressed in a language (Abdelrazeq et al., 2015). The tweets are essentially categorized as either being a positive or negative sentiment. It is essential to analyze the information generated by an academic organization through its microblogging communication since it defines the social media face of that organization.

It is the digital representation of that organization, the means by which its audience will perceive and evaluate the entity. It will dictate the type of word-of-mouth communication generated by its followers. More importantly, given the explosion of social media microblogging and the ease in which a tweet can be reposted or retweeted and shared, the message can be propagated instantly across the Internet, even reaching the significance of going viral. The term "going viral" has its foundations rooted in viral marketing, which refers to a broad array of online word-of-mouth strategies designed to encourage both online and peer-to-peer communication about a brand, product or service (Eckler & Rodgers, 2010). When a message or tweet goes viral, the tweet becomes so popular that it is shared repeatedly from person to person throughout the Internet. Messages or content that are deemed positive in nature, are more likely to be emailed, shared, or go viral (Berger & Milkman, 2011). Henceforth, it is in the best interest for academic institutions to manage the social media microblogging in a meticulous manner, considering the contents significance on audience's perception and behavior.

Significance of the Study

The relevance of effective communication, sending out or tweeting positive messages or messages that are perceived as positive, is a significant characteristic that colleges and universities need to adhere to in their microblogging activities. Social media microblogging, specifically through Twitter, has the unique quality of interactive communication between the provider of a message and its audience. When a college puts forth a message through a microblog, an exchange of ideas or dialogue commences between the creator of the message, known as the sender in communication models, and its audience or public, known as the receiver in communication models. In the past,

information or messages presented through an academic institution's web page was considered a linear form of communication. In the Linear Communication Model, communication is considered one directional, flowing from sender to a passive receiver (McCornack, 2016). Microblog style communication is applicable to two-way flow of communication, like interactive communication, where both the sender and the receiver are active participants in the communication process. Microblog style communication is analogous to the Transaction Model of Communication, in which "communication is viewed as being reciprocal, with participants continuously engaged in the sending and receiving of messages" (Velentzas & Broni, 2014, p. 127; West & Turner, 2010, p. 11). The microblogging concept facilitates two-way flow of interaction with participants switching between the roles of sender and receiver (Taylor, 2013). Thus, a dialogue is established through the microblog tweets from the academic institution and its audience.

Microblogging dialogic communication engages an organizations audience with the opportunity to analyze messages and provide feedback through comments, and/or forward the message to other individuals through reposting or retweeting. Therefore, Microblog Dialogic Communication facilitates a negotiated exchange of ideas and opinions allowing the public to query an organization and, more importantly, offer the organization an opportunity to respond to comments, questions, and concerns (Kent & Taylor, 1998). These are the foundations to build and maintain relationships with an organization's audience. Colleges need to establish effective Microblog Dialogic Communication to attract students, redefine their relationship with students and the communities they serve, and proactively build ongoing new relationships (McAllister & Spooner, 2008).

Since the desired objectives of institutions are to maintain favorable relationships with its audiences, and microblogging is used to directly interact and propagate messages that are relevant to the college and its followers, it is imperative that colleges apply beneficial and ethical communication through the discourse released on social media. Specifically, colleges need to consider conducting dialogic communication with its public as a means to building relationships (Kent & Taylor, 2002). The form of dialogic communication refers to a two-way exchange; it is both asynchronous and synchronous. Therefore, microblogging on Twitter can obtain a comment or a retweet instantaneously at the moment of the tweet or many hours apart, resulting in a dialogue that can be either asynchronous or synchronous (Muckensturm, 2013). Several other studies have identified a direct association between blogs and relationship-building outcomes attributed to dialogic practices (Kerstetter, 2014). Kent and Taylor (2002) defined five dialogic principles to facilitate relationship building through mediated, two-way, dialogic communication. These principles include "(1) *mutuality*, or the recognition of organization–public relationships, (2) *propinquity*, or the temporality and spontaneity of interactions with publics, (3) *empathy*, or the supportiveness and confirmation of public goals and interests, (4) *risk*, or the willingness to interact with individuals and public on their own terms, and (5) *commitment*, or the extent to which an organization gives itself over to dialogue, interpretation, and understanding in its interactions with publics" (Kent & Taylor, 2002, p. 24).

The notion that dialogue can facilitate relationship building when it is embedded into an organization's Internet interactions originated from a research titled "*Building Dialogic Relationships Through the World Wide Web*" and conducted by Kent & Taylor

(1998). The study yielded five dialogic communication principles which a website must contain for it to promote relationship building. These principles are summarized as follows: 1) The Dialogic Loop - relationship building commences with reliable information offered through the Internet, 2) The Usefulness of Information - organizations must strive to provide valuable information that is relevant to their publics, 3) The Generation of Return Visits - the organizations public needs to be motivated to frequently return to the website for current content, 4) The Intuitiveness/Ease of Interface - having a clear and easy to navigate Web sites, and 5) The Rule of Conservation of Visitors - preserving the visitors of a web site by minimizing external links that drive visitors away (Kent & Taylor, 1998). A college should consider these basic principles when engaging in their microblogging activities, thus facilitating a dialogic loop, a continual exchange of feedback and interaction between the sender and the receiver, to enhance relationship building. Specifically, a college should apply the strategies of the Microblog Dialogic Communication framework identified in the current study when conceptualizing their microblogging activities.

The final aspect of consideration for colleges implementing the dialogic loop concept for the purpose of developing and improving relationships with its students, alumni, community, and stakeholders, is the sentiment of the messages propagated through their microblogging. Messages posted on Twitter, and for that matter, across all social media platforms used by the college, should reflect a positive attitude with choice words and statements that communicate a favorable perception of the college among its existing followers and/or would-be followers. The process of sentiment analysis allows a college to evaluate the overall attitude of its microblogging activities through a

categorization system that designates messages as positive or negative. The sentiment analysis process provides colleges with a performance measurement mechanism for improvement and adjustment plans upon their microblogging, which directly reflects on student considerations and decision-making process for joining or continuing with a college to proceed with their studies (Abdelrazeq et al., 2015). The combined postings in social media microblog sites such as Twitter are central to its audiences' activities as it can impact their behaviors and influence their decision-making process (Beigi et al., 2016). Therefore, the sentiment of a microblog message plays a significant part for establishing a favorable and ethical microblogging activity that is targeted toward improving relationships with all the constituents of a college.

The aforementioned expresses the significance of positive communication when colleges use social media to interact with their audiences, in particular the concept of Microblog Dialogic Communication which requires a deliberate action in structuring messages that are positive in nature within the 280-character constraint for the sole intent of establishing an interaction with followers and would-be followers, directed toward forming an affinity that leads to long-term relationships. As noted in the inception on the use of communication, from the development of the first book by Ptah-hotep and leading to current times, communication comprises a powerful and significant mechanism to influence and establish specific sentiments. Much of the research conducted on positive communication focuses on applying the public relations concept of dialogic theory as developed by Pearson (1989) and further expanded by Kent and Taylor (1998, 2002) to an organization's communication objectives. The theory has been subsequently applied to interactions on the World Wide Web (Kent & Taylor, 1998) using traditional web pages,

and online communication on the Internet and its social media portals (Bortrec & Seltzer, 2009; Kent et al., 2003; Lee, et al., 2014; Rybalko & Seltzer, 2010) such as Facebook and Twitter. Given the significance on the unique capabilities of the medium known as social networking and microblogging, it is imperative to investigate the use of favorable communication through Microblog Dialogic Communication. Specifically, the use of Twitter as a microblogging vehicle to broadcast messages. The Twitter microblogging application limits the postings to 280 characters or one image or a video link, which if not formulated and/or even formatted properly, could impact the conveyance of the actual message to the audience, possibly affecting existing and future relationships.

Applying Kent and Taylor's model for relationship building through the World Wide Web to the microblogging activity on Twitter forces the concept of establishing specific guidelines and rules to social interactions, which are labeled as Microblog Dialogic Communication in the present study. Incorporating Kent and Taylors five dialogic communication principles (1998) as the foundation for Microblog Dialogic Communication gives rise to the following strategies:

1) Social Media Looping - communication generated in the form of tweets or any other social media interaction as a means to interact with the public should cause feedback or a response from the intended audience in the form of comments and/or produce likes or approval of the original message;

2) Information Relevancy - all interactions posted or tweeted by the organization will provide new, relevant, constructive, trustworthy, and engaging content that is intended to inform its public with relevant information and/or entertain its public in a captivating light-hearted manner;

3) Instigating Reaction - posts or tweets must provide communication that is interesting or engaging to the extent that it is propagated in the form of a re-post, re-tweet, or email forwarding to further circulate the original message. Social media, like no other broadcast medium, has the unique characteristic of furthering the delivery of a message to an endless amount of viewers. All communication generated by an organization through microblogging must be created with the intent that it will be re-posted or re-tweeted. Therefore, any message that obtains the qualities of being redistributed by its public has acquired its goal within the social media spectrum;

4) Use of Classifiers and Direct Messages - any content that is posted or tweeted in the social media networks of an organization should adhere to specific levels of accurate information that are straightforward and simple to process. Messages must include appropriate hashtag (#) references which serve to categorize the message within an intended topic. Using appropriate hashtags facilitate search capability and promotes hashtag activism and/or support for social causes. Furthermore, tweets should contain direct message lexicon which require a @ (at sign) followed by a user identifier address, that allows users to reply to someone else's tweets and/or allows the message to be sent to any user;

5) Perceived Informational Authority - communication created through Twitter and other social media outlets should include images and video links that strengthen the message tweeted. Images displayed and video links serve to enhance the intended purpose of the message and further the understanding of the information communicated, often serving as a valuable resource of additional information beyond the original tweet. The use of external links in social media is seen as a means to extend the content of the

tweet, taking full advantage of the social or interactive aspect that social media facilitates. External links also enhance the distribution of the tweet to reach other peripheral followers. The tweet source is considered the information tree root that leads to branches of additional knowledge, while the root continues to be the source of information authority.

An organization or an academic institution or college should consider the above mentioned principles defined by Kent and Taylor and extended in the current study for microblogging as fundamental guidelines when implementing Microblog Dialogic Communication framework. These principles establish the foundation for effective relationship building, not only for creating a Web site, but also for any interactive means of online communication through microblogging and other social media outlets. For the purpose of the present study, the above mentioned Microblog Dialogic Communication strategies are used as a measuring structure on the Twitter data analysis of urban and rural community colleges.

Furthermore, it is necessary to mention the demographic composition of the colleges being analyzed, specifically the ethnicity of the student population. As an example, one of the urban college's student population consists of 69.1% Hispanic/Latino, 15.0% Black/African-American, and 6.0% White (College Factual, 2018), while one of the rural college's student population consists of 67.6% White, 11.2% Hispanic/Latino, and 6.4% Black/African-American (College Data, 2018). Waller, Tiejien-Smith, Davis and Copeland (2008) conducted a study that compared the student demographics by gender and ethnicity composition for urban versus rural community colleges. The study revealed that "town and rural colleges enrolled higher percentages of

white non-Hispanic students than their sister institutions in city and suburban areas. City colleges served higher percentages of black enrollments than their town counterparts. City and suburban institutions were closely linked and indicated higher percentages of Hispanic, Asian/Pacific Islander, unknown, and non-resident alien enrollments than town and rural colleges" (Waller et. al., 2008, p. 5). The study noted that student ethnicity among urban colleges consisted of a diverse student population including a greater portion of minority students compared to rural colleges with more of a homogeneous student population that is primarily white non-Hispanic.

The current study considered the sentiment analysis of a cross-section of urban and rural community colleges, given that urban colleges consist of a diverse student population that includes a greater portion of minority students compared to rural colleges with more of a homogeneous student population that is primarily white American. The study compares these urban and rural college groups independently of each other and comparing between the groups to evaluate if the size of the college has an impact on their Twitter activity.

The intent of the present research is to provide a guide or training mechanism on the ways a college can structure its Microblog Dialogic Communication that promotes a positive sentiment, fostering relationship building aspects. The Microblog Dialogic Communication would enhance colleges' opportunities in improving its image and augmenting student retention and recruiting efforts.

The sentiment expressed through the Microblog Dialogic Communication should be of concern for colleges. Student perception and attitude towards their attending college could be adversely affected if improper Microblog Dialogic Communication is used by

the college. Furthermore, recruitment of future students could also be hindered given the Microblog Dialogic Communication sentiment, specifically since most microblog posts are easily shared with friends or are commented on by recipients, further propagating the message. Finally, alumni, prospective parents, and existing college community participants could be influenced by the Microblog Dialogic Communication if the sentiment is not properly structured to benefit the college's appearance.

Research Questions

The section explains the significance of the dissertation. The first part details the research questions that are addressed by the current work, and the second presents the hypotheses raised by the research questions.

The present study conducted a content analysis comparing the Microblog Dialogic Communication of a cross-section of urban and rural community colleges within the United States. The research performed a sentiment analysis on the microblogging messages of these colleges, using data-mining techniques, linguistic analysis, and Dialogic Loop Theory to measure if the messages are positive or negative, thus providing a perspective on the messages. The significance of the study, as mentioned above, analyzes the microblogging activity of a cross-section of urban and rural community colleges in the United States to see if they are using the Dialogic Loop Theory model of creating positive messages in their Twitter postings to engage their public in a favorably relationship building interaction, which could have a direct impact on student retention and future recruiting.

For the purpose of the present study, the following research questions are considered.

1. What is the microblogging sentiment of an urban college that interacts with a diverse ethnic student population? (Hypothesis 1)
2. What is the microblogging sentiment of a rural college that interacts with a homogeneous student population? (Hypothesis 2)
3. How does the microblogging activity of urban colleges differ from rural colleges with regard to their microblogging sentiment? (Hypothesis 3)

According to Kent and Taylor (2002), it is essential for an organization to perform dialogic communication with its public as a means to building relationships. Therefore, analyzing and knowing if the microblog activity of a college's tweets contain a positive or negative sentiment is an essential component to successfully engage in dialogic loop communication.

4. How does the microblogging activities of a college instigate positive outcomes with regard to the sentiment of retweets and comments? (Hypothesis 4 and 5)

Kerstetter (2014) and Muckensturm (2013) mention that relationship-building is obtained through dialogic practices, which result in an interaction between an organization and its public. The interaction occurs in the form of a retweet, comment, or any form of feedback enacted by the public. Thus, it is necessary to obtain a quantitative analysis on the activity associated with each tweet from the colleges. The analysis furthers the Dialogic Loop Theory of establishing meaningful interaction between an organization and its public (Kent & Taylor, 1998, 2002).

Dialogic Loop Theory contends that positive interactions with an organization's public promote greater relationship-building capabilities (Kerstetter, 2014). To further

consider the relationship-building capabilities of microblogging, the current study conducted a sentiment analysis on the plethora of retweets and comments associated with urban and rural colleges' Twitter activity.

5. How well does the Microblog Dialogic Communication framework explain differences that exist on relationship building capabilities in the Twitter activity between urban colleges and rural colleges? (Hypothesis 6)

The Microblog Dialogic Communication framework establishes specific measures for an organization to produce social media communication, in particular Twitter tweets, in such a way to maximize relationship building capabilities. The current study considered the five elements defined in the Microblog Dialogic Communication framework when analyzing the sentiment analysis of the tweets associated with a cross-section of urban and rural community colleges. The analysis resulted in descriptive statistics comparing the use of Microblog Dialogic Communication within the urban and rural colleges Twitter activity.

Hypotheses

The main hypothesis claimed by my dissertation are as follows:

Hypothesis 1 - An urban college's positive tweets do not surpass its negative tweets.

Hypothesis 2 - A rural college's positive tweets do not surpass its negative tweets.

Hypothesis 3 - The average number of negative tweets among urban colleges does not surpasses the average number of negative tweets of rural colleges.

Hypothesis 4- A college's positive sentiment tweets are not greater than its negative sentiment tweets.

Hypothesis 5 - The total positive sentiment tweets that received a retweet did not exceed the college's negative sentiment tweets that received a retweet.

Hypothesis 6 – The average number of negative tweets of urban colleges do not surpass the average number of negative tweets of rural colleges when analyzed using the Microblog Dialogic Communication framework.

Assumptions

The assumptions of the present content analysis study on the sentiment of college's Twitter activity to establish dialogic communication are as follows: 1) that a college utilizes Twitter and all communication posted on the Twitter microblogging site as a means to formulate an outreach program of information with the intent of establishing credible and reliable interaction with its publics; 2) that a college has knowledgeable employees or department that are responsible for the Twitter postings and conduct themselves in a professional matter; and 3) that a college's public regularly receives, interacts, provides feedback, retweets or reposts, and are active participants of the college's Twitter activity.

Delimitations

The current study is delimited by the following possible issues: 1) additional colleges could have been analyzed beyond the selected college's to obtain a broader perspective of the Twitter dialogic loop activity of academic institutions; 2) other social media activities of the colleges could have been considered, such as referencing the Facebook activity to further consider the dialogic loop communication established outside of the Twitter activity; 3) a questionnaire could be applied to the college's student population to obtain a qualitative analysis on Twitter activity perspectives; and 4) other

data analysis tools could have been applied to the corpus beyond the Naive Bayes Classifier such as Support Vector Machines and Maximum Entropy as alternatives to classifying the Twitter data set for sentiment.

Summary

A college's microblogging activity can have a direct impact on how a college is perceived by its public, including current students, prospective students, alumni, parents, donors, and stakeholders. A sentiment analysis assesses a college's microblogging activity, specifically the dialogic communication or interaction that a college engages to establish favorable relationship communication through its Twitter activities. The objective of the present study is to obtain an understanding of the attitudes and perceptions of a college's public from its Twitter activity and its ability to establish credible relationship building communication. Also, establishing a way to interact in a positive manner on social media such as Twitter by identifying the Microblog Dialogic Communication framework as an extension of Kent and Taylor's (1998) Dialogic Loop Theory for Web sites.

The significance of the study is to analyze the Microblog Dialogic Communication exhibited by a cross-section of urban and rural community colleges within the United States through a sentiment analysis which determines if the Twitter activity is perceived as positive or negative. The current study permits colleges to assess if its Twitter messages are developed in a favorable or positive orientation on behalf of the college, which could have an impact on increased student engagement and support, the college's retention and enrollment, and further future recruiting efforts. The present content analysis study includes a literature review on previous and existing research

conducted in the areas of Dialogic Loop Theory, content analysis, sentiment analysis, and social media and microblogging. The study also includes research questions, assumptions, and delimitations.

Definition of Terms

Content Analysis - A research method that examines patterns in communication artifacts, such as texts, pictures, audio or video (Macnamara, 2003). Neuman (1997) lists content analysis as a research methodology used for gathering and analyzing the content of text such as to words, meanings, pictures, symbols, ideas, themes, or any message that can be communicated in the form of anything written, visual, or spoken that serves as a medium for communication.

Dialogic Loop Theory - A dialogic communication approach to the process of relationship building between an organization and their publics. Kent and Taylor (1998) proposed five dialogic principles which could guide organizations to establish dialogic relationships with public through websites. McAllister (2008) defined it as a theory-based, strategic framework to facilitate relationship with public through the World Wide Web.

Machine Learning - A field of computer science that uses statistical techniques to give computer systems the ability to learn with data, without being explicitly programmed. Machine learning approach makes use of training data set to correctly identify the emotions of each word (Suchdev et al., 2014).

Microblog Dialogic Communication - The present study establishes as a new form for evaluating Twitter tweets, enhancing Dialogic Loop Theory to develop a model for social media communication. The study extends Kent and Taylor's 1998 Dialogic

Loop Theory and its five dialogic principles to a current model for social media communication through microblogging, incorporating five specific microblogging strategies.

Microblogging - A type of broadcasting aspect that is similar to text messaging, that lets users share brief blasts of information (usually in less than 200 characters) to friends and followers from multiple sources including web sites, third-party applications, or mobile devices. Popular microblogging applications include Twitter, friendfeed, Jaiku, Yammer, Tumblr, and Plurk (DeVoe, 2009).

Naive Bayes Classifier - is a machine learning term belonging to a family of simple probabilistic classifiers that apply Bayes' theorem with strong independence assumptions between features (Vidhya & Aghila, 2010).

Sentiment140 - A Twitter sentiment analysis tool that contains 1.5 million tweets which have been analyzed for positive and negative sentiment, used to train the machine learning classifier (Kumar, 2014; Sentiment140, 2018).

Twitter - An online news and social networking service on which users post and interact with messages no bigger than 240 characters long, known as tweets (Kwak, 2010). Twitter is the most widespread microblogging platform on social media, which was created in 2006 and that allows users to send and read text-based messages. It is used by people to express their opinions about a variety of topics (Grosse, 2012).

Twitter API - Also called Twitter application programming interface, is a means in which Twitter provides companies, developers, and users with programmatic access to Twitter data. Twitter API allows developers to look up tweets containing a specific word or a phrase, limited by Twitter to 1500 tweets at a time (Jhaveri et al., 2015).

Uniform Resource Locator - Also known as a URL or a web address, the uniform resource locator is a reference to the location of a web resource on a specific computer network, rather than identifying the resource by name or by some other attribute(s) of that resource (Berners-Lee et al., 1998).

CHAPTER II. LITERATURE REVIEW

Microblog Dialogic Communication refers to the concept of generating concise messages using positive oriented word choices with the intent of formulating lasting online favorable relationships with a user's public through standard microblogging applications such as Twitter. The complexity of the microblogging method is the need to generate a message within specific communication constraints. Primarily, microblogging places a restriction on the quantity of characters used in generating a message, such as Twitter that limits a posting to 280-characters. The character limitation places tremendous responsibility on the sender of a message to utilize specific selective words that convey the proper message with the adequate sentiment. The spontaneous nature of microblogging, such that senders of messages are constantly interacting with their audiences, posting information and replying to comments or re-posting of other information, further complicates the Microblog Dialogic Communication concept. For the purpose of effective Microblog Dialogic Communication, the sentiment of all these types of messages need to be considered prior to sending them out through various social media channels, specifically Twitter.

A microblogger user's image and reputation can severely be impacted negatively if the proper Microblog Dialogic Communication isn't considered. Once a message is thrust into the ever-reaching space of social media, its content is seen, shared, and commented on by all followers and any other subsequent re-posted followers of that message. Sharing, commenting, and re-posting is an ongoing chain reaction of cumulative events triggered from the original post. A misguided or improperly word chosen message can have substantial repercussions against the creator of the message.

The issue should be of great concern for colleges, since their microblogging and social media activities are critical to maintaining their student populations engaged in college-wide activities, attitudes, and values. Furthermore, a college's microblogging activities have tremendous outreach and influence on student retention, student recruitment efforts, and alumni relations. The microblogging activity can also influence ongoing future financial support, parent attitudes and awareness on the college, community involvement, and overall stakeholder interest and support. The image set forth through a college's Microblog Dialogic Communication determines the long-term relationships that the college will have with all of its constituents.

The current research serves as a point of consideration on the sentiment that colleges incorporate into their microblogging activities with the intent of peering through a social media window, specifically Twitter, considering the attitudes expressed in these activities. Initially, a greater understanding of the Dialogic Loop Theory is covered to augment the significance of relationship importance through social media. An understanding of content analysis research is presented to substantiate the methodological approach incorporated for the purpose of the present study. The concept of Sentiment Analysis is also discussed to provide an understanding on the issues of positive and negative messages when sending out posts through microblogging applications. Finally, a detailed analysis on Social Media and Microblogging is also included, providing in-depth details on this type of communication medium.

Dialogic Loop Theory

Dialogic Loop Theory or dialogic communication is a public relations concept for establishing strategic relationships through the use of continuous interaction. The main

premise for Dialogic Loop Theory is that an organization needs to effectively and ethically interact with its publics, establishing two-way communication in a manner that is honest, truthful, and trustworthy. Two-way communication fosters meaningful relationships that will garnish prolonged and reliable interactions, which are beneficial towards an organization's long-term goals. The concept of dialogue theory is attributed to Pearson's 1989 unpublished dissertation "*A Theory of Public Relations Ethics*." Pearson presented the idea that a managed interpersonal dialectic is necessary, in which an ethical dialogic system is mandated over monologic policies (Kent & Taylor, 2002). An ethical dialogic system is an open and honest communication type that is void of manipulation, disconfirmation, or exclusion, and that is continuous with the intent of establishing solid relationships through interactive communication.

Kent and Taylor (1989) proposed a theory-formulated framework using a dialogic approach to building relationships through the World Wide Web (WWW). They viewed the Internet, formerly known as the WWW, as a multi-channel environment that facilitates interactive communication that can be used to improve relationships among the various participants in the process. Interactive communication is the essence of dialogic communication, defined as a relationship building process between an organization and its public (Kent & Taylor, 1998). The concept of dialogic communication is further described as a relational interaction when there is an existing relationship, considering the attitudes held by all the participants, in a communication transaction. Furthermore, Kent and Taylor (1998) state that a dialogue must be created between the organization and its public to formulate effective dialogic relationships. The Internet and online environments such as microblogging sites provide organizations the opportunity to create dynamic and

lasting relationships by incorporating dialogic loop concepts into their Internet communication strategies.

The researchers assert that "relationships between public and organizations can be created, adapted, and changed" through the Internet (Kent & Taylor, 1998, p. 326). They identified five principles that need to be integrated into website development with a dialogic communication approach that promotes relationship building. These principles are as follows:

Principle One: The Dialogic Loop - The aspect of relationship building commences with reliable information offered through the Internet presence of an organization. Relationship building includes a firm commitment of resources and individuals to adequately and timely respond to questions, concerns and problems. Kent and Taylor (1998) state that for dialogic communication to occur, an organization must dedicate specific well-trained personnel to correctly respond and interact efficiently with its public through the various online media such as email, social media, and microblogging.

Principle Two: The Usefulness of Information - Kent and Taylor (1998, p. 327) emphasize that "content is what should drive an effective Web site, and not the smoke and mirrors and fancy graphics." Thus, for the purpose of dialogic communication, organizations must strive to provide valuable information that is relevant to their publics, thereby cultivating a relationship by addressing the interests, values, and concerns of their public (Kent & Taylor, 1998).

Principle Three: The Generation of Return Visits - Dialogic relationships are enhanced through updated information, featured content, interactive forums, and question

and answer sessions. These features maintain a Web site dynamic, giving off the need for visitors or the organizations public to frequently return for current content (Kent & Taylor, 1998).

Principle Four: The Intuitiveness/Ease of Interface - Kent and Taylor (1998) stress the significance of having clear and easy to navigate Web sites, with substantial relevant content that is informational, and containing minimal graphics. They mention that "text loads faster than graphics, and well typeset pages can actually be more effective attention getters than a graphic" (Kent & Taylor, 1998, p. 329). For the purpose of dialogic communication, the speed and ease in which information and a web site can be navigated, are key factors in formulating positive online relationships

Principle Five: The Rule of Conservation of Visitors - The final principle focuses on preserving the visitors of a web site by minimizing third-party links such as advertisements that can drive visitors away from the organizations site. Once a visitor leaves the site through another link he or she may never come back. Kent and Taylor (1998, p. 330) indicate that "if the goal of public relation is to create and foster relationships with publics, Web site should include only essential links with clearly marked paths for visitors to return."

An organization or an academic institution or college should consider the above mentioned principles defined by Kent and Taylor as fundamental guidelines when considering Microblog Dialogic Communication. These principles establish the foundation for effective relationship building, not only for creating a Web site, but also for any all interactive means of online communication including microblogging.

Agozzino (2015) considered the role of dialogic communication in regard to the most followed organizations that use a microblogging digital bulletin board known as Pinterest. The current research notes that the primary function of public relations is to establish strategic relationships with an organization's stakeholders by using two-way communication. Agozzino study on Pinterest, which is a microblogging site that consists of images rather than text, further defined dialogic loop as a discussion between an organization and their online users. The dialogic loop relationship aspect originated from the "opportunity for users to comment, opportunity for users to 'like' pins, and an opportunity for users to engage with pins through repining, posting, playing games, voting on pictures, and participating in contests" (Agozzino, 2015, p. 8).

McAllister (2013) conducted a content analysis study on the status of dialogic relationship-building tactics incorporated by community colleges in New Jersey through their respective Web sites. For the purpose of the present study, dialogic communication was considered as a means to relationship building within the practice of fundraising for nonprofit organizations and higher educational nonprofit settings. The study considers two-way symmetrical or dialogic model as the preferred communication strategy for nonprofits to "manage conflict, improve understanding, and build relationships with publics" (McAllister, 2013, p. 264). The research notes that dialogic communication fosters transparency, public involvement, trust, and understanding. These characteristics are aspects critical to the relationship building process that colleges need to implement with their Microblog Dialogic Communication objectives. McAllister stressed that dialogic communication involves intricate dedication and work, which is both time consuming and laborious. Although, the results can produce "greater organizational

rewards, increased public support, enhanced image/reputation, and decreased governmental interference" (McAllister, 2013, p. 265).

In another study, Muckensturm conducted a content analysis on the use of dialogic communication by the accommodation sector, which refers to the hospitality and tourism industry, specifically through the social media activity on Facebook. The author contends that it is essential to consider dialogic communication as it relates to social media since "social networks have the same amount of influence on consumers as television and more influence than newspapers" (Muckensturm, 2013, pg 3).

Muckensturm (2013) mentioned that dialogue is two-way communication that can be both asynchronous and synchronous, as it relates to blogging, since an interactive communication can occur at the moment a post is made or many hours after the post was made. Muckensturm notes that dialogue is an important aspect of an organizations customer relationship management, and that the accommodation sector's social media activity on Facebook needs to employ the principles of dialogic communication for effective interaction with its public.

Watkins (2017) used a content analysis to determine if dialogic communication increased relations, interaction, and positive attitudes between professional athletes and their public. The study considered the five principles defined by Kent and Taylor (1998) through the social media microblogging site known as Twitter. The author contends that two-way communication refers to the communication between an organization and its public, both of which are seeking mutual benefits. Furthermore, Watkins defines dialogue as an outcome that is generated by the interaction between an organization and its public. Thus, social media provides the communication outlet that facilitates interactivity, while

the relationship building interaction is obtained through the dialogue created in the microblogging process (Watkins, 2017). The research noted that despite there being limited dialogic communication, the process of posting tweets with useful information (one-way communication) served as a viable means in engaging the public in relationship building activities.

Much of the research conducted on dialogic communication and the opportunities that exist on Web sites and social media venues focus on the ample possibilities for relationship building, as noted by the previously mentioned studies. The research is mostly founded on the relationship building potential established by Kent and Taylor, which uses five dialogic principles as a guideline to foster dialogic communication between an organization and its public (Kent & Taylor, 1998). Other researchers have conducted studies on Web sites and social media activity, specifically Twitter and Facebook, for the sole purpose of analyzing the effectiveness of an organizations ability to generate two-way communication through these online components (Agozzino, 2015; Bortree & Seltzer, 2009; Lim & Lee-Won, 2017; Lovejoy, 2012; McAllister-Spooner 2009; McAllister, 2013; Muckensturm, 2013; Rybalko & Seltzer, 2010; Waters & Jamal, 2011; Watkins, 2017).

McAllister-Spooner (2009) conducted a ten-year survey on Kent and Taylor's dialogic Internet principles and its ability to extend dialogic theory to relationship building via Web sites. McAllister-Spooner's extensive research suggests that Web sites underutilize dialogic theory for relationship building functions by limiting the interactive potential of the Internet. Furthermore, Bortree and Seltzer (2009) noted that "organizations are not adequately utilizing the Web to generate dialogic communication;

there are gaps between organizational relationship-building goals, implementation of online strategy, and actual dialogic engagement" (p. 317). They considered social media to conduct a content analysis on 50 Facebook profiles to determine to what degree advocacy organizations effectively used dialogic strategies that generated actual dialogue between the organization and its stakeholders. The study noted that most of the advocacy organizations use social media to inform its public, rather than taking advantage of the potential relationship building dialogic strategies present in social networking. The findings by Bortree and Seltzer on the effectiveness of dialogic communication on social media were similar to the findings by the McAllister-Spooner study on Web site dialogic communication.

Another study conducted by Waters and Jamal, focused on how nonprofit organizations were using microblogging to communicate with their public. A content analysis was conducted on the Twitter tweets posted by 81 nonprofit organizations. The study found that nonprofit organizations were more likely to engage in asymmetrical communication, primarily using Twitter to convey one-way messages. Waters and Jamal (2011) state that nonprofit organizations are not benefiting from the "interactive nature and dialogic capabilities of the social media service, rather using Twitter as a means of sharing information instead of relationship building" (p. 323). A similar study was also conducted by Lovejoy (2012) in which a content analysis of 73 nonprofit organization's Twitter feeds were analyzed for dialogic communication. The study found that nonprofit organizations are not using Twitter to maximize stakeholder involvement, rather they use social media as a one-way communication channel.

Lim and Lee-Won (2017) recognized that research showed that "organizations tend to use Twitter primarily in a one-way, monologic manner" (p. 422) limiting the dialogic communication component of the microblogging capabilities. Thus, the authors conducted a study to investigate the persuasive effect of an organization's dialogic retweeting (the reposting of other Twitter users' messages). The study found that when an organization engaged in dialogic retweets, it expanded the conversation to additional publics, generated favorable attitudes toward the organization, and extended the organization's dialogic loop (Lim & Lee-Won, 2017).

Rybalko and Seltzer conducted a content analysis on the Twitter profiles of various Fortune 500 companies for the use of dialogic features. The research attempted to find an understanding of the dialogic potential of the microblogging site Twitter, noting that online communication was a perfect venue for organizations to nurture dialogue (Rybalko & Seltzer, 2010). The authors mention that organizations are not using Twitter for dialogic communication at its full potential. Rybalko and Seltzer's findings are similar to the findings of other studies which noted that organizations use social media to communicate through one-way messages and with limited dialogic communication activities (Bortree & Seltzer, 2009; Lim & Lee-Won, 2017; Lovejoy, 2012; McAllister-Spooner, 2009; Waters & Jamal, 2011). Although, the research revealed that "corporations seem to use their Twitter feeds more effectively than their websites in facilitating a dialogic loop" (Rybalko & Seltzer, 2010, p. 340). The findings also seem to indicate that organizations use the dialogic feature of conservation of visitors, or an attempt to encourage its public to stay or revisit a web site often, through Twitter to develop dialogic communication. Organizations attempted to keep users engaged by

frequently posting new tweets within 24 hours, thereby fostering relationship building behavior with its public.

Dialogic communication has been defined by various studies to mean the ethical interaction between an organization and its public that is honest, truthful, and trustworthy with the intent to foster prolonged relationships. The ethical interaction concept was considered in several studies within various social media outlets such as Pinterest (Agozzino, 2015), Internet web sites (McAllister-Spooner 2009; McAllister, 2013), Facebook (Bortree & Seltzer, 2009; Muckensturm, 2013), and Twitter (Lim & Lee-Won, 2017; Lovejoy, 2012; Rybalko & Seltzer, 2010; Waters & Jamal, 2011; Watkins, 2017) among others. Each of these studies has used similar methodologies, primarily content analysis, and has found some common results in regard to an organization's use of social media when applying dialogic communication for the purpose of relationship building.

Content Analysis

The research methodology used in the present study is a quantitative content analysis process that evaluates the overall attitude of a message or of the sender of the message on the microblogging site Twitter. The process is known as a sentiment analysis, which evaluates the specific words and phrases of a corpus. The term content analysis refers to a research process used to study a body of text for its use of words and phrases. Researchers use a content analysis to evaluate the intent and meaning of corpus by labeling and categorizing specific words and phrases. These words and phrases are coded following predetermined criteria for a semantic relation. The results are statistically tabulated to obtain a quantitative outcome on the specific use of words and or phrases as determined by the study (Rose et al., 2015). Content analysis yields significant and

valuable results when applied to the microblogging activity on Twitter. The relational analysis component of a quantitative content analysis is an ideal process for the objective of evaluating the use of a college's microblog activity on Twitter in regard to it communicating positive or negative messages to establish favorable relationship outcomes.

The present study conducts a quantitative analysis on the microblogging activity of a cross-section of urban and rural community colleges within the United States. A significant aspect of the current research is the application of machine learning, which is trained using a data source of 1.5 million tweets, called Sentiment140 that has analyzed the corpus of these tweets labeling them as positive and negative sentiment. The content analysis component, the use of Sentiment140 for the quantitative research used in the current study, is a common denominator across other studies in the sentiment analysis field, which allows the current study and future work to be linearly compared with other similar studies on sentiment analysis.

The technique of content analysis has existed since 1927, when introduced by Harold Lasswell to study propaganda (Newbold et al., 2002). It has been used to study text from a broad range of data sources from transcripts of interviews and discussions to the narrative and form of films, TV programs and the editorial and advertising content of newspapers and magazines. Media content analysis has been a primary research methodology to understand and study social issues in television and film (Macnamara, 2003). During the 1920s and 1930s, it became very popular to investigate the rapidly expanding communication content of movies. During the 1950s the investigation migrated toward the study of mass communication, and it continues today as a means to

examine content of all forms including Internet communication (Chaffee & Metzger, 2001; Macnamara, 2003).

Content analysis operates on the notion that verbal behavior is a part of human behavior and that symbols are a part of events, thus it is a technique that describes with objectivity and simplification, what is said on a given subject in a given place at a given time (Lasswell et al., 1952). A content analysis is a systematic analysis that classifies parts of a text through a structured, methodical coding scheme from which conclusions can be drawn about the message (Rose et al., 2015). A content analysis can be conducted on different types of corpora, including text from speeches, letters or articles whether digital or in print, as well as social media text such as from the microblog site Twitter. Content analysis is also applicable to text in the form of pictures, video, film or any other visual media. "Content analysis provides a structured way of analyzing data that are typically open-ended and relatively unstructured" (Rose et al., 2015, p. 195). The aim of a content analysis is to describe the features of the message content, such as in the study conducted by Adams and McCorkindale (2013) in which a content analysis was conducted on the Twitter pages of the 2012 presidential candidates to determine the candidates use of Twitter, if meaningful dialogue was done, and the level of transparency. Another example is Waters and Jamal (2011) content analysis study on how nonprofit organizations use Twitter. In their research, randomly selected organizations that use Twitter were analyzed by acquiring the text data for a specific period of time. Additional information was obtained from Twitter such as the number of followers and quantity of tweets. The tweets for each organization analyzed were then coded by two-coders to conduct statistical study and generate a description of the text.

A content analysis researcher uses an objective and systematic study of counting and recording events to produce a quantitative description of the symbolic content in a text (Neuman, 1997). A quantitative content analysis collects data from a large corpus such as from a newspaper publication, broadcast television or movie, and/or postings from a microblog site, determined by key words in context, and consider the circulation, frequency, and volume of the corpus to analyze. Quantitative content analysis analyzes the positive and negative words and phrases of the corpus to identify the tone, which is then processed using statistical methods in a quantitative approach. Qualitative content analysis examines the relationship between the text and its likely audience-derived meaning, recognizing that a text is polysemic or that it can have multiple different meanings to different audiences. Therefore, a qualitative content analysis tries to determine the likely meaning of texts to audiences. It pays attention to audience and contextual factors – not simply the text. Qualitative analysis of texts is necessary to understand the semantics of the information and audience interpretations, which is the paramount objective of a content analysis. Therefore, researchers and academics view quantitative content analysis as being a part of analyzing texts to discern the text's likely meanings to and impact on audiences (Curran, 2002; Gauntlett, 2002; Newbold et al., 2002).

The design aspect of a content analysis research involves several key stages. Rose, Spinks and Canhoto (2015) present a detailed layout of the process, which includes, "when appropriate, the formulation of a hypothesis in response to a research question" (p. 3). The next step involves identifying and obtaining the material to be analyzed, in other words, what is the source for the corpus material. To conduct a reliable

content analysis of a material, a sampling method is necessary. The process of sampling to conduct a content analysis is listed as: (a) consisting of selecting the media form to be analyzed, such as print, radio, television, film, blog, or microblog; (b) selection of a specific period, such as issues or dates to be analyzed; and (c) sampling of appropriate and significant content from within those media (Newbold et al., 2002). The most appealing method of sampling content for analysis is to conduct a census, which selects all possible items to measure. A census is preferred since it provides the most comprehensive representation of the items under consideration, but it is often impractical or not possible since a very large volume of corpus or media has to be analyzed consisting of a study over a long range period of many months or years. When a census is not possible, a copious or sample of the text/media to analyze must be selected as a representative sample of the overall population of items. The manner in which sampling is performed requires that items be selected in an objective and unbiased manner to ensure the reliability of the results obtained from the study. There are several methods of sampling for content analysis. Random sampling is a process of selecting every n th unit from the total population; while purposive sampling is a process that selects from key media only and not from less important media (Newbold et al., 2002). Other methods are quota sampling in which a selection of a proportion of articles from several regions or areas of either geographic, demographic, psychographic, or subject categories; and stratified composite sampling where a process of randomly selecting units for analysis is conducted from certain days or weeks over a period (Newbold et al., 2002).

The next step in the content analysis process is to decide what will be coded from the corpus material. The coded material can include words, phrases, sentences,

paragraphs, whole documents, images, audio, or visual data. As part of the coding process, a coding scheme is developed, or the process of establishing classification rules to assign the body of text. The coding scheme specifies the systematic manner in which a corpus is classified, which allows for future replicable coding of the data. It details how and what will be coded, assuring that each category is exhaustive and mutually exclusive (Rose et al., 2015). Once a coding scheme has been developed, multiple coders can be used to conduct the coding or textual data can be coded by using computer software programs that support text analysis. The results are then analyzed with statistical methods such as descriptive statistics or inferential statistics to "summarize the findings and /or test any hypothesis that have been formulated" (Rose et al., 2015, p. 5).

The above mentioned process defines the various steps taken in conducting a typical quantitative content analysis. The process is used to analyze large sets of documents or corpus with the objective of testing theoretical issues and obtaining a greater understanding of the data. It is a research method for "making valid inferences from data, attaining a condensed and broad description of the corpus, with the outcome serving as a categorized description of the issue being observed" (Elo & Kyngas, 2008, p. 108). As such, a quantitative content analysis can be applied to a wide variety of text with the purpose of obtaining a greater understanding of the material being observed. Specifically, a type of content analysis, called sentiment analysis, can be used to analyze the type of emotion expressed in the microblog site Twitter. For the purpose of the present study, a sentiment analysis is conducted on the Twitter feed of a cross-section of urban and rural community colleges. The process is described in the following section.

Sentiment Analysis

Sentiment analysis refers to a process whereby specific words and phrases are systematically evaluated for their intent (Beigi et al., 2016). Sentiment analysis attempts to determine the overall attitude of a message or of the sender of the message within the contextual framework of the communication. The analysis criteria are derived from specific word choices within a message that can either make a message appear to be positive or negative in regard to the attitude expressed. The analysis process is typically conducted by using data analytics tools and software that evaluate specific words in a large database of messages to identify patterns and establish word associations. The evaluation provides an insight into the communication style of a message, surmising the message attitude being communicated. Sentiment analysis is a widely used process that provides a systematic method of examining message content for its attitude, emotion, or opinion. Many businesses and organizations incorporate sentiment analysis to evaluate customer attitudes toward a product line or marketing objective, also applying it to examining social media interactions on Facebook or Twitter. The sentiment expressed through social media networks can have a direct impact on the customer relations with a business or organization. It is imperative that businesses and organizations regulate the types of messages communicated through social media networks such as Twitter to establish favorable relationships with its stakeholders.

The term sentiment analysis was defined by Nasukawa and Yi (2003) as a "technique to detect favorable and unfavorable opinions toward specific subjects within large numbers of documents" (p. 1). The authors create a distinction from previous work on sentiment analysis, which focused on identification of sentiment expressions and their

polarities, and their application of sentiment analysis. Primarily, past work considered the sentiment expressions applied to a document as a whole, while the authors considered the relationships between the sentiment expression and the subject. Therefore, specific text fragments are analyzed to identify a sentiment on a subject within a document (Nasukawa & Yi, 2003). The authors mention the importance for organizations to detect and analyze favorability within online documents since they often can influence public opinion and impact relationships.

Beigi, Hu, Maciejewski, and Liu (2016) classified sentiment analysis as an area of study that includes "sentiment classification, opinion mining, subjectivity analysis, polarity classification, and affect analysis" (p. 3). The term is further defined as a process used to extract subjective information from a given text for the purpose of analyzing its sentiments, attitudes, emotions and opinions using natural language processing and data analytics. The authors note that sentiment is categorized into positive or negative categories for the purpose of business intelligence, politics, sociology and the study of human opinions and attitudes (Beigi et al., 2016). The analysis can be applied to social networking sites, microblogs, Web applications, and videos since all express opinions on various topics and events (p. 1).

Jhavaeri, Chaudhari, and Kurup (2015) conducted a study on the sentiment analysis of two significant e-commerce websites in India by extracting the sentiments from their Twitter feed. The authors defined sentiment analysis as "a line of research that harnesses people's opinions and attitude in relation to different topics, products, events and attributes" (p. 14). The analysis considered the "opinionated word or phrase in the text and labeling it as positive, negative or neutral" (p. 14). The research was conducted

on Twitter because of 1) the number of messages that are shared on a daily basis, 2) the synchronous communication aspect of Twitter tweets, and 3) the 280 character limitation of Twitter that stresses the importance of using specific words to convey a message (Jhavaeri et al., 2015). The above mentioned reasons make Twitter a perfect candidate for sentiment analysis due to the unique nature of the messages and the ability to apply data analytics and natural language processing (NLP), or the processing of large structured set of texts.

In another study conducted to analyze social media content as an evaluation tool for universities, the aspect of an opinion derived from communication medium, such as Twitter, is an essential component. "Twitter is one of the most opinion-rich resources, where huge amounts of opinions on different topics are expressed" (Abdelrazeq et al., 2015, p. 49). The researchers used sentiment analysis tools to analyze the social media content of nine universities in Germany since the sentiment expressed through the microblog activities on Twitter can support university rankings, provide universities with a performance measurement mechanism for improvements, and facilitate student information on deciding to attend or choosing a specific university for their academic studies (Abdelrazeq et al., 2015). They define sentiment analysis as "a method used to analyze opinions, reactions, impressions, emotions and perspectives" (p. 52). A search routine and filter called Twitter API was used to extract the specific tweets associated with the search criterion established for the study. The search resulted in a data set of tweets matching the used keyword list. The data set of tweets were then passed through a filtering process to remove all irrelevant text, word abbreviations, emoticons, and other Twitter specific classifiers. The final step in the sentiment analysis approach is to pass the

filtered data set through a sentiment classifier that distinguishes the specific wording in the filtered tweets as being positive, negative, or neutral. The authors applied the Naive Bayes Classifier, which is "a probabilistic classifier applying Bayes' theorem with the assumption that features are conditionally independent from each other given a specific class" to categorize the text (Abdelrazeq et al., 2015, p. 56).

Dodd (2014) conducted a sentiment analysis on the opinions expressed on Twitter regarding a specific television program. The goal of the study was to discover perceptions on the television program, labeling the sentiment either positive or negative, while evaluating different machine learning algorithms. The authors placed an emphasis on evaluating different machine learning algorithms that can effectively realize a Twitter sentiment analysis. A significant aspect for the study was that the microblogging activities in Web 2.0 culture produces "massive amounts of data containing consumer sentiment that is predominantly unstructured text" (p. 1). To analyze such large data sets, the authors used a computational process of extracting information from a data set that structures the data and discovers patterns, called data mining. They also applied sentiment analysis, also known as opinion mining, which also discovers patterns in data that is analyzed to classify the sentiment in that text (Dodd, 2014). Since much of the sentiment in a Twitter feed is unstructured, it is necessary to use natural language processing techniques to "transform raw data into a form that can be processed efficiently by a computer" (p. 3). The study concluded that the use of natural language processing techniques was an effective method for conducting a sentiment analysis on the unstructured data set provided in a microblogging site such as Twitter.

Kumar (2014) defined sentiment analysis as a task finding process used to discover people's opinions and affinity towards a specific topic. These views are posted on social media sites such as Facebook and Twitter, which can be categorized into various types of sentiments like positive and negative. The objective of the study was to determine which text classifier produced the best results in classifying the sentiment of a substantial Twitter data set. The main obstacle in the classification process is that the data contains an abundance of noise in the form of slang, short words, acronyms, hashtags, emoticons and special characters, which makes it difficult to achieve good accuracy in the sentiment analysis process (Kumar, 2014). The author used various data preprocessing steps including Naive Bayes, Support Vector Machines (SVM) and Maximum Entropy (MaxEnt) to sort and classify the Twitter data set. Naive Bayes refers to a basic text classification algorithm that is derived from the Bayes theorem that considers the probability of an event and makes naive independent assumptions of the feature variables. The SVM is a classifier that constructs a hyperplane of analyzed data with a maximized separation that maps data into a category. The MaxEnt is a discriminative classifier used in Natural Language Processing, Speech and Information Retrieval problems which classifies the text (tweet, document, review) to a particular category, given a contiguous sequence of items from a given sequence of text in the form of a unigrams (single words), bigrams (two- word sets) or other n-gram features (Kumar, 2014). The study concluded that the text classification process incorporating these methods obtained higher performance results with the SVM data processing technique, followed by MaxEnt and Naive Bayes respectively.

Other studies have extended the sentiment analysis processing methods to generate further precise analysis results. Zhang and Lee (2011) state that the popularity of the microblog site Twitter makes it a valuable source to obtain public opinion for business marketing and social studies. The study performs a sentiment analysis on specific topics within a posted tweet using a pre-process lexicon-based approach, followed by an additional analysis using a sentiment classifier. The lexicon-based approach uses an opinion lexicon dictionary which consists of words that are commonly used to express positive or negative sentiments. The sentiment classifier used was Support Vector Machines which was the additional algorithm that categorized Twitter specific lexicon extracted from the initial preprocess such as emoticons, colloquial expressions, abbreviations, and unique spelled words (Zhang & Lee, 2011). The multi-tiered process generated improved sentiment analysis results, extending the learning-based sentiment analysis approaches currently used in most sentiment analysis studies.

Machine Learning Classifiers

The issue of classifying text by sentiment classification using machine learning techniques such as Naive Bayes, Maximum Entropy, and Support Vector Machines was considered by researchers Pang, Lee and Vaithyanathan (2002). They conducted a sentiment analysis using movie review data to determine if the overall sentiment was positive or negative. The objective of the study was to examine the effectiveness of applying machine learning techniques to the sentiment classification problem, which refers to the fact "topics are often identifiable by keywords alone, sentiment can be expressed in a more subtly manner, thus sentiment requires further comprehension than the traditional topic-based classification" (Pang et al., 2002, p. 1). According to the study,

machine learning techniques compared favorably over the human-generated baseline, with the Support Vector Machines process outperforming MaxEnt and Naive Bayes, although the differences were marginal (Pang et al., 2015, p. 6).

The machine learning algorithms most commonly used in the sentiment analysis process for text classification are Naive Bayes, Maximum Entropy, and Support Vector Machines.

Naive Bayes

The Naive Bayes classifier is a simple probabilistic classifier that applies Bayes' theorem with strong naive independence assumptions. Bayes' theorem dates back to the 1700s, when an English Reverend named Thomas Bayes attempted to compute a probability parameter, wanting to know how to infer causes from effects (Lesswrong, 2011). His work was published posthumously in 1764 by Richard Price and expanded upon in 1812 by Pierre-Simon Laplace who published the modern formulation in "Théorie analytique des probabilités" (Stigler, 1983; Stigler, 1986), although it has been argued that Nicholas Saunderson had discovered the theorem prior to Thomas Bayes. The theorem was widely rejected by mathematicians because of its simplistic approach and did not gain any real significant acceptability until the mid-1950s (Lesswrong, 2011). Despite the controversy, the Bayes' theorem is named after Thomas Bayes (Stigler, 1983).

Bayes' Theorem is a probability theory that describes the likelihood of an event, using prior knowledge of conditions related to the event. For example, if a ball is randomly dropped on a flat table, the location of the ball is noted. As additional balls are dropped on the table, their positions are also noted. The more balls are dropped, the better

the position of the first ball is predicted. Thus, a learning process ensues that permits an accurate probability to be determined (Bayesian, 2018; Berkson, 1930). The assuming nature of the model in the conditional independence within each class has caused a variety of names for the theorem, such as Idiot's Bayes, Naive Bayes, Simple Bayes, and Independence Bayes. These names commonly refer to the fact of its basic simplicity in that it ignores the interactions which might almost always be expected to exist (Hand & Yu, 2001).

Vidhya and Aghila (2010) provide an explanation of Naive Bayes machine learning approach as follows:

Naïve Bayes Model works with the conditional probability which originates from well-known statistical approach "Bayes Theorem", where as Naïve refers to "assumption" that all the attributes of the examples are independent of each other given the context of the category. In this context of text classification, the probability that a document d belongs to class c is calculated by the Bayes theorem as follows:

$$P(c / d) = \frac{P(c)P(d / c)}{P(d)}$$

The estimation of $P(d/c)$ is difficult since the number of possible vectors d is too high. This difficulty is overcome by using the naïve assumption that any two coordinates of the document is statistically independent. Using this assumption the most probable category 'c' can be estimated. (p. 206)

Furthermore, Danso, Atwell and Johnson (2013) offered a description on Naive Bayes as a simple machine learning technique that analyzes "the relationship between each feature and the class for each instance to derive a conditional probability for the relationship between feature values and the class" (p. 3).

Naive Bayes is calculated using the joint probabilities of features that estimate the probability that a given set of text is associated to a specific Class, such as the sentiment being considered as positive or negative. In the graph

above, the Feature quality would represent the probability that a specific text has been associated with a given Class or sentiment.

Multinomial Naive Bayes

A Naive Bayes model assumes that each of the features it uses are conditionally independent of one another given some class, in order to predict the category of a given sample. It is a probabilistic classifier, therefore Naive Bayes will calculate the probability of each category using the Bayes theorem, and the category with the highest probability will be output.

The term Multinomial Naive Bayes simply lets us know that each observed feature is a multinomial distribution, rather than some other distribution. The multinomial distribution works well for data which can easily be turned into counts, such as word counts in text.

In summary, Naive Bayes classifier is a general term which refers to conditional independence of each of the features in the model, while Multinomial Naive Bayes classifier is a specific instance of a Naive Bayes classifier which uses a multinomial distribution for each of the features (Vidhya & Aghila, 2010). For the purpose of the present study, the terms Naive Bayes and Multinomial Naive Bayes are used synonymously.

Maximum Entropy

Another probability distribution estimation technique used in machine learning algorithms is Maximum Entropy (MaxEnt). The MaxEnt approach is "used for a variety of natural language task such as language modeling, part-of-speech tagging, and text segmentation" all-of-which are sentiment analysis

components (Nigam, Lafferty, & McCallum, 1999, p. 1). The Maximum Entropy approach was described by Kumar (2014) as follows:

The MaxEnt classifier uses a model very similar to the Naive Bayes model but it does not make any independence assumption, unlike Naive Bayes. The MaxEnt classifier's goal is to classify the text (tweet, document, reviews) to a particular class, given unigrams, bigrams or others as features. If $w_1; w_2; \dots; w_m$ are the words that can appear in a document, according to bag-of-words model, each document can be represented by 1s and 0s indicating if the word w_i is present in the document or not.

The parametric form of the MaxEnt model can be represented as below:

$$P(c|d, \lambda) = \frac{\exp[\sum_i \lambda_i f_i(c, d)]}{\sum_c [\sum_i \lambda_i f_i(c, d)]}$$

Here, c is the class to be predicted, d is the tweet, and λ is the weight vector. The weight vector defines the importance of a feature. (p. 4)

According to Pang et al.,(2002), the MaxEnt classification is "an alternative technique which has proven effective in a number of natural language processing applications. Additionally, unlike Naive Bayes, MaxEnt makes no assumptions about the relationships between features, therefore MaxEnt might potentially perform better when conditional independence assumptions are not met" (p. 4).

Support Vector Machines

Support Vector Machines (SVM) is a classification technique that is highly effective at traditional text classification, generating better results than Naive Bayes and MaxEnt, since it is capable of handling noisy and sparse data sets (Danso et al., 2013; Pang et al., 2002). The SVM approach was described by Joachims (1998) as follows:

Support vector machines are based on the *Structural Risk Minimization* principle from computational learning theory. The idea of structural risk minimization is to find a hypothesis h for which we can guarantee the lowest true error. The true error of h is the probability that h will make an error on an unseen and randomly selected test example. An upper bound can be used to connect the true error of a hypothesis h with the error of h on the training set and the complexity of H (measured by VC-Dimension), the hypothesis space containing h . Support vector machines find the hypothesis h which (approximately) minimizes this bound on the true error by effectively and efficiently controlling the VC-Dimension of H . (p. 2)

The main idea of SVM is to select a hyper-plane that separates the positive and negative examples while maximizing the minimum margin (Kumar, 2014; Zhang & Lee, 2003; Pang et al., 2002; Joachims, 1998). "SVM employs a technique of 'maximal-margin-hyper-plane', where the maximum linear distance between Classes in the features space is estimated and separated from each other (Danso et al., 2013, p. 3). Support Vector Machines is suitable for text classification and out performs other machine learning algorithms since 1) it is capable of handling large text features; 2) it is able to limit irrelevant text features; 3) it is well suited for problems with dense concepts and sparse instances; and 4) can effectively find linear separators in corpus, which tends to be linearly separated (Joachims, 1998).

As expressed by the research, sentiment analysis is a comprehensive process that involves computational and natural language processing to identify and characterize large corpus in terms of emotion or attitude. A machine learning algorithm, such as Naive Bayes, Maximum Entropy, or Support Vector Machines, is used to classify a data set of text into a specific class, positive or negative as noted in various studies (Abdelrazeq et al., 2015; Danso et al., 2013; Dodd, 2014; Jhavaeri et al., 2015; Joachims, 1998; Kumar, 2014; Pang et al., 2002; Vidhya & Aghila, 2010; Zhang & Lee, 2003; Zhang et al., 2011). In particular, sentiment analysis is used to study attitudes and opinions expressed through various social media sites, primarily the microblog site Twitter. The results of a Twitter sentiment analysis for an organization can have a direct impact on the organization's communication output as it attempts to regulate effective interaction with its public while trying to establish favorable relationships.

Social Media and Microblogging

The Dialogic Loop Theory and sentiment analysis previously discussed is applied to the method of communication conducted on social media, in particular the microblogging site Twitter. Of specific concern for the present study is the intent and types of interaction that occur through social media. Therefore, it is necessary to identify the origins of social media as it developed as a derivative of the communication uniqueness created by the Internet, along with the emerging technology and communication reach in regard to Twitter.

In particular, the social media activity of colleges, specifically Twitter, is relevant to the present study, given the significant outcomes that social media can generate. There are several colleges and universities that use social media as a means to interact with their public. According to uniRank (2018), an international higher education directory and search engine, 48.8% of the colleges and universities in their directories, or 6,415 out of 13,146, use an official Twitter page to interact with their students, staff, alumni, and stakeholders. BestColleges (2018) ranks the top colleges and university that use social media: Harvard University with 843,000 Twitter followers and over five million Facebook followers, Stanford University with 609,000 Twitter followers and 1.2 million Facebook followers, and Yale University with 404,000 Twitter followers and 1.3 million Facebook followers. Colleges and universities increasingly see their presence, visibility and footprint on the social media as central to their reputation and international standing (Corren, Nelson & Weigle, 2017). Social media has provided a ubiquitous communication method that is incessant in nature, which captivates our attention and provides a voice to all that aspire to express their opinion or commentary. It has changed

social behavior and the way individuals interact on a daily basis. Its origin is structured on how individuals communicated using technology. Miller, Costa, Haynes, McDonald, Nicolescu, Sinanan, Spyer, Venkatraman and Wang (2016) mention that individuals communicated using media in two distinct ways 1) using public broadcasting such as television, radio and newspapers to reach a broad audience with a message, or 2) dyadic communication or communication between two people in the form of a conversation such as in a face-to-face interaction using a telephone, or using party line telephony in a group-based interaction manner. During this time, it was a difficult task for a single individual to extend a message to a specific attentive group of individuals, which could effectively interact and provide feedback to the original message. The development of the Internet and the use of email merged these two methods of communication, allowing individuals to broadcast an email or message to small or large groups of people via bulletin boards, specialized forums, chat rooms and blogging sites (Miller et al., 2016). Furthermore, the development of text messaging and the concept of instant messages being shared by groups of people, along with the rise of the smartphone, propelled the possibility of social interactions with groups of individuals that shared a common interest in a specific topic.

From a global perspective, social media commenced in Korea with the launching of the Cyworld platform in 1999. The Cyworld platform required users to establish a relation, which implied that they were "socially bound to the principles of reciprocity, such as commenting upon each other in a manner that evoked kin relationships" (Miller et al., 2016, p. 13). In terms of North America, the first social media platform that established similar interactive properties was MySpace, which was founded in 2003.

Since then several other popular social media platforms have been conceived, such as Facebook (2004), Twitter (2006), Instagram (2010), and Snapchat (2011).

Smits and Mogos (2013) define social media in reference to a transitional stage in the Internet, in which Internet content went from manager generated to user generated. The transitional period was labeled Web 2.0 by Tim O'Reilly in 2001 and refers specifically to the Internet activities associated with social media. These activities include "an emphasis on collaborative learning as well as on user engagement through participation, and one which allows immediate publication and wide distribution of user generated content" (Smits & Mogos, 2013, p. 3). The authors also mention the importance of the volume of users that interact and communicate within the social media platforms, continuously creating and sharing content in a collaborative way.

Ruhi (2014) defined social media as "a variety of online channels and platforms that facilitate collaborative creation and dissemination of information" (p. 2). The objective of social media, according to Ruhi, is to allow individuals to be creative, share their experiences, observations, and comments, and interact with other like-minded individuals. These activities are associated to an individual's activity stream and occur among established online relationships and their social network of friends, followers and fans, which Ruhi labels as "their social graph of connections" (2014, p. 2).

Social media is also defined as a medium that allows ordinary individuals, as opposed to professional journalists, to create user-generated news. Additionally, it encourages and assists public interaction, sharing of information, and collaboration (Murthy, 2012). Another element of social media is its ability to disseminate information and news. Individuals that use social media "often consume media produced by people

they have found of interest, leading to interactions with strangers" (p. 1061). Thus, social media interactions lead to an expansion of friends and relationships, beyond the personally known networks of people.

Therefore, it can be noted that social media is a venue which allows individuals to interact in a communal manner, expressing their feelings, thoughts, ideas, comments, experiences, and news to friends, acquaintances, and like-minded individuals. The interactive nature of social media provides a relational component through direct feedback from the message recipients, while facilitating the ability to propagate these messages to extended networks of individuals through re-posts and re-tweets. It is essential to note that the significance of a social media interaction is not founded on the platform utilized to share messages, rather on the content being shared. Miller et al., (2016) noted that "it is the content rather than the platform that is significant when it comes to why social media matters" (p. 1). It is precisely what is shared or communicated in a post or a tweet that generates the activity within a social media network. The post or tweet shared is particularly of interest to the current study since the sentiment analysis conducted on a Twitter feed for a specific community college can determine if the college's audiences are receiving positive or negative messages. Audiences will formulate a favorable or unfavorable opinion of these social media interactions, which could have an impact on the college's ability to establish effective relationships with its publics.

For the purpose of the present study, I consider the social media platform of Twitter as a means of communication and interaction. Twitter is a microblogging platform that allows users to send and read text-based messages of up to 280 characters. These messages or tweets enable individuals to express their opinion about different

topics, as well as mark those tweets with character identifiers or descriptors that are either hashtags (#) and/or at sign (@) (Grosse et al., 2012). Twitter messages are comprised of several descriptors since the platform developed its own markup language used in the posting, forwarding, and relationship building aspect of a message (Kwak, 2010). The "RT" indicates a retweet, which refers to forwarding or reposting a message; the "#" refers to a hashtag, which are added to a tweet and are used to classify and sort tweets according to specific topics or categories; the "@" used in conjunction with a user name (Ex: @username0), refers to a message being directed specifically as a reply to another user; emoticons or emojis, refer to small digital images or icons used to express an idea or an emotion; semantic identifiers or colloquial expressions that use informal words, phrases, or spellings, including slang words or exaggerated spelled words such as "SMH for shaking my head", "TBH for to be honest", "IMO for in my opinion", ":-)", "perrrrreeect", and "lovvvve"; and external web links (Balahur, 2013; Jhavaeri et al., 2015; Moss, 2013; Zhang et al., 2011).

Weller, Bruns, Burgess, Puschmann, and Mahrt, (2014) mention that "Twitter is an Internet-based communication technology that allows users to distribute short messages (tweets) of 280 characters or fewer on the World Wide Web or through smartphone apps" (p. 5). Other features include picture upload and display, and the abbreviation of URLs to assist in the 280 character limitation. Of particular importance is Twitter's dependency on the social connections, established through shared messages. A "sender-audience" relationship is established through the concept of "following" a user or subscribing to another individuals Twitter feed, thus their tweets or posted messages will appear alongside those of other followed individuals in a reverse-chronological timeline

of messages (Weller et al., 2014, p. 5). Unlike other social media platforms which require a reciprocal connection when being followed, Twitter users can follow any other user, while the user being followed does not need to accept, approve, or follow back the user (Kwak, 2010; Lee, Park, & Moon, 2010; Weller et al., 2014). According to Lee, De Zuniga, Coleman, and Johnson (2014), Twitter's "simple one-way subscription concept enables Twitter to combine the aspects of both social networking and news media within a single structure" (p. 793). The feature extends the communication reach of tweets beyond the known social networks of a user, expanding on the relationship building aspect of Twitter.

Alivídrež and Rodríguez (2016) state that Twitter as part of social media is used for creating and maintaining interpersonal relationships. Twitter is a communication tool that "is used to attract the attention of users, create emotional bonds with followers, and mobilize people to undertake concrete actions" (Alivídrež & Rodríguez, 2016, p. 90). Additionally, Lee et al., (2014) state that social media is dialogic, expanding ways for information-sharing and relationship building practices. The authors contend that the information flow within social media is "multidirectional, interconnected, and difficult-to-predict, with content that is transformed into shared data objects that is freely accessible, searchable, and traceable to anyone" (p. 793). These characteristics, along with the ability to retweet or forwarding of another users tweets to their own followers generates bidirectional relationship building aspects along with the broadcast of the message.

Of significance is the ability to analyze these messages for opinion and/or emotion expressed. Twitter, being the most popular microblogging communication tool,

serves as an adequate venue to assess public opinion on a variety of different issues. "As the audience of microblogging platforms and services grows every day, data from these sources can be used in opinion mining and sentiment analysis tasks" (Grosse et al., 2012, p. 2). Zhang et al., (2011) also view Twitter as a popular and rapidly growing microblogging platform in which "an increasing number of people are willing to post their opinions" (p. 2). Thus, Twitter is considered a valuable online source for measuring public opinion in regard to business marketing, social studies, and sentiment analysis. Jhavaeri et al., (2015) view Twitter data as a valuable source of information for analysis on opinion-related data that can provide productive insight in regard to a company's marketing strategies and decision making process. The authors use sentiment analysis as a means to capturing individual's collective opinions and attitudes through Twitter's publicly accessible application programming interface (API). The Twitter API facilitates the data-mining process by "allowing developers to search through Twitters database of tweets for specific words or phrases" (Jhavaeri et al., 2015, p. 15). The availability of the Twitter API, a searchable database of all tweets, provides an ample setting to gather the collective opinions and attitudes expressed by individuals, businesses, and organizations. The searchable database characteristic makes Twitter a valuable source of information for research across various levels of analytics, particular sentiment analysis, which is the focus of the present study.

Summary

The current study consists of several distinct areas of research. An extensive literature review was conducted covering the theoretical framework used to evaluate the research questions of the present study. The research noted that the Dialogic Loop Theory

has been incorporated in several studies attempting to show the two-way communication interactions between an organization and its audience, with the intent to foster meaningful relationships through microblogging. Furthermore, the quantitative content analysis methodology was defined as a systematic process used to evaluate the overall attitude of a corpus by labeling and categorizing specific terms within the corpus. As part of the content analysis methodology, a sentiment analysis process was identified as a process specifically used to evaluate the meaning of a message, if positive or negative in regard to the attitude expressed in the microblogging site Twitter. Several studies were cited which used a content analysis type of methodology in the research process for a Twitter analysis. Within the analysis, it was noted that the studies used data analytic tools and software such as Naive Bayes, Support Vector Machines (SVM) and Maximum Entropy (MaxEnt) to sort and classify the Twitter data set. Of particular concern for the current study is the communication interaction that occurs in social media, specifically the microblogging site Twitter. The literature research identified the origins of social media and the significance of the emerging technology of the microblogging site Twitter. The research revealed the specific components of Twitter and provides several examples of studies that consider the communication reach in regard to Twitter, its relevance to communication, and the potential relationship building aspects.

These areas of research serve as the foundation for the study being conducted, which is to conduct a sentiment analysis comparing the Microblog Dialogic Communication of a cross-section of urban and rural community colleges within the United States. The study uses data-mining techniques, linguistic analysis, and Dialogic Loop Theory to measure the relationship capability of their Twitter messages along with

the attitudes expressed through these messages. The process provides a perspective on the effectiveness of the college's Twitter communication. The current study should be of great concern for colleges since the messages posted on Twitter create a perception among their public and are critical to maintain their student populations engaged in college wide activities, attitudes, and values.

CHAPTER III. METHODOLOGY

The purpose of the present study is to conduct a sentiment analysis on the Twitter activity from a cross-section of urban and rural community colleges within the United States. The current research uses the Dialogic Loop Theory to analyze the college's Twitter activity as a means to establish favorable relationships with its public as determined by the positive or negative sentiment of the posted messages. The current research uses content analysis of the Twitter activity from the selected community colleges. Berelson (1952) defines content analysis as "a research technique for the objective, systematic, and quantitative description of the manifest content of communication" (p. 18) with the intention of being able to describe the basic characteristics of a particular content. Therefore, a set of procedures can be used to infer the sentiment of content or text (Weber, 1985). A content analysis is appropriate for the current study since it is suitable for conducting analysis on large quantity of data (Kerstetter, 2014). Since Twitter generates large quantities of desirable corpus, a content analysis is the most appropriate methodology.

The research process employed consists of using a data-mining search routine and filter associated with the Twitter API (Application Programming Interface) that allows the capture of a copious of the Twitter activity from a cross-section of urban and rural community colleges in the United States. These Twitter capabilities captured tweets from specific colleges and for a specific time span. Once a collection of raw tweets is downloaded and archived, it must be preprocessed to remove any irrelevant text, abbreviations, and other Twitter specific annotations. Once the data set of tweets is stripped of unnecessary content, the features were extracted and used for the training and

classification process using a machine learning approach towards a sentiment classifier. A sentiment classifier is trained from unigram and bigram, and other features extracted from tweets obtained from a training corpus. Once the sentiment classifier has been trained with the training data, the actual college tweets underwent a similar process of filtering and feature extraction before passing them through a sentiment classifier that categorized each tweet as being positive or negative. The final step in the machine learning process uses Naive Bayes as the sentiment classifier. The Naive Bayes Classifier computes the probability of each tweet's sentiment using the defined features. The Naive Bayes employs supervised learning in which the samples provided (tweets) must be tagged or labeled using the class they belong to, such as positive or negative. The tagging process requires manually performing this task or using an existing set of tagged tweets, such as Sentiment140, which contains over 1.5 million tagged tweets.

A significant component of utilizing a machine learning algorithm is the training phase, which requires a balanced number of samples associated with positive and negative tweets to avoid model over fitting. The model constructed using the Naive Bayes algorithm needs to represent a wide-range of positive and negative tweets to avoid the model skewing the classification process towards favoring one class over the other. Once the results have been obtained associated with the sentiment classification process, then statistical analysis is applied to evaluate the significance of the classification process. Figure 4 illustrates the general process described, indicating the various stages and transitions in building the model and utilizing it for tweet sentiment classification applied in the present study.

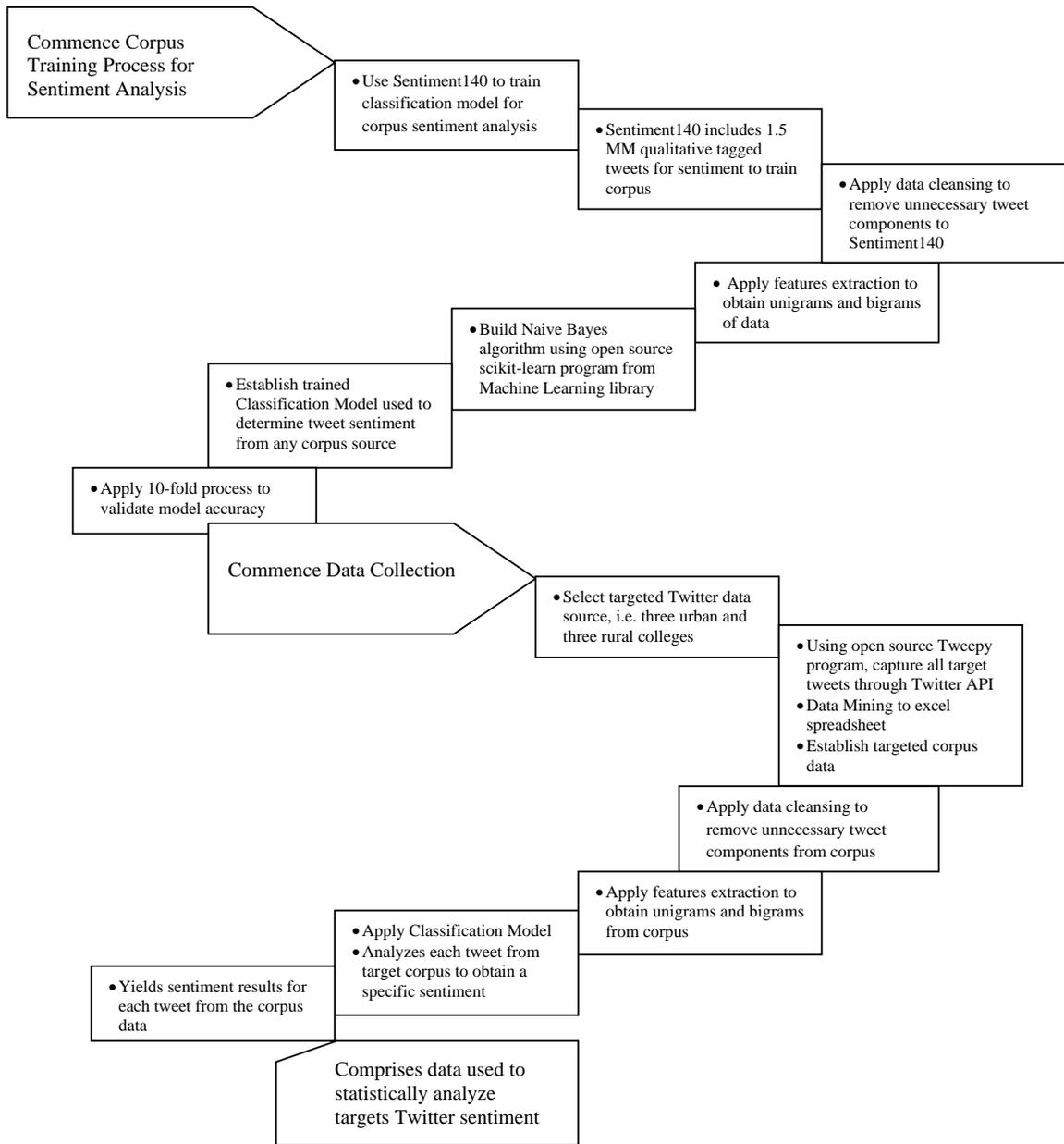


Figure 4. Sentiment Classification Process

Data Source

The microblogging Twitter platform served as the data source for the present study. It consists of the tweet universe from which I obtained the data from to conduct the research. The focus is college’s microblogging sentiment and, as such, I selected a cross-section of urban and rural community colleges to perform the analysis. The

selection of these college consists on colleges that are associated with an active social media engagement strategy. According to Kent and Taylor's (1998) Dialogic Loop Theory, an organization must invest in having a dedicated well-trained staff to continuously interact with its public to engage in relationship building. The colleges selected both have an active marketing department that promotes institutional awareness and brand management across several media platforms, including their online media activity such as Twitter. Specifically, the National Center for Educational Statistics was searched along with the web site Campus Explorer, which is a free online service offering college rankings and data on over 8,500 US colleges, to obtain a list of the top five largest community college's in the Unites States. Three urban area colleges were randomly selected from the list. Additionally, the web site College Data was used to identify the best small colleges from the United States. A list of rural area colleges was extracted, using a 500 to 10,000 student population criteria. Three rural area colleges were randomly selected from the list. The Twitter activities associated with these colleges were electronically captured using the Twitter API to conduct the sentiment analysis.

Data Set - Twitter API - Twitter Data Collection

The objective of the present study is to analyze the sentiment of tweets associated with a cross-section of urban and rural community colleges. Twitter is a microblogging social network platform where users share messages with a maximum size of 280 characters. These short messages or tweets establish an extensive repository of data in regard to each Twitter user. For the purpose of the study, it is necessary to data mine the tweets associated with these institutions. To perform data-mining on Twitter, an automated manner must be developed to efficiently acquire a copious of the tweets

associated with these institutions. First, an account was registered in the Twitter website to gain access to the Twitter API. Similar to Abdelrazeq et al., (2015), Pak and Paroubek (2010), and Zhang et al., (2011), the Twitter API was used to search and acquire all of the tweets associated with a specific identifier (name, id, etc.). For the present study, the identifier information for the college's being analyzed was used as the search criteria. Additionally, an application program is used to interface with Twitter to extract all of the tweets and import the extensive data source to an excel spreadsheet for further analysis. These steps are detailed to provide support for future research and extending the present work.

Register Data Extraction Application with Twitter - SpeechBusComm

To obtain access to the Twitter API, it is necessary to establish a Twitter account. Once the Twitter account was activated, it is required to create an application or app called SpeechBusComm that interacts with the Twitter API, covered in section 3.3, Data Extraction Application. The initial action is to register the Data Extraction Application by accessing the Twitter developer website <https://dev.twitter.com/> as in Figure 5 and clicking on *My App* link on the top of the page. A page that allows an application to be registered with Twitter is obtained, click the “*Create New App*” button, which requires additional information as shown in Figure 6 to define the application for Twitter. Once these steps are completed, Twitter will provide four pieces of information that are incorporated into the Data Extraction Application to authenticate the app with Twitter. Access to the Twitter API is then obtained where different data operations can be performed. Figure 7 shows the information, since this is a live Twitter application, the actual codes that authenticate and identify the application to Twitter have been redacted.

The codes are the *Consumer Key (API Key)*, *Consumer Secret (API Secret)*, *Access Token*, and *Access Token Secret*, which are used in the Data Extraction Application.

These key value strings must be kept private since they provide the application access to Twitter on behalf of the account. The default permissions are read-only. The next step is to proceed with the data extraction from Twitter for the focused organizations.

Another aspect to consider when extracting data using Twitter API are the rate limits imposed by Twitter. Rate limits refers to the Twitter restrictions on the amount of Twitter information that can be downloaded using Twitter's API to prevent downloadable data sets (Rate limiting, 2017). For the purpose of the present study, the Twitter data was acquired in blocks of 3000 tweets by executing the Data Extraction Application in different 15-minute windows. Twitter imposes time limits to download bulk data of tweets, before throttling it off at around 3000 tweets. Therefore, data collection for all of the tweets associated with the targeted organization were obtained by executing the application various times, using a cumulative starting point to augment previously acquired tweets with the next 3000 block of tweets until obtaining a copious of tweets for each organization.

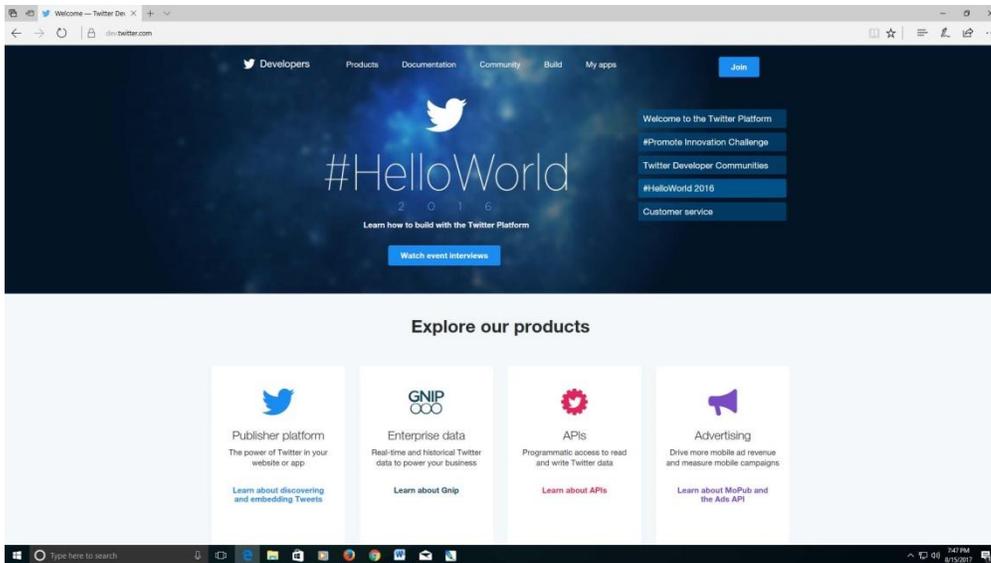


Figure 5. Twitter Developer Website

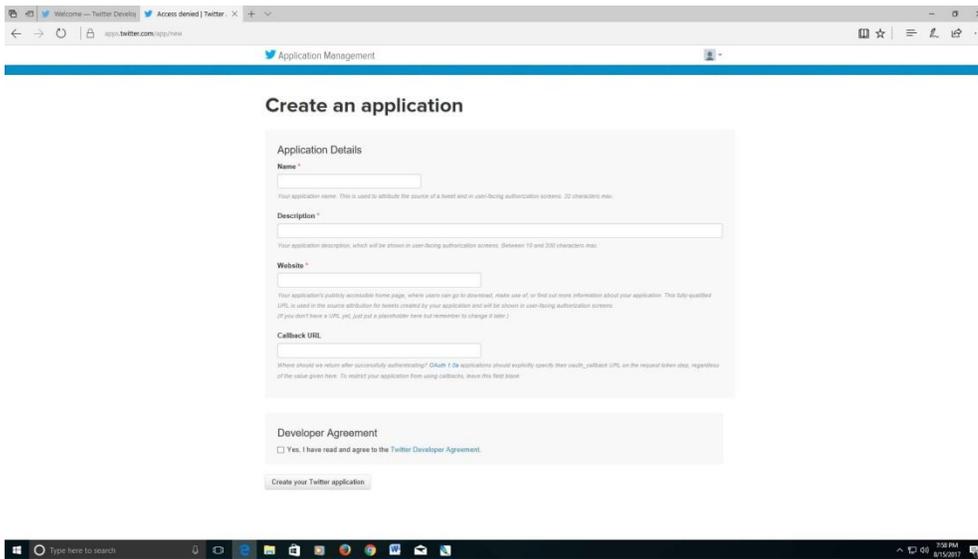


Figure 6. Twitter Application Creation

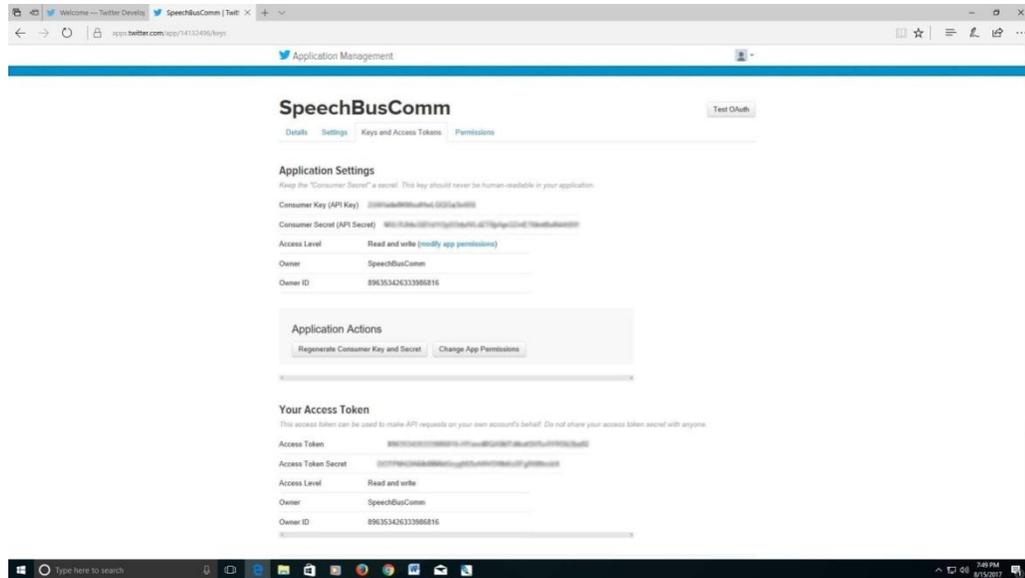


Figure 7. Application Customer Key and Token

SpeechBusComm - Data Extraction Application

The application implemented uses Twitter's Representational State Transfer (REST) APIs to interact with their service. The application utilizes Tweepy, which is a Python library that enables Python code to interact with Twitter. The application is built using Python scripting language. Tweepy is an open-sourced library that is hosted on GitHub (Tweepy: Twitter for Python, 2018), a repository of developer code, which enables Python to communicate with the Twitter platform and use its API. Once Tweepy is installed, it is then possible to use the Python program to extract the tweets. The system used to execute the python code runs the operating system Ubuntu 16.04. To install Tweepy, the following command was executed: *sudo apt-get install python-tweepy*

The application is called `tw_csv_dump.py` and is written in python. It was written with comments to make it readable for future use. For every tweet obtained, there are over 250 data items that can be extracted from a status object (Tweet data dictionary,

2018), what Twitter calls information returned from a tweet. It consists of information like, Geographic location, user ID, source, and many more. The items selected are Tweet_ID, Tweet_Created_At, and Tweet_Text and are illustrated in Figure 8, which shows a few obtained items from the extracted file. The comma separated fields correspond respectively to the previously mentioned items extracted from the Status Object. The name of the file will have the corresponding organization or college name, for example an organization called Student College will have a file named StudentCollege_tweets.csv.

```
896068903578640384,2017-08-11 18:00:49,Student Community College Named as One of the Top Colleges for
Hispanics by @HispanicOutlook: https://t.co/xsKeh60ghj https://t.co/yJwc3lt6o2

896057472271605760,2017-08-11 17:15:23,Fall 2017 Semester is just around the corner! #BeSCC and register for
classes today. https://t.co/wjhVJrwyj4A

896023721076568064,2017-08-11 15:01:17,SCC's West Campus Presents Henry Watle's Exhibition Under the Ocean
Pool on Sept. 28:... https://t.co/LpdhuRmo6

895695684149678084,2017-08-10 17:17:46,"SCC in Focus is a training ground for future journalists - meet one
amazing alumna, Chastity Bryan! https://t.co/sZb6uldI2x"
```

Figure 8. *Sample extracted tweets*

The program is started as:

```
> python tw_csv_dump.py StudentCollege
```

Since the Twitter API limits the number of tweets extracted to 3000 for a rate limiting window every 15 minute, the Tweet ID of the last obtained tweet is used to run the application again from the point the application left off to obtain the subsequent 3000 tweets. The process continues until all the tweets for the organization have been extracted. The application would be started in as follows:

```
> python tw_csv_dump.py StudentCollege 895695684149678084
```

The tweet extraction does not begin at the top of the tweet list or the latest tweet, but at the tweet located at 895695684149678084 (the value is not an actual location, but the number represents the relative position of other tweets in the tweet chain. The code will read tweets in blocks of 200 until 3000 are reached or the end of the tweets are reached. Figure 9 illustrates the code for the program.

```

#!/usr/bin/env python
# encoding: utf-8

import tweepy #https://github.com/tweepy/tweepy
import csv
import sys

#Twitter API credentials – need to be filled in from Twitter Application registration page
#consumer_key = ""
#consumer_secret = ""
#access_key = ""
#access_secret = ""

maxid = ""

def get_all_tweets(screen_name):
    #Twitter only allows access to a users most recent 3240 tweets with this method

    #authorize twitter, initialize tweepy
    auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_key, access_secret)
    api = tweepy.API(auth)

    #initialize a list to hold all the tweepy Tweets
    alltweets = []
    new_tweets = []
    outtweets = []

    #make initial request for most recent tweets (200 is the maximum allowed count)
    if maxid == "":
        new_tweets = api.user_timeline(screen_name = screen_name,count=200)
    else:
        new_tweets = api.user_timeline(screen_name = screen_name,count=200,max_id=maxid)

    #save most recent tweets
    alltweets.extend(new_tweets)
    #save the id of the oldest tweet less one
    oldest = alltweets[-1].id - 1
    #keep grabbing tweets until there are no tweets left to grab
    while len(new_tweets) > 0:
        print "getting tweets before %s" % (oldest)
        #all subsiquent requests use the max_id param to prevent duplicates
        new_tweets = api.user_timeline(screen_name = screen_name,count=200,max_id=oldest)
        #save most recent tweets
        alltweets.extend(new_tweets)
        #update the id of the oldest tweet less one
        oldest = alltweets[-1].id - 1
        print "...%s tweets downloaded so far" % (len(alltweets))
    #transform the tweepy tweets into a 2D array that will populate the csv
    outtweets = [[tweet.id_str, tweet.created_at, tweet.text.encode("utf-8")] for tweet in alltweets]
    #write the csv
    with open('%s_tweets.csv' % screen_name, 'wb') as f:
        writer = csv.writer(f)
        writer.writerow(["id","created_at","text"])
        writer.writerows(outtweets)
        pass

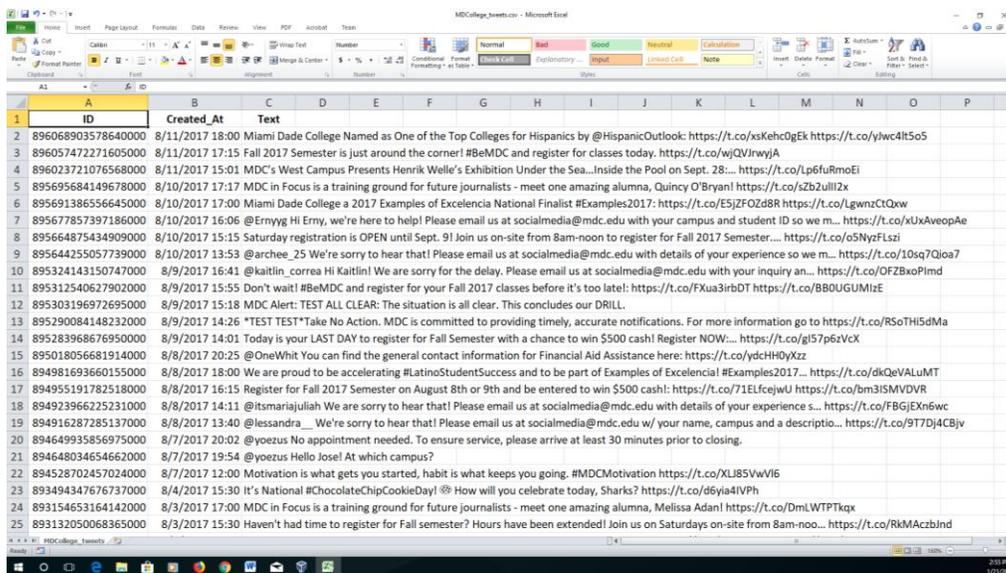
if __name__ == '__main__':
    #pass in the username of the account you want to download
    if len(sys.argv) < 2:
        print 'Must specify a twitter account name'
    else:
        if len(sys.argv) > 2 and sys.argv[2] != "":
            maxid = sys.argv[2]
            get_all_tweets(sys.argv[1])

```

Figure 9. *Python Program Code*

SpeechBusComm – Load CSV Files into Excel

The previous described process generates several CSV files per organization, if the total number of tweets is greater than 3000 tweets. These files are imported into Excel and constitute the collective tweets or copious of tweets from each of the college's being analyzed for the present study. Once the datum has been imported into an Excel spreadsheet format, it is ready for the preprocessing phase of sentiment analysis. Figure 10 depicts the layout and composition of a few of the tweets loaded into Excel.



ID	Created_At	Text
896068903578640000	8/11/2017 18:00	Miami Dade College Named as One of the Top Colleges for Hispanics by @HispanicOutlook: https://t.co/xsKehc0gEk https://t.co/yhw4it5o5
896057472271605000	8/11/2017 17:15	Fall 2017 Semester is just around the corner! #BeMDC and register for classes today. https://t.co/wjQV1rvwyjA
896023721076568000	8/11/2017 15:01	MDC's West Campus Presents Henrik Welle's Exhibition Under the Sea...Inside the Pool on Sept. 28:... https://t.co/Lp6fuRmoEI
895695684149678000	8/10/2017 17:17	MDC in Focus is a training ground for future journalists - meet one amazing alumna, Quincy O'Bryan! https://t.co/szB2ulll2x
895691386556645000	8/10/2017 17:00	Miami Dade College a 2017 Examples of Excelencia National Finalist #Examples2017: https://t.co/E5jZFOZd8R https://t.co/LgwnzCtQxw
895677857397186000	8/10/2017 16:06	@Ernyng Hi Erny, we're here to help! Please email us at socialmedia@mdc.edu with your campus and student ID so we m... https://t.co/xUxAveopAe
895664875434909000	8/10/2017 15:15	Saturday registration is OPEN until Sept. 9! Join us on-site from 8am-noon to register for Fall 2017 Semester.... https://t.co/o5NyzFLszl
895644255057739000	8/10/2017 13:53	@archie_25 We're sorry to hear that! Please email us at socialmedia@mdc.edu with details of your experience so we m... https://t.co/10sq7Qioa7
895324143150747000	8/9/2017 16:41	@kaitlin_correa Hi Kaitlin! We are sorry for the delay. Please email us at socialmedia@mdc.edu with your inquiry an... https://t.co/OFZBxoPlmd
895312540627902000	8/9/2017 15:55	Don't wait! #BeMDC and register for your Fall 2017 classes before it's too late! https://t.co/FXua3irbDT https://t.co/BB0UGUM1zE
895303196972695000	8/9/2017 15:18	MDC Alert: TEST ALL CLEAR. The situation is all clear. This concludes our DRILL.
895290084148232000	8/9/2017 14:26	*TEST TEST*Take No Action. MDC is committed to providing timely, accurate notifications. For more information go to https://t.co/RSoTHI5dMa
895283968676950000	8/9/2017 14:01	Today is your LAST DAY to register for Fall Semester with a chance to win \$500 cash! Register NOW... https://t.co/gI57p6zVcX
895018056681914000	8/8/2017 20:25	@OneWhit You can find the general contact information for Financial Aid Assistance here: https://t.co/ydcHH0Ykzz
894981693660155000	8/8/2017 18:00	We are proud to be accelerating #LatinoStudentSuccess and to be part of Examples of Excelencia! #Examples2017... https://t.co/dkQeVALUMT
894955191782518000	8/8/2017 16:15	Register for Fall 2017 Semester on August 8th or 9th and be entered to win \$500 cash! https://t.co/71ELfcejwU https://t.co/bm3ISMVDVR
894916287285231000	8/8/2017 14:11	@itsmariajullah We are sorry to hear that! Please email us at socialmedia@mdc.edu with details of your experience s... https://t.co/FBGjEXn6wc
894916287285231000	8/8/2017 13:40	@lessandra__ We're sorry to hear that! Please email us at socialmedia@mdc.edu w/ your name, campus and a descriptio... https://t.co/9T7Dj4CBjv
894649935856975000	8/7/2017 20:02	@yoezus No appointment needed. To ensure service, please arrive at least 30 minutes prior to closing.
894648034654662000	8/7/2017 19:54	@yoezus Hello Jose! At which campus?
894528702457024000	8/7/2017 12:00	Motivation is what gets you started, habit is what keeps you going. #MDCMotivation https://t.co/XL85VwV16
8934943476737000	8/4/2017 15:30	It's National #ChocolateChipCookieDay! 🍪 How will you celebrate today, Sharks? https://t.co/d6ya4IVPh
893154653164142000	8/3/2017 17:00	MDC in Focus is a training ground for future journalists - meet one amazing alumna, Melissa Adan! https://t.co/DmLWPTKqax
893132050068365000	8/3/2017 15:30	Haven't had time to register for Fall semester? Hours have been extended! Join us on Saturdays on-site from 8am-noon... https://t.co/RkMacczbjnd

Figure 10. Excel Tweet Layout and Composition Sample

Data Preprocessing

The next step in the sentiment analysis process is to prepare the obtained corpus data for the analysis. The preprocessing phase is essentially a cleaning of the data, removing all information that is considered irrelevant to the sentiment of the message. The cleaning of the data is one of the most significant components of developing an effective analyzer, since it scrubs each tweet to contain the main essence of its meaning,

removing superfluous text or symbols, while rendering a set of words. There are several components to a Twitter tweet which do not provide any useful information for the sentiment analysis process and thus must be removed or extracted from the data. Some of these components are:

- Retweets which are messages that have been reposted or forwarded on Twitter are removed from the data set. These messages are prefixed with a \RT acronym and create data redundancy.
- Twitter tags that serve as replies or mentions do not provide any value to the sentiment analysis and therefore are removed. These are typically tagged with the prefix of an ampersand sign (@) followed by a username, such as "@username1", and serve to identify the target of the message.
- Hashtag (#) markup language such as "#BeHappy" is used to identify keywords, topics, and organize tweets according to specific categories within Twitter. Hashtags are an essential part of the Twitter language, allowing tweets to be searchable. These tags possibly contain valuable information for the sentiment analysis, thus the prefix # symbol is removed from the tweets, leaving only the word.
- Web links in the form of an active URL such as "http://a.co/fDeawsB", are commonly used to redirect the Twitter user to an external URL source and other general links to images, blogs or documents are removed from the data set.
- Repeated letters that exceed two similar letters are filtered out of the tweets. Due to Twitters character limitations, users often over emphasize words by repeating letters such as "Happyyyyy". The two letter constraint replaces the exaggerated

word spelling with a more accurate spelling "Happy" while allowing for word spellings that have repeated letters such as the word "cool".

- Punctuations and other miscellaneous data such as exclamation marks (!), question marks (?), numbers (123), quotes (""), semicolons (;), colons (:), commas (,) and other unique symbols are not considered applicable for the sentiment analysis and are removed from all tweet data.
- Emoticons are often sources of abbreviated symbols used to represent an emotion on Twitter. User often express a happy feeling with a smiley emoticon such as :-), ;-), :) or ;) and express unhappy feelings with a sad emoticon such as :-(or :(.
These emoticons serve as a user sentiment representation thus the specific symbol combinations replaced by a word interpretation of the symbol combination of "happy" and "unhappy" respectively.
- Word abbreviations such as "bc" representing the word "because" and acronyms like "lol" meaning laughing-out-loud, are both popular forms of expression on Twitter are not considered relevant to the sentiment analysis and are removed from all tweets data.
- Stop-words are commonly used words such as "the", "and", "or", "while", etc. that are used as conjunctions in sentences. These words do not contribute to the sentiment of the tweet and are removed from the data. A python program is used to extract the stop-words using a stop-words corpus obtained from the Natural Language Tool Kit (NLTK, <http://www.nltk.org/>) at the URL location <https://pythonspot.com/nltk-stop-words/>. Some stop-words were removed from the 153 stop-words corpus since they represented an important part in identifying

negative sentiment, such as "nor", "not", and "neither". These words were not removed from the tweet data.

- Stemming refers to the process of reducing inflected or derived words to its word stem or root form. For example the words "talking, talker, and talked" are converted to the root word "talk". Stemming increases the probability of the word since all derived iterations of a specific word point to the same root form word. The NLTK package for stemming used for this task was the LancasterStemmer (or PorterStemmer).

A clean data set is obtained once all tweets have been purged of inconsistencies and irrelevant information. A clean data set improves the quality of the data allowing for further analysis.

Feature Extraction

A very significant component in the construction of a machine learning classifier is the selection of features that serve as the input to the classifier algorithm. The goal of the feature selection phase was to establish a set of features that are the most effective during the classification process. The properties extracted from tweets serve as the source for feature selection, i.e., the words composing a tweet, the size of a tweet, hashtags, lexicons, emoticons, etc. These features are the input for the classifier. The different feature selection approaches can be associated to the manner in which the tweet sentences are analyzed to correlate them to the tweet sentiment (positive or negative). For example, the words composing a tweet can be viewed as a Bag-Of-Words, if they are used as single words, then they are considered as *unigram*, if they are combined as double-words,

then they are referred to as *bigrams* or *trigrams* as three combined words (Go et al., 2009). In addition, other features consist of part-of-speech tags and lexicons (Kouloumpis et al., 2011) and hashtags and emoticons (Davidov et al., 2010). The significance of using n-grams, where $n=1$, compared to $n \geq 2$ is that it can cover a greater range of different tweet text applied to different areas. These unigrams can be used in different situations to capture more general sentiment. Unigrams lack the ability to capture sentiment expression patterns as is possible with n-grams, where $n \geq 2$, such that as n becomes greater, more specific sentiment is achievable (Abdelrazeq et al., 2015). The motivation to use a specific n-gram is linked to research that has shown that $n \geq 2$ can capture domain specific sentiment as intended in the current study that focuses on the college domain. From the literature, unigrams have shown effective outcomes in sentiment classification of specific domains, i.e. movie reviews (Pang & Lee, 2008) and bigrams and trigrams have been effective in product review polarity classification (Dave et al., 2003).

The approaches adopted consist of a combination of both of these schemes as tweet features, since the unigrams' coverage of the data and the bigrams' ability to capture the sentiment expression patterns have demonstrated the most effective outcome (Pak & Paroubek, 2010). According to published results, using part-of-speech reduces the performance of the sentiment classifier (Go et al., 2009) while converting emoticons to their text sentiment leads to better classification (Kouloumpis et al., 2011). Therefore, the features used to build the sentiment classifier consisted of both unigram and bigrams, encompassing the emoticon converted into text.

Sentiment Classifier

The literature yielded various alternatives to using Twitter sentiment classifiers as a means to perform an automated analysis. These approaches have primarily deployed supervised classifiers, which require a training dataset that consist of the tagged or labeled tweets. The three most popular or common classifiers used from the machine learning field consist of Naïve Bayes, Support Vector Machines and Maximum Entropy. Among these classifiers, Naïve Bayes has more often demonstrated better performance when applied to a broader class of domains (Go et al., 2009). As such, I used the Naïve Bayes classifier as it is a proven method used for text classification. It is a probabilistic classifier applying Bayes' theorem which makes the assumption that its features are conditionally independent of each other when computing the classification probability. The algorithm computes the probability of having a tweet with the sentiment, either positive or negative, if it contains a specific feature. It is calculated according to the probability of the feature's existence in all tweets, and the probability of finding the feature in the tweets that belongs to that specific sentiment.

Training Data

The classifier is only as effective as the training data utilized. It is necessary to expose the algorithm to a broad set of manually tagged or labeled tweets so that the classifier can learn what feature values are most probabilistically associated with the sentiment classes of positive and negative. As such, I adopted a training dataset which has been previously tagged and is sufficiently large to train the classifier and sufficiently robust to handle the quantity of tweets from the domain of college tweets. In addition, I must be careful of over fitting the classifier, where the training data, if not balanced,

could skew the classifier to one class over the other, generating low accuracy. Therefore, I used the training dataset Sentiment140 (Sentiment140, 2018) which consists of over 1.5 million tweets that have been manually tagged to positive or negative. From the Sentiment140 dataset, I extracted the positive and negative sentiment samples of equal proportions, referred to as the basic dataset. From the basic dataset, two datasets are extracted, one for training and the other for classifier validation, known respectively as the *training dataset* and the *testing dataset*.

To perform classifier validation, a 10-fold cross validation was employed to evaluate the effectiveness of the Naïve Bayes classifier. For a 10-fold cross validation, the basic dataset is divided into 10 equal parts of tweets, nine of the parts are selected as the training dataset and one of the parts is the testing dataset (Refaeilzadeh, 2009). The results obtained are recorded, which indicate the performance of the classifier for that one test. Then, another training dataset is created which selects a different nine parts of the basic dataset and another one part is used for testing. The process is conducted 10 times, each time selecting a previously not used part for testing and the remaining nine parts become the training dataset. The final results are the average of the 10 classifier evaluation results. Thus, 10-folds means 90% of full data is used for training (and 10% for testing) in each fold or part test. The process is a compromise motivated by the fact that 90% is a large portion of the basic dataset which should produce good results as it is close to 100% of the basic dataset. The use of other fold sizes, for example 5, would lead to less data being used to train the model, 80% and a fold of 20 would increase the percentage of the training dataset, but would require far more computation resources as I would have to compute the performance 20 times. Therefore, a common technique that is

adopted for the current study is to perform 10-fold cross validation to evaluate the effectiveness of the classifier since it uses a higher percentage for training. The results after each iteration in the 10-fold cross validation is a confusion matrix that quantifies false-positive and true-negative thus misclassified data, which would indicate the number of positive tweets and negative tweets incorrectly classified. These were utilized in evaluating the classifier performance.

Statistical Analysis

Once the results were obtained from the sentiment classification process, statistical tests were conducted to evaluate the significance of the classification process. The Twitter data was obtained through the use of a software acquisition process. These data were subsequently coded using a machine learning process that applied a data source of 1.5 million tweets which had previously been analyzed for sentiment. Three separate data analysis classifiers were then used to code the data using the sentiment. The coding identified tweets as positive sentiment tweets or negative sentiment tweets for each of the three urban colleges and three rural colleges. Further data analytics included retweets, likes, and the various social media components on the inclusion of URLs, at signs (@), and hashtags (#), within the tweet population.

The independent variable considered for the current study was the college type, which was categorized as either urban or rural, and the dependent variable was the sentiment type, or positive sentiment and negative sentiment, obtained from the tweet population. Subsequently, statistical tests were run to see if the tweet sentiment were associated with the urban or rural colleges. The quantitative data were entered into SPSS and analyzed for findings (Morgan et al, 2007). Descriptive statistics, frequencies and

cross tabulations were used for comparing counts and percentages, and Pearson's chi-square test and Phi were used for testing the hypotheses. The p level was set at .05 which is the standard significance level.

CHAPTER IV: RESULTS

Participants

The participants for the Twitter analysis in the current research were selected from two separate lists. The first list presented large colleges located in urban areas in the United States. The list was obtained from the National Center for Educational Statistics and a college ranking online web site called Campus Explorer. The second list presented small colleges located in rural areas in the United States. The second list was obtained from a college search and admission assistance web site called College Data. From these two lists, three urban area and three rural area colleges were selected.

The urban colleges were labeled U1, U2, and U3, and the rural colleges were labeled R1, R2, and R3, respectively. As seen in Table 2 below, the student enrollment for the urban colleges averaged 82,848 enrolled students for a given academic year, with U1 having 85,759 students enrolled for the 2016-17 academic year, U2 having 89,130 students for the Fall 2018 academic year, and U3 having 51,190 student enrollment for the 2017-18 academic year. The rural colleges considered for the current research had an average student enrollment of 3,493, with R1 having 2,042 enrolled students for the 2016-17 academic year, R2 having 6,571 enrolled students for the 2018 academic year, and R3 having 1,828 student enrollment for the 2018-19 academic year.

Table 2
Student Enrollment Totals by Gender

	Male	%	Female	%	Unreported	%	Total Enrolment
U1	36,484	42.5	49,191	57.4	84	0.1	85,759
U2	35,150	39.4	53,980	60.5	-	-	89,130
U3	24,807	48.5	26,383	51.5	-	-	51,190
R1	1,051	51.5	991	48.5	-	-	2,042
R2	3,385	51.5	3,186	48.5	-	-	6,571
R3	903	49.4	925	50.6	-	-	1,828

SOURCE:

https://www.mdc.edu/ir/Fact%20Book/FB_Credit_Student_Enrollment_Profile.pdf
<http://www.lonestar.edu/about-us-institutional-research.htm>
<http://www.lonestar.edu/images/OD%20FA18%20Enrollment%20Report.pdf>
https://www.nvcc.edu/oir/_files/factbooks2013-2018.pdf
https://provost.williams.edu/files/williams_cds_1617.pdf
<https://www.dartmouth.edu/~oir/data-reporting/factbook/enrollment.html>
<https://www.bowdoin.edu/ir/pdf/common-data-set-2018-19.pdf>

The gender demographic of the urban schools were predominantly composed of female students, with U1 consisting of 57.4% female and 42.5% male, U2 consisting of 60.5% female and 39.4% male, and U3 consisting of 51.5% female and 48.5% male. The results were in sharp contrast to the rural schools which were predominantly composed of male students. The rural school demographic consisted of R1 with 51.5% male and 48.5% female, R2 with 51.3% male and 48.7% female, and R3 with a narrow exception of 50.6% female and 49.4% male (See Table 2).

One of the key components considered by the current research was in the area of the ethnic composition between the urban and rural colleges. It was anticipated that ethnic differences existed between the student bodies of urban colleges when compared to rural colleges. As was predicted, most of the participants from the U1-college were Hispanic, Table 3. According to the Credit Student Enrollment Profile generated by the

U1-college, 72.1% of the student population was Hispanic, 16.4% Black Non-Hispanic, 6.6% White Non-Hispanic, 1.2% Asian, and 3.7% other or unreported (U1 Fact Book, 2018). The ethnic analysis of U2 for Fall 2018 also registered a majority of Hispanic students, with 40.2% being Hispanic, 30.8% White Non-Hispanic, 14.8% Black Non-Hispanic, 7.7% Asian/Pacific Islander, and 6.5% other or unreported (U2 Fast Facts, 2018). The ethnic analysis for the U3-college was different with 37.6% students being White Non-Hispanic, 23.2% Hispanic, 16.5% Asian, 15.9% Black Non-Hispanic, and 6.8% other or unreported (U3 Fact Book, 2017).

Table 3
Ethnic Demographic on Urban Colleges

	U1		U2		U3	
	#	%	#	%	#	%
White Non-Hispanic	5,620	6.6%	27,465	30.8%	19,272	37.6%
Hispanic	61,872	72.1%	35,875	40.2%	11,855	23.2%
Black Non-Hispanic	14,068	16.4%	13,152	14.8%	8,150	15.9%
Asian	1,010	1.2%	6,872	7.7%	8,422	16.5%
Other	643	0.7%	5,766	6.5%	2,633	5.1%
Unreported	2,546	3.0%	0	0.0%	858	1.7%
	85,759	100.00%	89,130	100.0%	51,190	100.0%

Regarding the ethnic composition of rural colleges, a majority of White Non-Hispanic student population was expected. As foreseen, the three rural colleges registered high in the percentage of White Non-Hispanic. Table 4. Student enrollment for the R1-college was 53.2% White Non-Hispanic, 12.2% Hispanic, 7.3% Black Non-Hispanic, 12.7% Asian, and 14.6% other or unreported (R1 Common Data Set, 2018). The ethnic analysis of R2 also registered a majority of White students, with 51.3% being White Non-Hispanic, 9.6% Hispanic, 6.2% Black Non-Hispanic, 20.4% Asian, and 12.5% other or unreported (R2 Fact Book, Fall 2018). The ethnic analysis for the R3-college was 60.6%

White Non-Hispanic, 10.2% Hispanic, 8.4% Black Non-Hispanic, 7.4% Asian, and 13.4% other or unreported (R3 Common Data Set, 2018).

Table 4
Ethnic Demographic on Rural Colleges

	R1		R2		R3	
	#	%	#	%	#	%
White Non-Hispanic	1,104	53.2%	3,371	51.3%	1,109	60.6%
Hispanic	253	12.2%	631	9.6%	189	10.2%
Black Non-Hispanic	153	7.3%	408	6.2%	147	8.4%
Asian	264	12.7%	1,342	20.4%	137	7.4%
Other	302	14.6%	381	5.8%	234	12.8%
Unreported	0	0.0%	438	6.7%	12	0.6%
	2,076	100.0%	6,571	100.0%	1,828	100.0%

Data Set Analysis

As previously mentioned, the objective of the study was to conduct a sentiment analysis on the Twitter activity of three urban and three rural community colleges. Once the colleges were selected, the acquisition of tweets posted on Twitter from these colleges microblogging activity was performed. The Twitter API provided access to the data set of tweets associated with a specific entity or Twitter user, for the present study the community colleges being analyzed were used, and thus was searched to acquire the tweets (Abdelrazeq et al., 2015; Pak and Paroubek, 2010; and Zhang et al., 2011). The data set of tweets was saved to a local file and yielded the following results:

Table 5
Raw Tweet Acquisition Values

College	Date Range		Activity Duration in Months	Average Monthly Tweets	# of Tweets
	Start	End			
U1	2/24/2016	6/5/2018	28	114.2	3200
U2	8/8/2013	6/5/2018	58	55.2	3204
U3	9/4/2015	6/5/2018	33	98.2	3242
R1	10/8/2014	6/5/2018	44	73	3213
R2	2/1/2016	6/5/2018	28	115.3	3230
R3	9/5/2017	6/5/2018	9	359.8	3238
Total				3221	19327

As noted in Table 5, the total quantity of tweets obtained from the six colleges was 19,327 tweets with an average number 3,221 tweets among the three urban and three rural colleges. The quantity of tweets acquired was mostly consistent among the colleges, although the Twitter activity or time span of tweets posted by the colleges for a similar quantity of tweets significantly varied for both the urban and rural colleges. Tweet acquisition for urban college U1 was at 3,200 tweets during a time span of 28 months or approximately 114.2 tweets per month, U2-college resulted in 3204 tweets during a 58-month period or approximately 55.2 tweets per month, and U3-college had 3,242 tweets during a 33-month period or approximately 98.2 tweets per month. U1-college registered the most activity, while U2-college had significantly less Twitter activity than the other two urban colleges.

Regarding the rural colleges, R1-college had 3,213 tweets during a 44-month period or approximately 73 tweets per month, R2-college had 3,230 tweets during a 28 month period or 115.3 tweets per month, and R3-college had 3,238 tweets during a 9 month period or approximately 359.8 tweets per month. R3-college had an overwhelming activity of tweets given the acquired monthly quantities when compared to all the other urban and rural colleges, while both R1 and R2-colleges had comparable tweet activity with R2 having slightly more tweet activity.

Model Building Process

To conduct a sentiment analysis on the Twitter corpus obtained from the three urban and three rural community colleges, it was first necessary to build or prepare a sentiment analysis model or classifier. A preprocessing step was performed on the training dataset, which refers to the Sentiment140 (2018) data source of 1.5 million

tweets that have previously been analyzed for sentiment. Preprocessing is the phase of cleaning up the corpus from unnecessary symbols, characters, hashtags, links, repeated letters, punctuations, emoticons, word abbreviations, stop-words, stemming, and retweets. Once the training dataset was prepared for analysis, the dataset was converted into a bag-of-words that was used to build three separate sentiment analysis models. The three classifier models generated were unigram, bigram and combogram. The unigram classifier is the partitioning of the training dataset into single word sentiments. The bigram classifier is the partitioning of the training dataset into double-words or two-word phrases. Finally, the combogram classifier is a cohort of the unigram and bigram classifiers.

A critical aspect of the process in establishing these three classifier models was the validation step, since it is necessary to validate that the model functions as it is intended in analyzing the sentiment of the corpus. The validation was accomplished by conducting a 10-fold cross validation on the models. The 10-fold cross validation process partitions the training data into ten different unique parts, and then cycles a comparison to verify how each part performs in evaluating the sentiment of all other parts. The outcome of the validation step is the average of the ten classified evaluation parts.

The results after each iteration produces a confusion matrix, which is a table that describes the performance of a classification model, used predominantly in the field of machine learning to visually evaluate the performance of an algorithm (Visa et al., 2011; Fawcett, 2006). The confusion matrix indicates the number of positive tweets and negative tweets incorrectly classified with a percentage indicating the overall sentiment

prediction accuracy. The percentage or test's accuracy is known as the F1-Score (also F-score or F-measure). A good F1-Score means that you have low false positives and low false negatives (Sasaki, 2007). An F1-Score of 1.0 is considered perfect accuracy or 100% accuracy. The higher the F1-Score, the better accuracy that the classifier model possesses in evaluating the sentiment of a corpus. Table 6 shows the Confusion Matrix Accuracy chart.

Table 6
Confusion Matrix Accuracy Chart

		Predicted Class	
		P	N
Actual Class	P	True Positives (TP)	False Positives (FN)
	N	False Positives (FP)	True Negatives (TN)

A confusion matrix with an F1-Score was obtained for each of the classifier models being validated. The unigram model demonstrated a 0.76 F1-Score, the bigram model had a 0.73 F1-Score, and the combogram model demonstrated the highest accuracy with a 0.78 F1-Score, as is observed in Table 7.

Table 7
Data Confusion Matrix

	Unigram Model		Average F1-Score
	P	N	
P	607244	187964	0.76
N	197023	597262	
	Bigram Model		
	P	N	
P	594452	200756	0.73
N	222672	571613	
	Combogram Model		
	P	N	0.78
P	627263	167945	
N	187137	607148	

All three classifier models demonstrated a high accuracy rate after the 10-fold cross validation step was concluded. Since model validation was obtained, the next step was to proceed with the preprocessing and feature extraction to the actual corpus obtained from all the urban and rural colleges. Each individual tweet from the corpus was then processed through each of the three classifier models: the unigram, bigram, and combogram classifiers.

The classifier models output a tweet sentiment analysis of positive or negative for each tweet from the corpus. The first analysis conducted was for the unigram classifier and the results generated for each of the three urban and the three rural schools. The same

analysis was then done using the bigram classifier. Finally, the process was repeated one last time using the combogram classifier.

Unigram Classifier

Table 8 below highlights the sentiment analysis obtained from applying the unigram classifier to the corpus of tweet data for both the urban colleges and rural colleges.

The total quantity of acquired tweets for U1 was 3,200 tweets. The tweet corpus was preprocessed, resulting in 3,157 tweets that were applied to the sentiment analysis using the unigram classifier. The sentiment analysis obtained for the U1 tweet corpus resulted in 2,343 positive tweets and 814 negative tweets. There were 43 other tweets that were discarded for lack of contextual data. For the U2 tweet corpus, which consisted of 3,204 acquired tweets, there were 2,173 positive tweets compared to 954 negative tweets out of the 3,127 tweets after the preprocessing phase. A total of 77 other tweets were discarded for lack of contextual data. The U3 tweet corpus, which consisted of 3,242 acquired tweets. There was a total of 3,141 tweets after the preprocessing phase, which resulted in 2,190 positive tweets and 951 negative tweets, with 101 other discarded tweets for lack of contextual data.

A review of the data indicates that urban colleges generated significantly higher quantities of positive sentiment tweets in comparison to the negative sentiment tweets. The outcome was consistent with the concept of relationship-building activities of the Dialogic Loop Theory (Kent & Taylor, 1998).

The same unigram classifier was executed on the rural colleges. The R1 tweet corpus, which consisted of 3,213 acquired tweets, yielded a sentiment analysis of 2,636

positive tweets and 539 negative tweets from the total 3,175 tweets after the preprocessing phase. There was a total of 38 tweets that were discarded for lack of contextual data. The R2 tweet corpus, which consisted of 3,230 acquired tweets, resulted in 2,595 positive tweets and 566 negative tweets, with 69 other discarded tweets for lack of contextual data out of a total 3,161 tweets after the preprocessing phase. The R3 tweet corpus consisted of 3,238 acquired tweets. There was a total of 3,224 tweets after the preprocessing phase, which produced 2,681 positive tweets and 543 negative tweets, with 14 other discarded tweets for lack of contextual data.

The data analysis indicates that rural colleges also generated significantly higher quantities of positive sentiment tweets in comparison to the negative sentiment tweets. Rural colleges also outperformed urban colleges in outputting appropriate levels of sentiment in their microblogging. Rural colleges demonstrated higher positive sentiment tweets and lower negative sentiment tweets than the urban colleges. Therefore, rural colleges' microblogging activities are consistent with Kent and Taylor's (1998) Dialogic Loop Theory for establishing relationship-building opportunities.

Table 8
Unigram Sentiment Classifier Data

	Positive	%	Negative	%	Total Processed Tweets	Acquired Tweets
U1	2343	74.22	814	25.78	3157	3200
U2	2173	69.49	954	30.51	3127	3204
U3	2190	69.72	951	30.28	3141	3242
R1	2636	83.02	539	16.98	3175	3213
R2	2595	82.09	566	17.91	3161	3230
R3	2681	83.16	543	16.84	3224	3238

Figure 11 visually presents the sentiment analysis results obtained from the unigram classifier for both the urban and rural colleges. The bar chart clearly presents the

high positive sentiment tweet output by both urban and rural colleges, and also highlights the tweet output differences in regard to positive and negative tweets between the two college types.

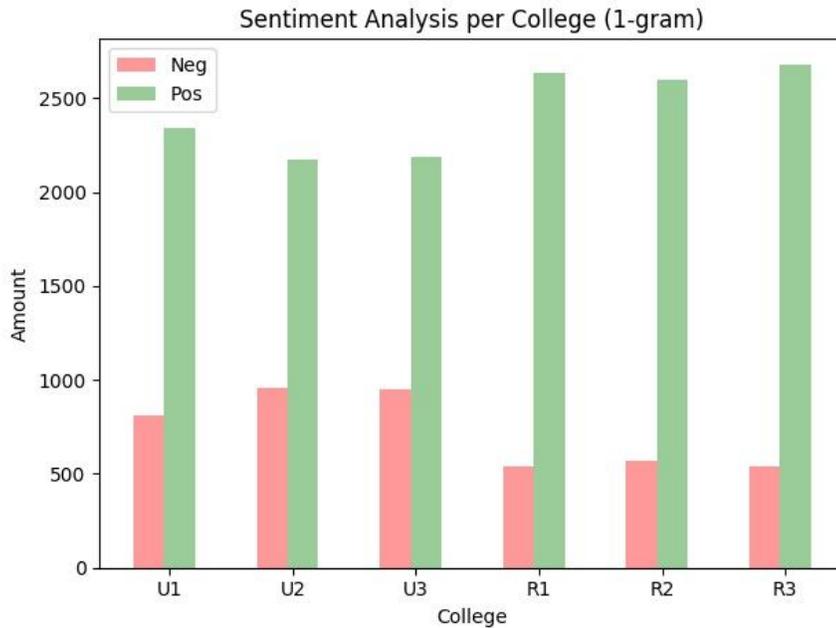


Figure 11. Sentiment Analysis per College Unigram

Bigram Classifier

The bigram classifier was subsequently executed on the college corpus for both the urban and rural colleges. Table 9 presents the sentiment analysis results obtained using the bigram classifier on all the college tweet corpus.

The sentiment analysis results of the bigram classifier for the U1 tweet corpus, which consisted of 3,200 acquired tweets, resulted in 2,345 positive tweets and 812 negative tweets, with 43 other discarded tweets for lack of contextual data. The total tweets analyzed was 3,157 after the preprocessing phase. There were 3,204 acquired tweets for the U2 tweet corpus with 3,127 tweets after the preprocessing phase. The

bigram classifier yielded 1,744 positive tweets compared to 1,383 negative tweets with 77 other discarded tweets for lack of contextual data. The U3 tweet corpus, which consisted of 3,242 acquired tweets, resulted in 1,784 positive tweets with 1,357 negative tweets out of a total 3,141 tweets after the preprocessing phase. There was also 101 other tweets that were discarded for lack of contextual data.

The sentiment analysis data for the urban colleges using the bigram classifier demonstrated a consistent pattern regarding positive sentiment and negative sentiment tweets for urban college U1 when compared to the unigram classifier results. Although, the bigram classifier yielded significantly lower positive sentiment tweets and significant higher negative sentiment tweets for the U2 and U3-colleges when compared to the unigram classifier results.

The bigram classifier results for the rural colleges, yielded a sentiment analysis for the R1 tweet corpus, which consisted of 3,213 acquired tweets, of 1,935 positive tweets and 1,240 negative tweets from the total 3,175 tweets after the preprocessing phase, with 38 other discarded tweets for lack of contextual data. The R2 tweet corpus, which consisted of 3,230 acquired tweets, resulted in 1,908 positive tweets and 1,253 negative tweets out of a total 3,161 tweets after the preprocessing phase, with 69 other discarded tweets for lack of contextual data. The R3 tweet corpus, which consisted of 3,238 acquired tweets, produced 2,026 positive tweets with 1,198 negative tweets out of a total 3,224 tweets after the preprocessing phase, with 14 other discarded tweets for lack of contextual data.

The sentiment analysis outcome using the bigram classifier for the rural colleges resulted in significantly lower positive sentiment and significantly higher negative

sentiment tweets for all the rural colleges when compared to the unigram classifier results.

Table 9
Bigram Sentiment Classifier Data

	Positive	%	Negative	%	Total Processed Tweets	Acquired Tweets
U1	2345	74.28	812	25.72	3157	3200
U2	1744	55.77	1383	44.23	3127	3204
U3	1784	56.80	1357	43.20	3141	3242
R1	1935	60.94	1240	39.06	3175	3213
R2	1908	60.36	1253	39.64	3161	3230
R3	2026	62.84	1198	37.16	3224	3238

The sentiment analysis data for each urban and rural college is presented in Figure 12. The bar chart presents a balanced outcome comparison for the sentiment of all colleges with comparable high positive sentiment and high negative sentiment tweet output by both urban and rural colleges.

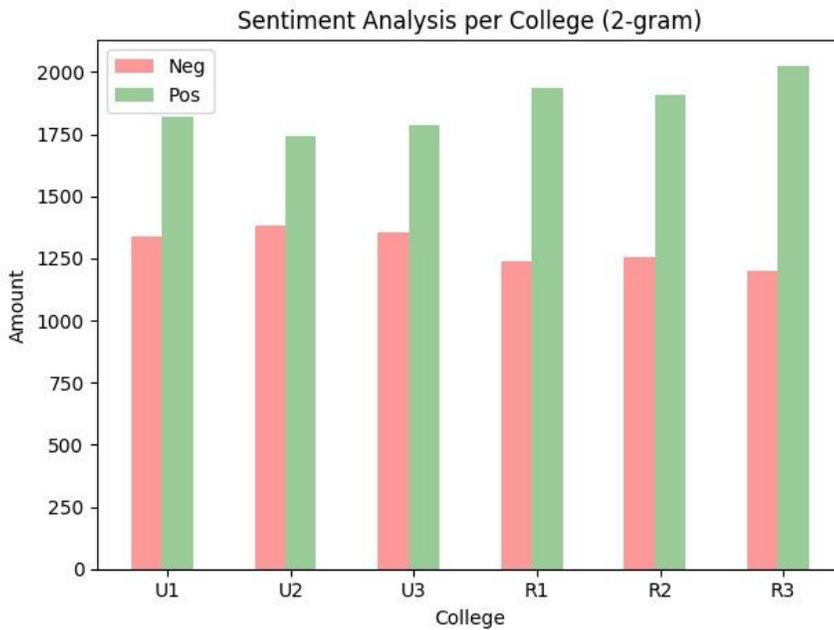


Figure 12. Sentiment Analysis per College Bigram

Combogram Classifier

The final step in the sentiment classification process was to run the combogram, which is the union of both the unigram and bigram, against the college corpus. The results of the urban and rural college sentiment analysis using the combogram classifier is presented in Table 10.

The sentiment analysis results for the U1 tweet corpus, which consisted of 3,200 acquired tweets, was 2,188 positive tweets and 969 negative tweets out of a total 3,157 tweets after the preprocessing phase, with 43 other discarded tweets for lack of contextual data. The U2 tweet corpus, which consisted of 3,204 acquired tweets, yielded 2,142 positive tweets compared to 985 negative tweets out of the 3,127 tweets after the preprocessing phase, with 77 other discarded tweets for lack of contextual data. The U3 tweet corpus, which consisted of 3,242 acquired tweets, resulted in 2,213 positive tweets with 928 negative tweets out of a total 3,141 tweets after the preprocessing phase, with 101 other discarded tweets for lack of contextual data.

The sentiment analysis outcome using combogram classifier for the urban colleges resulted in similar results to the sentiment analysis outcome using the unigram classifier. With the combogram classifier analysis, the urban colleges generated significantly higher quantities of positive sentiment tweets in comparison to the negative sentiment tweets.

The results for the rural colleges from the combogram classifier generated a sentiment analysis for the R1 tweet corpus. There were 3,213 acquired tweets, with a total of 3,175 tweets after the preprocessing phase. The sentiment analysis resulted in 2,605 positive tweets and 570 negative tweets, with 38 other discarded tweets for lack of

contextual data. The R2 tweet corpus, which consisted of 3,230 acquired tweets, resulted in 2,569 positive tweets and 592 negative tweets out of a total 3,161 tweets after the preprocessing phase, with 69 other discarded tweets for lack of contextual data. The R3 tweet corpus, which consisted of 3,238 acquired tweets, produced 2,610 positive tweets with 614 negative tweets out of a total 3,224 tweets after the preprocessing phase, with 14 other discarded tweets for lack of contextual data.

The combogram classifier yielded similar sentiment analysis outcome for all the rural colleges compared to the sentiment analysis outcome using the unigram classifier. With the combogram classifier analysis, the rural colleges generated slightly lower quantities of positive sentiment tweets and slightly higher negative sentiment tweets in comparison to the unigram classifier results. Regardless, the rural colleges generated significantly higher quantities of positive sentiment tweets in comparison to the negative sentiment tweets.

Table 10
Combogram Sentiment Classifier Data

	Positive	%	Negative	%	Total Processed Tweets	Acquired Tweets
U1	2188	69.31	969	30.69	3157	3200
U2	2142	68.50	985	31.50	3127	3204
U3	2213	70.46	928	29.54	3141	3242
R1	2605	82.05	570	17.95	3175	3213
R2	2569	81.27	592	18.73	3161	3230
R3	2610	80.96	614	19.04	3224	3238

The bar chart in Figure 13 demonstrates the sentiment analysis results obtained from the combogram classifier for both the urban and rural colleges. The bar chart presents the high positive sentiment tweet and lower negative sentiment tweet output by both urban and rural colleges. It should be noted that the chart clearly identifies the

differences between the two college types in regard to the positive sentiment and negative sentiment tweet output.

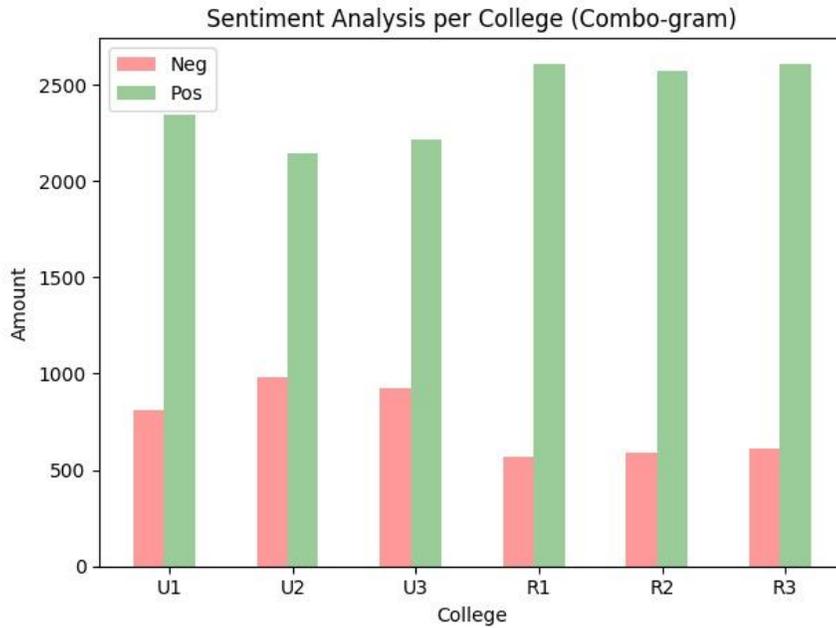


Figure 13. Sentiment Analysis per College Combogram

When considering the sentiment analysis output for the three classifiers: the unigram, bigram, and combogram, it is noted that the unigram classifier and the combogram classifier performed with similar sentiment analysis results across all the urban and rural colleges, with higher values for positive sentiment tweets compared to the negative sentiment tweets. It is also apparent that the two-word classifier or bigram generated a proportional sentiment analysis output across all the urban and rural colleges, with comparable sentiment analysis output for both the positive and negative tweets. Table 11 highlights the aforementioned differences among the different classifiers with the sentiment analysis output for each urban and rural college.

Table 11
Sentiment Analysis Results by Classifier

	Unigram Classifier			Bigram Classifier			Combogram Classifier		
	Positive	Negative	Total	Positive	Negative	Total	Positive	Negative	Total
U1	2343	814	3157	2345	812	3157	2188	969	3157
U2	2173	954	3127	1744	1383	3127	2142	985	3127
U3	2190	951	3141	1784	1357	3141	2213	928	3141
R1	2636	539	3175	1935	1240	3175	2605	570	3175
R2	2595	566	3161	1908	1253	3161	2569	592	3161
R3	2681	543	3224	2026	1198	3224	2610	614	3224

Table 12 presents a percentage breakdown of each classifier for the sentiment analysis of the college corpus. When only considering the unigram classifier between the three urban and three rural colleges, it is noted that the rural colleges consistently had significant higher positive sentiment tweets, ranking in the low 80s percentile and significant lower negative sentiment tweets that ranked in the tens percentile. The urban colleges positive sentiment tweets ranked in the high 60s and low 70s percentile while the negative sentiment tweets ranked in the mid-20s and low 30s percentile.

When considering the bigram classifier, the sentiment analysis among all the urban and rural colleges were somewhat consistent in regard to both the positive sentiment tweets and the negative sentiment tweets. The U2 and U3-colleges ranked in the high 50s percentile, with U1-college ranking in the mid-70s percentile for positive sentiment tweets. Regarding negative sentiment tweet ranking, U2 and U3-colleges ranked in the mid-40s percentile while U1-college ranked in the mid-20s percentile.

The rural colleges ranked in the low 60s percentile for positive sentiment tweets, with negative sentiment tweets ranking in the high 30s percentile. Although, the rural colleges did rank slightly higher in positive sentiment tweets than the urban colleges, with the sole exception of urban college U1, which registered a very high percentage of

positive sentiment tweets and a much lower negative sentiment tweets than all the other urban and rural colleges.

The combogram classifier produced the most consistent results among all the colleges in regard to the positive sentiment tweets and the negative sentiment tweets. All the rural colleges ranked in the 80s percentile for positive sentiment tweets while registering a negative sentiment tweet in the tens percentile. The urban colleges ranked in the high 60s and low 70s percentile for positive sentiment tweets, with negative sentiment tweets ranking in the high 20s and low 30s percentile. Similar to the other two classifiers, the combogram classifier also registered the rural colleges with a higher positive sentiment tweets and lower negative sentiment tweets than all the urban colleges.

Table 12
Sentiment Analysis Percentage Results by Classifier

	Unigram Classifier		Bigram Classifier		Combogram Classifier	
	% of Positive	% of Negative	% of Positive	% of Negative	% of Positive	% of Negative
U1	74.22	25.78	74.28	25.72	69.31	30.69
U2	69.49	30.51	55.77	44.23	68.50	31.50
U3	69.72	30.28	56.80	43.20	70.46	29.54
R1	83.02	16.98	60.94	39.06	82.05	17.95
R2	82.09	17.91	60.36	39.64	81.27	18.73
R3	83.16	16.84	62.84	37.16	80.96	19.04

Another important factor regarding the sentiment analysis of the college corpus was to analyze the various Twitter components of tweets. The analysis of the Twitter components pertained to research questions 4 and 5. These components refer to the retweets, likes, URL (Uniform Resource Locator) links, at signs (@), and hashtags (#) used within the tweets. The components are specifically significant for the adoption of the relationship building aspect of the Microblog Dialogic Communication framework.

The aggregate of each component is listed in Table 13 which contains the quantities for each college, while Table 14 provides a percentage analysis of these quantities. It is necessary to note that the amounts listed for the Total Count reflects the collective quantity of tweets for both the positive sentiment tweets and negative sentiment tweets obtained after the preprocessing phase. The tweets for two of the urban colleges, U1 and U2, received 4.51 and 5.88 retweets per tweet respectively, while rural college R2 was retweeted at 22,142 and received 7.00 retweets per tweet, which was the highest retweet frequency over all other colleges. It can be observed that overall, the urban colleges obtained higher retweet frequencies when compared to the rural colleges. With regard to the “likes” component, two of the rural colleges obtained high quantities of “likes” with R2 excelling over all other colleges with 41,154 “likes” or having 13.02 “likes” per tweet. Only urban college U1 received a high quantity of “likes” with 17,044 “likes” or 5.40 “likes” per tweet. The rural colleges received higher frequencies of “likes” in comparison to the urban colleges when looking at the data as a whole.

Table 13
Total Tweet Social Media Components

	Total Count	Total Retweets	Total Likes	Total URLs	Total At Signs (@)	Total Hashtags (#)
U1	3157	14239	17044	3221	3333	1906
U2	3127	18391	8970	2862	3925	2523
U3	3141	9037	5198	2628	2255	2600
R1	3175	7877	10719	2633	3115	1764
R2	3161	22142	41154	2763	4072	1730
R3	3224	7911	8305	2476	3128	1777

Table 14
Total Social Media Component Frequencies

	Count	Retweets Frequency	Likes Frequency	Urls Frequency	At Signs (@) Frequency	Hashtags (#) Frequency
U1	3157	4.51	5.40	1.02	1.06	0.60
U2	3127	5.88	2.87	0.92	1.26	0.81
U3	3141	2.88	1.65	0.84	0.72	0.83
R1	3175	2.48	3.38	0.83	0.98	0.56
R2	3161	7.00	13.02	0.87	1.29	0.55
R3	3224	2.45	2.58	0.77	0.97	0.55

With regard to the URLs, @ signs, and hashtags used, all of which are critical components for relationship building according to the Microblog Dialogic Communication framework, all the urban colleges and rural colleges incorporated these components in a large portion of their tweet activities for the test group assessed. The tweet frequency of use that incorporated URLs was relatively high for all urban and rural colleges, with U1 registering the highest at 1.02 URLs per tweet, followed by U2 with .92 URLs per tweet. The highest ranking rural college with URL frequency was R2 with .87 URLs per tweet.

The @ sign use also registered very high for tweets for all urban and rural colleges by almost registering an @ sign frequency use of 1.00 per tweet, with the exception of U3 which only had .72 @ signs per tweet. The highest registering college with @ signs in their tweets was R2, which had 1.29 @ signs per tweet, followed by U2 and U1 with 1.26 and 1.06 @ signs per tweet respectively.

With regard to hashtags in tweets, all the rural colleges registered a frequency use in the range of .50 hashtags per tweet. In other words, only one hashtag for every two tweets, which was significantly lower than the urban colleges, with U1 registering the

lowest frequency use of hashtags of .60 hashtag per tweet, U2 with .81 hashtags per tweet, and U3 with the highest hashtag frequency of .83 hashtags per tweet.

Overall, all the urban and rural colleges incorporated most of the social media components of URLs, @ signs, and hashtags within their tweets with rural college R2 registering the highest use of @ signs in their tweets with a quantity of 4,072 or frequency of use of 1.29 @ signs per tweet, while also registering the lowest hashtag use among all the colleges with a quantity of 1,730 or .55 hashtags per tweet. The use of the social media components was proportional among all the urban and rural colleges, with minor outlier exceptions.

Another significant aspect to consider was obtained by factoring in the sentiment analysis as part of the tweet component analysis. In Table 15, the retweet component is evaluated on the basis of the sentiment where the retweet occurred using the unigram classifier. Therefore, when considering urban college U1, it can be noted that the 2,343 positive sentiment tweets received 10,249 retweets or they were retweeted 4.37 times per tweet compared to the 814 negative sentiment tweets that were retweeted 3,903 times or 4.79 times per tweet. The majority of urban and rural colleges obtained a proportional frequency count of positive sentiment tweets that were retweeted in the 2.00 range of retweets per tweet. The exceptions were urban college U1 as previously noted, and rural college R2, which registered the highest positive sentiment tweets that were retweeted with 17,671 retweets or a frequency of 6.81 retweets per tweet.

Table 15
 Percentage of Tweet Social Media Components for Unigram Classifier

	Unigram					
	Positive			Negative		
	Count	Retweets	Frequency Count	Count	Retweets	Frequency Count
U1	2343	10249	4.37	814	3903	4.79
U2	2173	5798	2.67	954	12446	13.05
U3	2190	4763	2.17	951	4207	4.42
R1	2636	6544	2.48	539	1273	2.36
R2	2595	17671	6.81	566	4015	7.09
R3	2681	6768	2.52	543	1116	2.06

Regarding the negative sentiment tweets that were retweeted, urban colleges U2 and U3 obtained significantly high negative sentiment tweets at 954 and 951 respectively, with urban college U2 receiving the highest quantity of retweets at 12,446 or a frequency count of 13.05 retweets per tweet, followed by U3 with 4,207 retweets or 4.42 retweets per tweet. Urban college U1 had a high quantity of negative sentiment tweets at 814 with a moderate quantity of retweets at 3,903, but at a higher frequency count of 4.79 retweets per tweet than urban college U3 values. Rural college R2 registered a quantity of negative sentiment tweets of 566, which was comparable to the quantities registered for R1 and R3 at 539 and 543 respectively, but with a significantly higher quantity and frequency count of negative sentiment retweets with 4015 or 7.09 retweets per tweet. The other two rural colleges, R1 and R3, both registering low quantities and frequency counts of negative sentiment tweets that were retweeted in the 2.00 range of retweets per tweet.

Table 16 presents the retweet component on the basis of the sentiment where the retweet occurred using the bigram classifier. The results for the positive sentiment tweets that received a retweet were similar to the unigram classifier outcomes with urban college U1 and rural college R2 receiving the highest quantity of retweets at 7,816 or a frequency count of 4.29 retweet per tweet and 11,641 or 6.10 retweet per tweet, respectively. All the

other urban and rural colleges registered a proportional frequency count of positive sentiment tweets that were retweeted in the 2.00 range of retweets per tweet.

Table 16
Percentage of Tweet Social Media Components for Bigram Classifier

	Bigram					
	Positive			Negative		
	Count	Retweets	Frequency Count	Count	Retweets	Frequency Count
U1	1821	7816	4.29	1336	6336	4.74
U2	1744	5158	2.96	1383	13086	9.46
U3	1784	5104	2.86	1357	3866	2.85
R1	1935	4536	2.34	1240	3281	2.65
R2	1908	11641	6.10	1253	10045	8.02
R3	2026	5160	2.55	1198	2724	2.27

Regarding the negative sentiment tweets that were retweeted, both the urban and rural colleges had consistent negative sentiment tweets ranging from 1,198 to 1,383. Although urban college U2 and rural college R2 obtained significantly high negative sentiment retweets with a quantity of 13,076 retweets with a frequency count of 9.46 retweets per tweet and 10,045 retweets with a frequency count of 8.02 retweets per tweet, respectively. Urban college U1 also registered a high level of negative sentiment retweets with 6,336 or 4.74 retweets per tweet, while the other urban college U3 and rural colleges R1 and R3 ranged at a quantity of negative sentiment retweets at 3,866, 3,281, and 2,724 respectively, all registering with a frequency count in the 2.00 range of retweets per tweet.

Table 17 presents the retweet component on the basis of the sentiment where the retweet occurred using the combogram classifier. The positive sentiment tweets for all urban and rural colleges were similar in the quantity count, ranging between 2,142 and 2,610. The results for the positive sentiment tweets that received a retweet were significantly higher for urban college U1 with 9,326 or a frequency count of 3.98

retweets per tweet and rural college R2 with 17,345 or a frequency count of 6.75 retweets per tweet. The other urban and rural colleges were similar in retweet count and all registered a frequency count in the 2.00 range of retweets per tweet.

Table 17
Percentage of Tweet Social Media Components for Combogram Classifier

	Combogram					
	Positive			Negative		
	Count	Retweets	Frequency Count	Count	Retweets	Frequency Count
U1	2345	9326	3.98	812	4826	5.94
U2	2142	5889	2.75	985	12355	12.54
U3	2213	4747	2.15	928	4223	4.55
R1	2605	6263	2.40	570	1554	2.73
R2	2569	17345	6.75	592	4341	7.33
R3	2610	6602	2.53	614	1282	2.09

With regard to the negative sentiment tweets, the three urban colleges registered high tweet counts with U1 at 812, U2 at 985, and U3 at 925, while the rural colleges obtained lower negative sentiment tweets of R1 at 570, R2 at 592, and R3 at 614. Urban college U2 obtained the highest level of negative sentiment retweets with 12,355 with a frequency count of 12.54 retweets per tweet, followed by U1 with 4,826 retweets or 5.94 retweets per tweet, and U3 with 4,223 retweets or 4.55 retweets per tweet. The highest negative sentiment retweets for the rural colleges was R2 with 4,341 retweets with a frequency count of 7.33 retweets per tweet, while the other two rural colleges registered much lower retweets with R1 at 1,554 and R3 at 1,282 and a negative sentiment retweet frequency count ranging in the 2.00 retweets per tweet.

The final aspect considered for the present research, specific to the Microblog Dialogic Communication framework, are the social media components for followers, friends, and favorites. Table 18 presents the aggregate of each of these components. Most all the colleges registered a high follower count, with U1 being the highest for the urban

colleges at 19,437 followers, U2 with 12,140, and U3 with the lowest count of all the colleges at 6,851. The rural colleges all ranked higher than the urban colleges. Rural college R2 had a significant higher follower count than all other colleges with 60,125, while R1 had 15,850 and R3 with 13,035 followers. There was a significant disparity in regard to the “friends” component with all urban colleges registering extremely low counts with U1 at 163, U2 at 306, and U3 at 543 friends, when compared to the rural colleges which all ranked at higher levels with R1 at 2,478, R2 at 2,538, and R3 at 2,168 friends. The favorites or like component yielded contrasting results among all the colleges. The urban colleges ranged from a significantly low 1,699 for urban college U2 to a very high 10,375 for urban college U1, while U3 had 3,075 favorites. Rural college R1 ranked at 2,563, R2 at 7,959, and R3 with the highest quantity than all other colleges at 16,689 favorites.

Table 18
Total Twitter Social Media Interaction

	Followers	Friends	Favorites
U1	19437	163	10375
U2	12140	306	1699
U3	6851	543	3075
R1	15850	2478	2563
R2	60125	2538	7959
R3	13035	2168	16689

The aforementioned data analysis values are all derived from the three different sentiment classifiers, which are labeled unigram, bigram, and combogram. When considering the sentiment analysis values obtained from these three sentiment classifiers, it should be noted that the combogram provided the most consistent results. Although, the unigram and bigram both outputted reliable values regarding the positive and negative sentiment for the microblogging activity of all the urban and rural colleges. It is noted that

the combogram classifier, which utilizes a combination method that analyzes the data with both the unigram and bigram arrangement of words, generated equilibrium values considered more reliable and consistent. Thus, the subsequent research analysis is based solely on the sentiment analysis values obtained from the combogram classifier.

Research Analysis

In regard to the research questions for the present study, the sentiment analysis data was used to gather a greater understanding of the urban and rural colleges' ability to establish dialogic communication with its public for relationship building purposes. A key component is garnished by obtaining the microblog activity sentiment, whether positive or negative. The data analysis from the current research was used specifically to comprehend the following research questions:

1. What is the microblogging sentiment of an urban college that interacts with a diverse ethnic student population?
2. What is the microblogging sentiment of a rural college that interacts with a homogeneous student population?
3. How does the microblogging activity of urban colleges differ from rural colleges with regard to their microblogging sentiment?"
4. How does the microblogging activities of a college instigate positive outcomes with regard to the sentiment of retweets and comments?
5. How well does the Microblog Dialogic Communication framework explain differences that exist on relationship building capabilities in the Twitter activity between urban colleges and rural colleges?

Results for Research Question # 1

The first research question was “What is the microblogging sentiment of an urban college that interacts with a diverse ethnic student population?”

The gender data analysis of the urban college's student enrollment as seen in Table 19, revealed that the urban colleges had an average student population of 75,359, which consisted of 43.47% male and 56.47% female.

Table 19
Urban Colleges Student Enrollment Totals/Averages by Gender

	Male	%	Female	%	Unreported	%	Total Enrolment
U1	36,484	42.5	49,191	57.4	84	0.1	85,759
U2	35,150	39.4	53,980	60.5	0	0	89,130
U3	24,807	48.5	26,383	51.5	0	0	51,190
Average	32,147	43.47	43,185	56.47	84	0.10	75,360

The ethnic composition of the urban colleges was predicted to consist of a diverse cultural population, which was demonstrated through the ethnic demographic data in Table 20. The student demographic data consisted of a large Hispanic population at 45.2%, Black Non-Hispanic at 15.7%, White Non-Hispanic at 25.0% and Asian at 8.5%.

Table 20
Student Demographic Data by Urban College

	U1		U2		U3		Average	
	#	%	#	%	#	%	#	%
White Non-Hispanic	5,620	6.6%	27,465	30.8%	19,272	37.6%	52,357	25.0%
Hispanic	61,872	72.1%	35,875	40.2%	11,855	23.2%	109,602	45.2%
Black Non-Hispanic	14,068	16.4%	13,152	14.8%	8,150	15.9%	35,370	15.7%
Asian	1,010	1.2%	6,872	7.7%	8,422	16.5%	16,304	8.5%
Other	643	0.7%	5,766	6.5%	2,633	5.1%	9,042	4.1%
Unreported	2,546	3.0%	0	0.0%	858	1.7%	3,404	1.6%
Total	85,759	100%	89,130	100%	51,190	100%	226,079	100%

Figure 14 presents the demographic population for all the urban colleges. The bar chart clearly presents the overwhelming levels of Hispanic students for both U1 and U2-colleges, along with the other represented demographic levels.

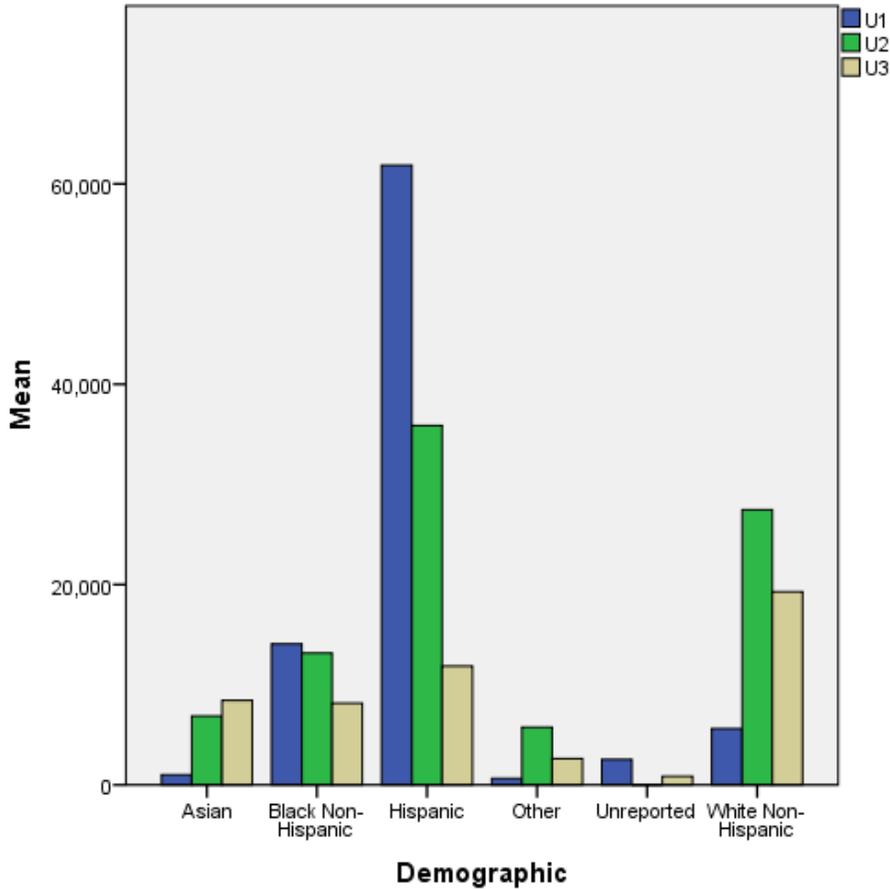


Figure 14. Demographic Data by Urban College

With regard to the microblogging sentiment of urban colleges, a total of 9,425 tweets were analyzed with 69.42% being of positive sentiment and 30.58% of negative sentiment. According to Table 21, the urban colleges registered on average 2,181 positive sentiment tweets with 961 negative sentiment tweets. The fact that urban colleges consist of a diverse ethnic student population could make it harder to structure microblogging messages that are widely accepted or are perceived as positive by a majority of students.

The diverse ethnic student population could be the cause for the higher negative sentiment tweets that urban colleges demonstrated. In other words, the microblogging sentiment of an urban college that interacts with a diverse ethnic student population was only 38.84% more positive than negative.

Table 21
Urban Colleges Combogram Classifier Tweet Sentiment Totals/Averages

	Positive	%	Negative	%	Total
U1	2188	69.31	969	30.69	3157
U2	2142	68.50	985	31.50	3127
U3	2213	70.46	928	29.54	3141
Totals	6543		2882		9425
%	69.42		30.58		
Average	2181		961		3142

Results for Research Question # 2

The second research question was “What is the microblogging sentiment of a rural college that interacts with a homogeneous student population?”

When considering the rural colleges, Table 22, the data analysis revealed that the average student population was 3,480, which consisted of 50.80% male and 49.20% female.

Table 22
Rural Colleges Student Enrollment Totals/Averages by Gender

	Male	%	Female	%	Unreported	%	Total Enrolment
R1	1,051	51.5	991	48.5	0	0	2,042
R2	3,385	51.5	3,186	48.5	0	0	6,571
R3	903	49.4	925	50.6	0	0	1,828
Average	1,780	50.80	1,701	49.20	0	0.00	3,480

As predicted, the rural colleges’ ethnic composition was much more homogeneous than the urban colleges. As seen in Table 23, there was a majority of White

Non-Hispanic at 55.0%. Other ethnic demographic representations for rural colleges included Asian at 13.5%, Hispanic at 10.7%, Black Non-Hispanic at 7.3%, and Other ethnicity at 11.1% student populations.

Table 23
Student Demographic Data by Rural College

	R1		R2		R3		Average	
	#	%	#	%	#	%	#	%
White Non-Hispanic	1,104	53.2%	3,371	51.3%	1,109	60.6%	5,584	55.0%
Hispanic	253	12.2%	631	9.6%	189	10.2%	1,073	10.7%
Black Non-Hispanic	153	7.3%	408	6.2%	147	8.4%	708	7.3%
Asian	264	12.7%	1,342	20.4%	137	7.4%	1,743	13.5%
Other	302	14.6%	381	5.8%	234	12.8%	917	11.1%
Unreported	0	0.0%	438	6.7%	12	0.6%	450	2.4%
Total	2,076	100.0%	6,571	100.0%	1,828	100.0%	10,475	100.0%

Figure 15 presents the demographic population for all the rural colleges. The bar chart clearly presents the overwhelming levels of White Non-Hispanic students for all three colleges, along with the other represented demographic levels.

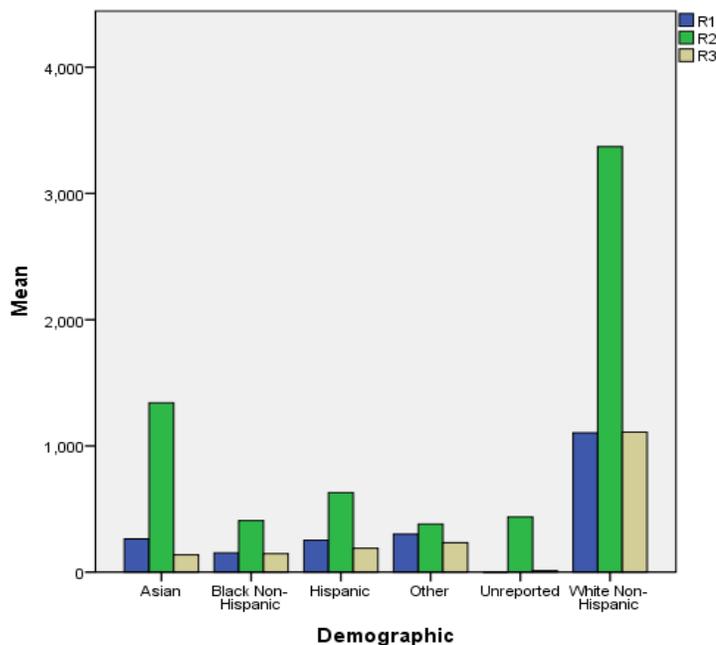


Figure 15. Demographic Data by Rural College

With regard to the microblogging sentiment of rural colleges, a total of 9,560 tweets were analyzed with 81.42% being of positive sentiment and 18.58% of negative sentiment. As is noted in Table 24, the rural colleges registered on average 2,595 positive sentiment tweets with 592 negative sentiment tweets. Therefore, rural colleges, on average, tweeted a significant portion of positive tweets, 81.42%, when compared to negative tweets, which was only 18.58%. A possibility for the high percentage of positive tweets could be attributed to the homogenous student population that rural colleges interact with when microblogging. As a result of the singularity of the student demographic, it's feasible that it is easier to structure microblogging messages that are better received by the student population, due to minimal ethnic diversity. In other words, the microblogging sentiment of a rural college that interacts with a homogeneous ethnic student population was 62.84% more positive compared to negative.

Table 24

Rural Colleges	Combogram Classifier		Tweet Sentiment		Totals/Averages
	Positive	%	Negative	%	
R1	2605	82.05	570	17.95	3175
R2	2569	81.27	592	18.73	3161
R3	2610	80.96	614	19.04	3224
Totals	7784		1776		9560
%	81.42		18.58		
Average	2595		592		3187

Results for Research Question #3

The third research question was “How does the microblogging activity of urban colleges differ from rural colleges with regard to their microblogging sentiment?”

The underlining premise for the research conducted in the current study is that the sentiment of urban colleges will differ from the sentiment of rural colleges, given the

complexity of the multi-cultural aspects and the size of urban colleges. As is noted in Table 25, the quantity of tweets analyzed for all urban colleges and all rural colleges was analogous with the aggregate of urban college tweets at 9,425 and the aggregate of rural colleges at 9,560. The data analysis revealed that on average the rural colleges generated higher positive sentiment tweets than the urban colleges. With regard to negative sentiment tweets, the rural colleges generated lower negative sentiment tweets than the urban colleges.

Specifically, the rural colleges averaged 2,595 or 81.42% positive sentiment tweets, while the urban colleges averaged 2,181 or 69.42% positive sentiment tweets. With respect to the negative sentiment tweets, the rural colleges averaged 592 or 18.58% negative sentiment tweets and the urban colleges averaged 961 or 30.58% negative sentiment tweets. In other words, the microblogging sentiment of a rural college had 11.9% more positive tweets than urban colleges. Also, when comparing the difference between the positive and negative tweets, rural colleges had 63.84% more positive than negative tweets compared to urban colleges, which had 39.84% more positive than negative tweets. The comparison demonstrates that rural colleges tended to have higher positive tweets compared to negative tweets.

Table 25
Urban and Rural College Combogram Classifier Average Sentiment Analysis

	Total Tweets	Average Positive Sentiment Tweets	%	Average Negative Sentiment Tweets	%
Urban	9425	2181	69.42	961	30.58
Rural	9560	2595	81.42	592	18.58

A Pearson's Chi-Square analysis was conducted on the college type and the tweet sentiment variables. As noted in Table 26, the urban colleges underperformed in regard to generating positive sentiment tweets as determined by the chi square expected count, while the rural colleges over performed in regard to the positive sentiment tweets. When considering the negative sentiment tweets, the urban colleges over performed as determined by the chi square expected count, while the rural colleges underperformed in negative sentiment tweet category.

Table 26
Sentiment Type and College Type Variables Cross Tabulation

			College Type					
			Urban	%	Rural	%	Total	
Sentiment	Positive	Count	6528	69.5%	7784	81.4%	14312	75.5%
		Expected	7093		7218		14312	
	Negative	Count	2867	30.5%	1776	18.6%	4643	24.5%
		Expected	2301		2341		4643	
Total	Count		9395	100.0%	9560	100.0%	18955	100.0%
	Expected		9395		9560		18955	

Pearson Chi-Square=365.176, $p < .000$ (statistically significant)

Pearson's Chi-Square Test resulted in 365.176 and a statistically significant p value of .000 between the college type (urban or rural) and the tweet sentiment (positive or negative). Rural colleges tended to have higher positive sentiment tweets compared to urban colleges, which tended to have lower positive sentiment tweets. Additionally, rural colleges also tended to have lower negative sentiment tweets compared to urban colleges that tended to have higher negative sentiment tweets.

It is possible that the colleges' student demographics has a direct association with the tweet sentiment generated by the urban or rural colleges. The elevated positive

sentiment tweets of rural colleges may be attributed to a more homogenous student population, since the tweets target a larger like-minded student population. The opposite is also true for urban colleges, which experienced higher negative sentiment tweets, given that urban colleges consist of a diverse student population, which require more inclusive style tweets. The results set the groundwork for further research to ascertain if urban colleges require more inclusive style tweets.

In other words, most of the rural colleges have a propensity to tweet more positive and less negative sentiment tweets than most of the urban colleges. Below is Figure 16, which illustrates the differences that exist between the urban colleges and the rural colleges in regard to the positive sentiment and negative sentiment tweets in comparison to the expected sentiment tweet counts.

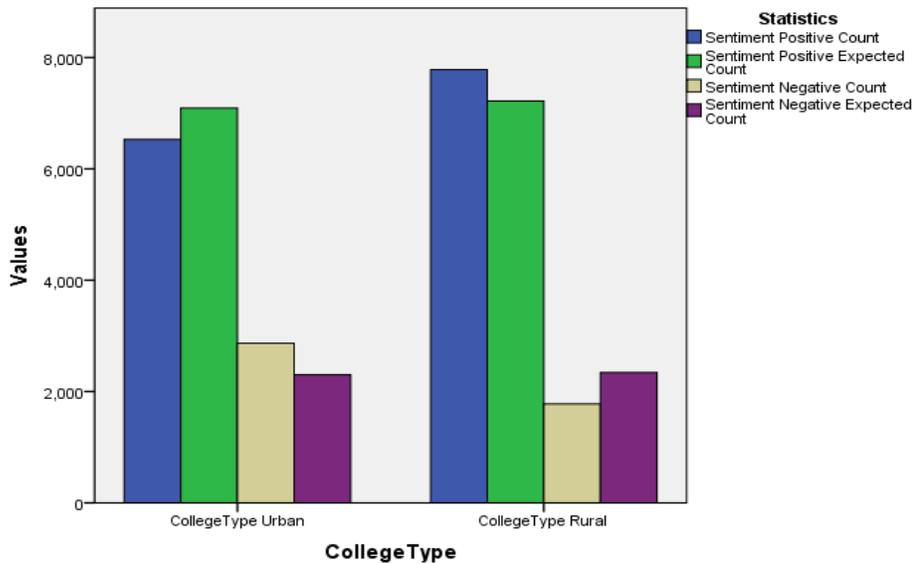


Figure 16. Comparison of Urban vs Rural Tweet Sentiment

The chart clearly presents the tweet sentiment output differences between the two college types. Urban colleges underperformed in regard to the positive sentiment tweets while outperformed in regard to the negative sentiment tweets with respect to the

expected values. The outcome represents unfavorable microblogging activity for urban colleges since the positive sentiment tweets were lower than expected while the negative sentiment tweets were higher than expected. Contrary to the urban colleges, rural colleges outperformed in regard to the positive sentiment tweets while underperformed in regard to the negative sentiment tweets with respect to the expected values. The results represents a favorable microblogging activity for rural colleges since the positive sentiment tweets were higher than expected while the negative sentiment tweets were lower than expected.

Results for Research Question #4

The fourth research question was “How does the microblogging activities of a college instigate positive outcomes with regard to the sentiment of retweets and comments?”

Relationship-building is enacted by dialogic practices through positive interactions with an organizations’ public. Within the Twitter capacity, the dialogic aspects are identified by the number of retweets, comments, or any form of feedback or interaction that is obtained when a tweet is posted. The data analysis revealed that rural colleges significantly outperformed the urban colleges. Rural colleges, on average, had higher positive sentiment retweets and considerably lower negative sentiment retweets than the urban colleges.

Specifically, as noted in Table 27, urban colleges tweeted 9,425, of which 6,700 were ranked as positive sentiment tweets. These positive sentiment tweets received 19,962 retweets, thus each positive tweet obtained 2.98 retweets. On average, the urban colleges tweeted 2,223 positive tweets, which received 6,654 retweets or each tweet received 2.96 retweets. With regard to negative sentiment tweets, urban colleges had 2,725 tweets that

were ranked as negative sentiment. These negative sentiment tweets obtained 21,404 retweets, thus each negative tweet was retweeted 7.85 times. Overall, the urban colleges tweeted 908 negative sentiment tweets, which received 7,135 retweets or each tweet received 7.68 retweets.

Table 27
Urban Retweet Count

College Type	Tweet Sentiment and Retweet Count					
	Tweet Count	Positive		Negative		
		Retweets	Retweet Frequency	Tweet Count	Retweets	Retweet Frequency
U1	2345	9326	3.98	812	4826	5.94
U2	2142	5889	2.75	985	12355	12.54
U3	2213	4747	2.15	928	4223	4.55
Totals	6700	19962	2.98	2725	21404	7.85
Average	2233	6654	2.96	908	7135	7.68

Table 28 contains details on the rural colleges, which had approximately a similar tweet total, with 9,560 tweets, and slightly higher positive sentiment tweet total, with 7,784 positive sentiment tweets, when compared to the urban colleges. But contrary to the urban colleges, the positive sentiment tweets for rural colleges received much higher retweets with 30,210 retweets, thus each positive tweet was retweeted 3.88 times. On average, the rural colleges tweeted 2,595 positive tweets, which received 10,070 retweets or each tweet received 3.90 retweets. The rural colleges had 1,776 negative sentiment tweets which obtained 7,177 retweets, thus each negative tweet was retweeted 4.04 times. The quantity of negative sentiment tweets that were retweeted for rural colleges was much lower than those for urban colleges. Overall, the rural colleges tweeted 592 negative sentiment tweets, which received 2,392 retweets or each tweet received 4.05 retweets.

Table 28
Rural Retweet Count

College Type	Tweet Sentiment and Retweet Count					
	Positive			Negative		
	Tweet Count	Retweets	Retweet Frequency	Tweet Count	Retweets	Retweet Frequency
R1	2605	6263	2.40	570	1554	2.73
R2	2569	17345	6.75	592	4341	7.33
R3	2610	6602	2.53	614	1282	2.09
Totals	7784	30210	3.88	1776	7177	4.04
Average	2595	10070	3.90	592	2392	4.05

A Pearson's Chi-Square Test, Table 29, was conducted on the college type (urban or rural) and the retweet count sentiment (positive or negative). The test yielded a statistical significance between the college type and the tweet sentiments that received retweets. The retweet analysis was similar to the tweet sentiment analysis results, with rural colleges tending to have higher retweets for the positive sentiment tweets compared to urban colleges, which had significantly lower retweets of positive sentiment tweets. Even more significant were the analysis for the retweets on the negative sentiment tweets, in which rural colleges had considerably lower retweets of negative sentiment tweets than the urban colleges that experienced dramatically high retweets of negative sentiment tweets.

Table 29
Retweet Count Sentiment and College Type Variables Cross Tabulation

			CollegeType					
			Urban	%	Rural	%	Total	
Retweet	Positive	Count	19962	48.3%	30210	80.8%	50172	63.7%
		Expected Count	26354		23819		50172	
	Negative	Count	21404	51.7%	7177	19.2%	28581	36.3%
		Expected Count	15013		13569		28581	
Total	Count		41366	100.0%	37387	100.0%	78753	100.0%
	Expected Count		41366		37387		78753	

Pearson Chi-Square=8997.049, $p < .000$ (statistically significant)

These outcomes can be attributed to the differences in school size from urban and rural colleges. Given that the urban colleges analyzed for the current study had extensive student populations when compared to the rural colleges, the tweet composition for the urban colleges may consider a cultural diversity component that requires a more inclusive type message that is appealing to a broader ethnic student population. Tweets that are less inclusive may generate more negative attention from the student population, which in turn might garnish an increase level of retweets. The rural colleges target a more homogenous student demographic that allows the colleges the opportunity to structure their tweets with less inclusive requirements and which are specific to the majority of their student population. Furthermore, the college size and location also influence student following and participation. It is likely that smaller rural colleges are located in less populace areas which have fewer distractions for students, thus the tweets generated by rural colleges have a greater influence and garnish greater attention than those by urban colleges, which most likely are situated in large cities which have a multitude of events,

activities, and other student distractions. Further research could be conducted on the aforementioned tweet styles to obtain a better understanding on the urban and rural student tweet appeal and retweet trend.

In other words, most of the rural colleges have a propensity to retweet more positive and less negative sentiment tweets than most of the urban colleges. Below is Figure 17, which illustrates these differences:

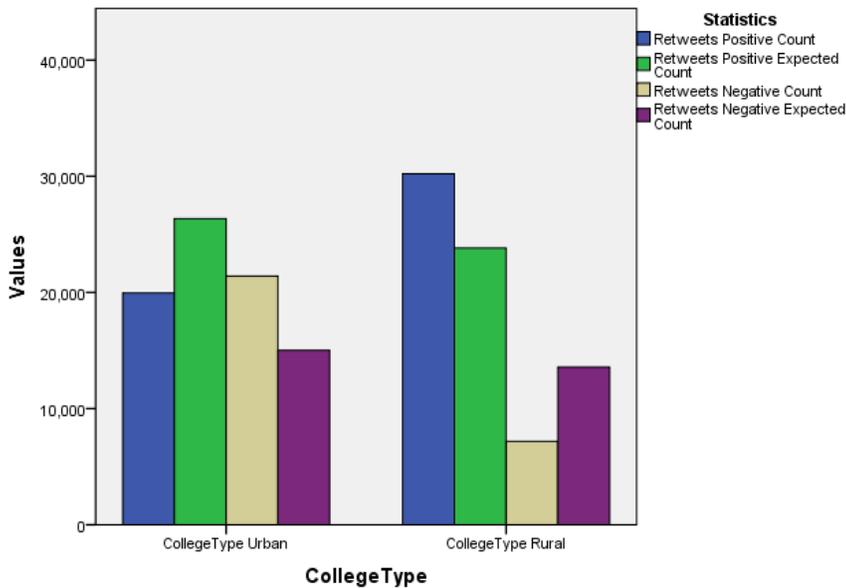


Figure 17. Comparison of Urban vs Rural Retweet Sentiment

The chart presents a comparison of the retweets for positive sentiment tweets and negative sentiment tweets between the two college types. The results for the retweet counts using sentiment were similar to the tweet sentiment analysis outcomes for both college types. Urban colleges underperformed in regard to retweets for positive sentiment tweets while outperformed in regard to retweets for negative sentiment tweets with respect to the expected values. The outcome means that less people retweeted the positive sentiment tweets while more people retweeted the negative sentiment tweets. The results

for urban college retweets represent an unfavorable microblogging activity since the positive sentiment retweets were lower than expected for positive sentiment tweets while higher than expected for the negative sentiment tweets. In direct contrast to the urban colleges, rural colleges outperformed in regard to retweets for the positive sentiment tweets while underperformed in regard to retweets for the negative sentiment tweets with respect to the expected values. The rural college results signify that more people retweeted the positive sentiment tweets while less people retweeted the negative sentiment tweets. The results indicate a favorable microblogging activity for rural colleges since the retweets for the positive sentiment tweets were higher than expected while lower than expected for the negative sentiment tweets.

Results for Research Question # 5

The fifth research questions was “How well does the Microblog Dialogic Communication framework explain differences that exist on relationship building capabilities in the Twitter activity between urban colleges and rural colleges?”

To maximize the relationship-building capabilities of Twitter, social media communication needs to include specific components detailed in the Microblog Dialogic Communication framework. As such, the current study considers the elements of social media communication for the urban and rural colleges. The data analysis used for the Microblog Dialogic Communication framework assessment did not apply any classifier, rather the raw data was considered to accurately assess the colleges’ performance.

According to Table 30, the analysis revealed that all colleges incorporated the necessary social media communication elements that include URL links, at signs (@), and hashtags (#) used within the tweets. Specifically, the urban colleges registered 9,425

tweets which contained 8,711 URL links, 9,513 "at signs" (@), and 7,029 hashtags (#). These tweets obtained a total of 41,667 retweets and 31,212 "likes". The urban colleges had a total of 38,428 followers with 1,012 friends. The rural colleges registered 9,560 tweets which contained 7,872 URL links, 10,315 "at signs" (@), and 5,271 hashtags (#). These tweets obtained a total of 37,930 retweets and 60,178 "likes". The rural colleges had a total of 89,010 followers with 7,184 friends.

Table 30
Aggregate Social Media Components by College Type

	Count	Retweets	Likes	URLS	@	#	Followers	Friends	Favorites
Urban	9425	41667	31212	8711	9513	7029	38428	1012	15149
Rural	9560	37930	60178	7872	10315	5271	89010	7184	27211

Considering the Microblog Dialogic Communication framework, urban colleges had higher use of social media communication elements than the rural colleges with higher use of URL links and hashtags. It is important to note that hashtags are used specifically to categorize or draw attention to a specific topic or theme. The urban colleges also had a higher retweet count than rural colleges despite having significantly less followers and friends. In contrast, rural colleges used slightly higher "at signs" (@), which are used to tag other Twitter account users and specifically notify them of the tweet. Rural colleges obtained a significant higher quantity of "likes" than the urban colleges and had significantly higher followers and friends.

Likes Component

As noted in Table 31, descriptive statistics obtained from the colleges use of social media components revealed that rural colleges had a significant higher "likes" frequency count than urban colleges. When considering the colleges separately, the

majority of the rural colleges experienced higher “likes” than the urban colleges, with R2 obtaining a significant frequency count above all the colleges.

Table 31
Frequency Totals for Component: Likes by College Type

College Type	Frequency	Percent	Aggregate Frequency	Aggregate Percent
Urban 1	17044	54.6		
Urban 2	8970	28.7		
Urban 3	5198	16.7		
Total			31212	34.2
Rural 1	10719	17.8		
Rural 2	41154	68.4		
Rural 3	8305	13.8		
Total			60178	65.8
Total Urban & Rural			91390	100

Note: Aggregate Percent = Aggregate Frequency / Total Urban & Rural Aggregate Frequency times 100

Figure 18 illustrates the aggregate levels using the “likes” social media component, while Figure 19 illustrates the frequency count of the “likes” social media component by individual college.

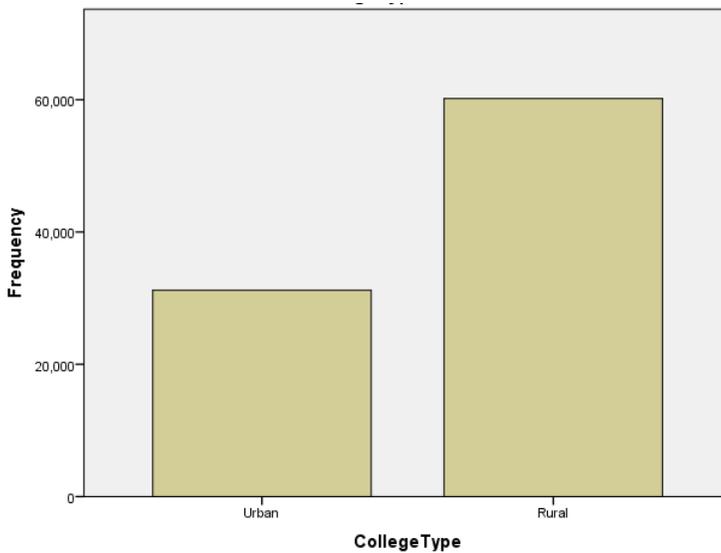


Figure 18. Comparison of Social Media Likes by Aggregate College Types

Figure 18 chart demonstrates that the aggregate of urban colleges obtain a considerable quantity of “likes” compared to the aggregate of rural colleges, which obtained a significantly higher quantity of “likes”.

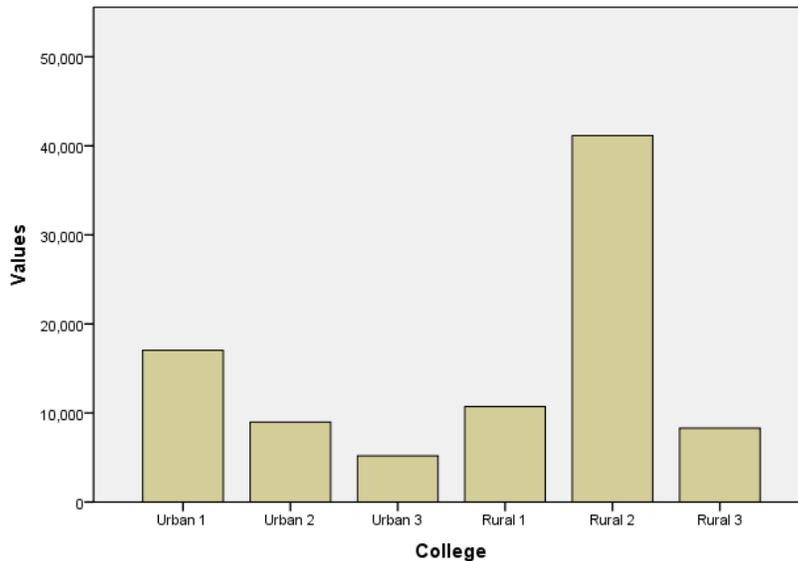


Figure 19. Comparison of Social Media Likes by College

When considering the colleges independently of each other, it can be observed in Figure 18 that only rural college R2 obtained a disproportionate quantity of “likes” compared to all the other colleges. Furthermore, urban college U1 obtained a higher quantity of “likes” than the other urban and rural colleges. Urban colleges U2 and U3 and rural colleges R1 and R3 obtained a proportional level of “likes” among each other.

URL Component

Table 32 presents the descriptive statistics obtained from the colleges’ use of the URL social media component. The analysis revealed that the urban colleges had a slightly higher URL count than the rural colleges. When considering the colleges

separately, the majority of the urban colleges used a higher frequency of URLs than the rural colleges, with U1 obtaining a significant frequency count above all the colleges.

Table 32
Frequency Totals for Component: URLs by College Type

College Type	Frequency	Percent	Aggregate Frequency	Aggregate Percent
Urban 1	3221	37.0		
Urban 2	2862	32.9		
Urban 3	2628	30.2		
Total			8711	52.5
Rural 1	2633	15.9		
Rural 2	2763	16.7		
Rural 3	2476	14.9		
Total			7872	47.5
Total Urban & Rural			16583	100

Note: Aggregate Percent = Aggregate Frequency / Total Urban & Rural Aggregate Frequency times 100

Figure 20 illustrates the aggregate levels using the URL social media component, while Figure 21 illustrates the frequency count of the URL social media component by individual college.

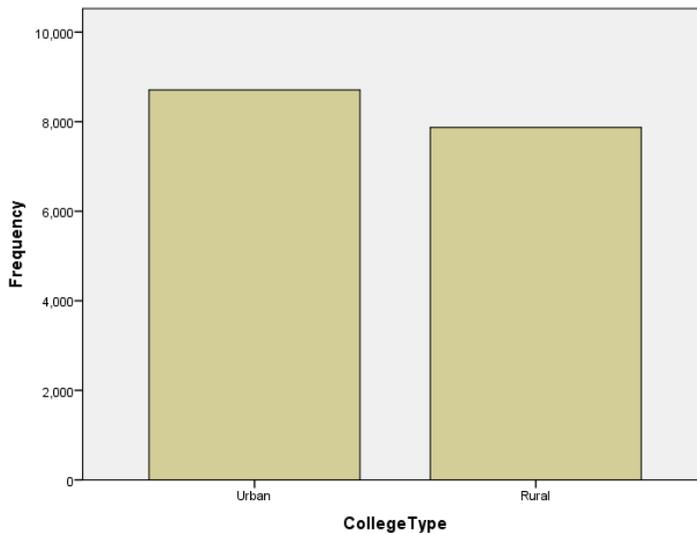


Figure 20. Comparison of Social Media URLs by Aggregate College Types

The above bar chart presents a slightly higher use of URL's in the microblogging activity for the aggregate of urban colleges compared to the aggregate of rural colleges. Therefore, it can be determined that urban colleges are using social media to an advantage in relationship-building capabilities than the rural colleges.

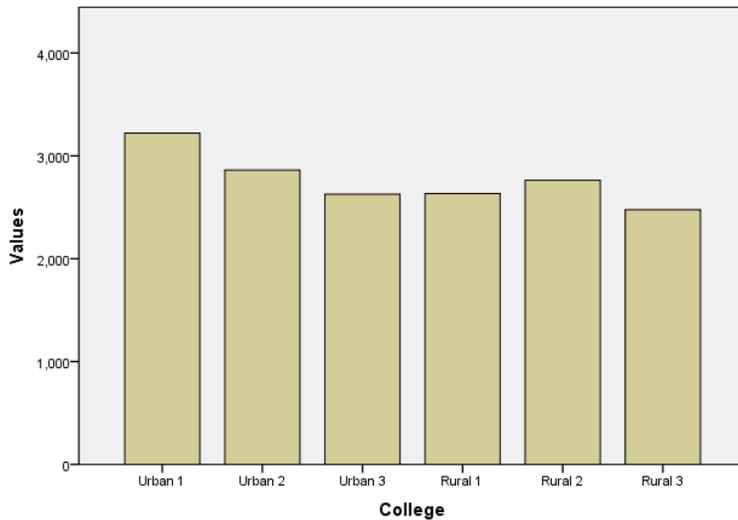


Figure 21. Comparison of Social Media URLs by College

When considering the colleges independently of each other, it can be observed in Figure 21 that only urban college U1 obtained a slightly higher use of URL's in the microblogging activity compared to all the other colleges. The other urban colleges, U2 and U3, and all the rural colleges obtained a proportional level on the use of URL's among each other. Therefore, it can be noted that when the analysis is not in aggregate, both the urban and rural colleges are using the URL social media component to an advantage in relationship-building capabilities.

Hashtag Component

Table 33 presents the descriptive statistics obtained from the colleges' use of the hashtag (#) social media component. The analysis revealed that the urban colleges had a

significantly higher hashtag count than the rural colleges. When considering the colleges separately, all the urban colleges used a higher frequency of hashtags than the rural colleges, with U3 obtaining a significant frequency count above all the colleges.

Table 33
Frequency Totals for Component: Hashtags (#) by College Type

College Type	Frequency	Percent	Aggregate Frequency	Aggregate Percent
Urban 1	1906	27.1		
Urban 2	2523	35.9		
Urban 3	2600	37.0		
Total			7029	57.1
Rural 1	1764	33.5		
Rural 2	1730	32.8		
Rural 3	1777	33.7		
Total			5271	42.9
Total Urban & Rural			12300	100

Note: Aggregate Percent = Aggregate Frequency / Total Urban & Rural Aggregate Frequency times 100

Figure 22 illustrates the aggregate levels using the hashtag social media component, while Figure 23 contains the frequency count of the hashtag social media component by individual college.

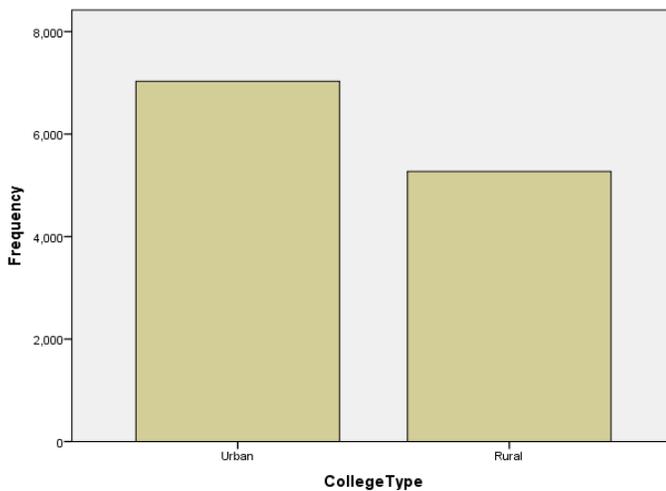


Figure 22. Comparison of Social Media Hashtags (#) by Aggregate College Types

The social media hashtags comparison by aggregate analysis demonstrates a significantly higher use of hashtags in the microblogging activity for the aggregate of urban colleges compared to the aggregate of rural colleges. Consequently, it can be determined that urban colleges are using the hashtag social media component to an advantage in relationship-building capabilities much better than the rural colleges.

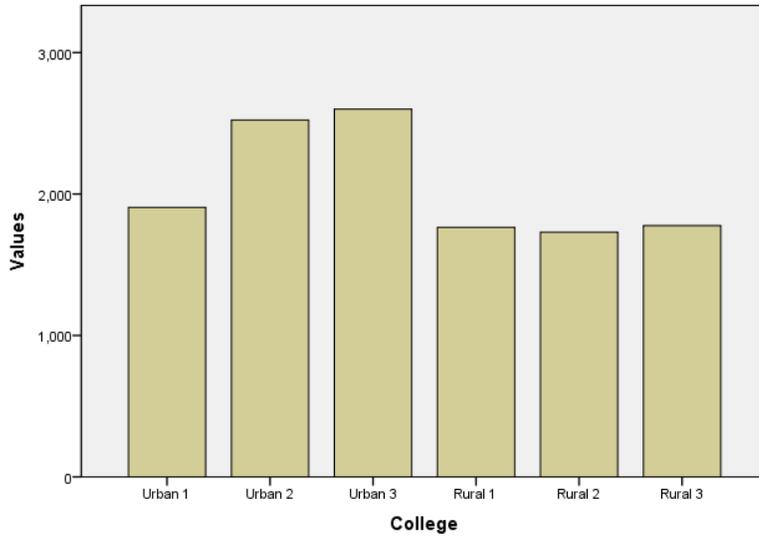


Figure 23. Comparison of Social Media Hashtags (#) by College

It can be observed in Figure 23 that all the urban colleges obtained significantly higher use of hashtags in the microblogging activity than all the rural colleges, when considering the colleges independently of each other. All the rural colleges obtained a proportional level on the use of hashtags among each other. Thus, urban colleges significantly use the hashtag social media component to an advantage in relationship-building capabilities.

At Sign Component

The final social media component considered is the "at sign" (@) component. The descriptive statistics obtained from the colleges use of the "at sign" (@) social media component is presented in Table 34. The analysis revealed that the rural colleges had a significantly higher "at sign" (@) count than the urban colleges. When considering the colleges separately, a majority of the rural colleges used a higher frequency of "at signs" than the urban colleges, with R2 obtaining a slightly higher frequency count above all the colleges.

Table 34
Frequency Totals for Component: At Signs (@) by College Type

College Type	Frequency	Percent	Aggregate Frequency	Aggregate Percent
Urban 1	3333	35.0		
Urban 2	3925	41.3		
Urban 3	2255	23.7		
Total			9513	48
Rural 1	3115	30.2		
Rural 2	4072	39.5		
Rural 3	3128	30.3		
Total			10315	52
Total Urban & Rural			19828	100

Note: Aggregate Percent = Aggregate Frequency / Total Urban & Rural Aggregate Frequency times 100

Figure 24 illustrates the aggregate levels using the "at signs" (@) social media component, while Figure 25 illustrates the frequency count of the "at sign" (@) social media component by individual college.

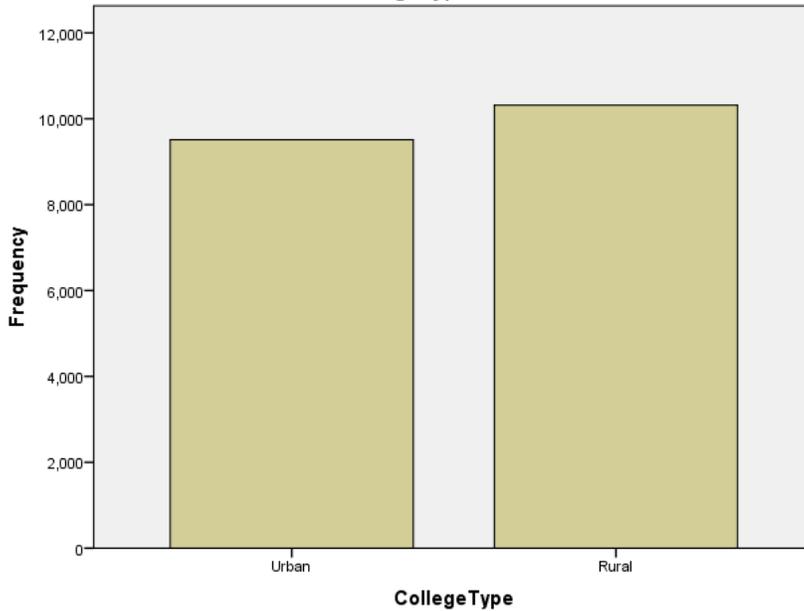


Figure 24. Comparison of Social Media At Signs (@) by Aggregate College Types

The above bar chart presents a slightly higher use of "at signs" (@) in the microblogging activity for the aggregate of rural colleges compared to the aggregate of urban colleges. Therefore, it can be determined that rural colleges are using the "at sign" (@) social media component to an advantage in relationship-building capabilities than the urban colleges.

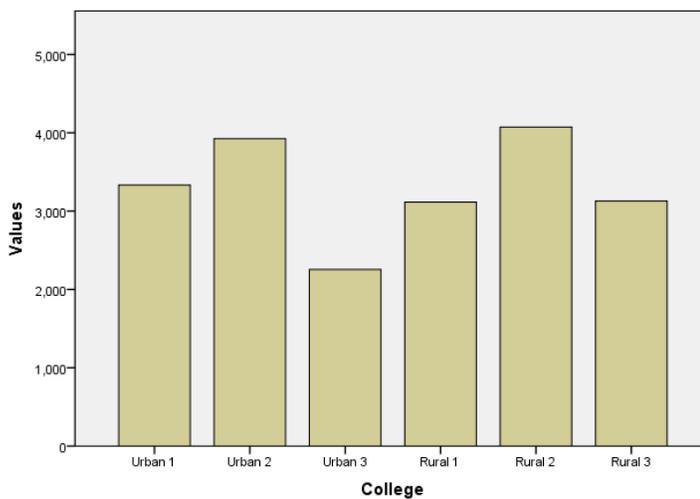


Figure 25. Comparison of Social Media At Signs (@) by College

It can be observed in Figure 25 that when considering the colleges independently of each other, all the urban colleges and rural colleges obtained a proportional level on the use of "at signs" (@) among each other. Thus, both urban colleges and rural colleges use the "at sign" (@) social media component to an advantage in relationship-building capabilities.

Research Question # 5 Analysis

The analysis revealed that both the urban and rural colleges applied the standard use of the social media components on their tweet activity considering the Microblog Dialogic Communication framework. Overall, urban colleges had higher use of social media communication elements than the rural colleges, with urban colleges including more hashtags and URL links than the rural colleges. Rural colleges did rank high on the use of the "at sign" (@) component, which could be attributed to the size of the student population and the need to directly engage with specific individuals of interest within the community or student population, such as parents or decision makers. The elements of social media interaction for the purpose of extending relationships and engaging in Microblog Dialogic Communication were present in the data, although further study could be done on the aforementioned social media components, specifically with larger data samples.

Hypothesis

The main hypothesis claimed by my dissertation are as follows:

Hypothesis 1 - An urban college's positive tweets do not surpass its negative tweets.

The data analysis revealed that the urban college's positive sentiment tweets which averaged 2181 surpassed their negative tweets that averaged 961 as depicted in Table 21, thus hypothesis 1 is rejected.

Hypothesis 2 - A rural college's positive tweets do not surpass its negative tweets.

The data analysis revealed that the rural college's positive sentiment tweets which averaged 2595 surpassed their negative tweets that averaged 592 as depicted in Table 24, thus hypothesis 2 is rejected.

Hypothesis 3 - The average number of negative tweets among urban colleges does not surpasses the average number of negative tweets of rural colleges.

The data analysis revealed that the urban college's negative sentiment tweets which averaged 961 did surpass the rural colleges negative sentiment tweets that averaged 592 as depicted in Table 25, thus hypothesis 3 is rejected.

Hypothesis 4- A college's positive sentiment tweets are not greater than its negative sentiment tweets.

The data analysis revealed that the positive sentiment tweets did surpass the negative sentiment tweets for all colleges. The average urban positive tweets was 2181 while the average urban negative tweets was 961, and the average rural

positive tweets was 2595 while the average rural negative tweets was 592 as depicted in Table 25, thus hypothesis 4 is rejected.

Hypothesis 5 - The total positive sentiment tweets that received a retweet did not exceed the college's negative sentiment tweets that received a retweet.

The data analysis revealed that the total positive sentiment tweets that received a retweet which was 50172 far exceeded the college's negative sentiment tweets that received a retweet that was 28581 as depicted in Table 29, thus hypothesis 5 was rejected.

Hypothesis 6 – The average number of negative tweets of urban colleges do not surpass the average number of negative tweets of rural colleges when analyzed using the Microblog Dialogic Communication framework.

As noted in research question #5, urban colleges had higher use of social media communication elements than the rural colleges, according to the Microblog Dialogic Communication framework. The data analysis revealed that the urban college's negative sentiment tweets which was 21404 did surpass significantly the rural colleges negative sentiment tweets that was 7177 as depicted in Table 29, thus hypothesis 6 is rejected.

Summary

The results of the content analysis study were presented in this chapter. The chapter contained demographic information regarding the colleges used in the study along with the statistical analysis used that included descriptive statistics, Pearson's Chi-Square and quantitative analysis on the college's Twitter activity, the retweets obtained

from the Twitter activity, and the social media components used within the college's Twitter activity.

The overall determination obtained from the data analysis was that rural colleges outperformed urban colleges in regard to their tweet sentiment production and are appropriately applying the Microblog Dialogic Communication framework when creating tweets. Rural colleges are inclined to turn out more positive sentiment tweets while generating lower negative sentiment tweets. The rural college tweet production enhances the overall colleges' image and relationship building capabilities. Furthermore, rural colleges also experienced higher positive sentiment tweets that were retweeted than the urban colleges, indicating that the rural college's messages or tweets obtain a higher impact and reached more people than those tweets created by urban schools. Rural colleges also had considerably lower retweets of negative sentiment tweets than the urban colleges, which experienced much higher retweets of negative sentiment tweets. The process of retweeting negative sentiment tweets can also have an unconstructive impact on a college's image and relationship-building capabilities.

The most favorable outcome for the urban colleges that was uncovered from the data analysis was the use of the social media components on their tweet activity considering the Microblog Dialogic Communication framework. It was observed that urban colleges incorporated higher use of social media communication elements such as URLs, at signs (@), and hashtags (#) than the rural colleges. The process of relationship-building, according to Dialogic Loop Theory, is enhanced by the use of these components, which was clearly being applied by the urban colleges. Despite the use of the social media components in their tweet activity, the urban colleges still obtained

lower positive tweets and lower retweets of the positive tweets, while registering higher negative tweets and higher retweets of those negative tweets. The urban college's outcomes are contrary to what was expected using the Microblog Dialogic Communication framework. A possible justification is the urban school's demographic diversity, which could make the tweet formation difficult since it must be more inclusive with its word choices than the rural colleges that deal with a more homogenous demographic.

The following chapter contains the restatement of the problem, summary of the study, discussion of the results, practical implications, limitations, recommendations for future research and a conclusion.

CHAPTER V: DISCUSSION

The chapter provides a restatement of the problem, summary of the study, and a discussion of the results. In addition, it contains practical implications, limitations, recommendations for future research and a conclusion.

Restatement of the Problem

Kent and Taylor's (1998) established the importance of relationship-building through the use of proper web page development with their Dialogic Loop Theory. Other researchers (McAllister, 2013, Muckensturm, 2013; Watkins, 2017) considered Dialogic Loop Theory as a way to establish relationships by conducting two-way communication in the form of comments, likes, following, and reposting. The gradual evolution of technology and the development of Web 2.0 with social media and increased web interaction have rendered the Dialogic Loop Theory outdated. Specifically, organizations need to establish certain guidelines to appropriately take advantage of the relationship-building capabilities that social media offers, thus it is necessary to augment the Dialogic Loop Theory to compensate for the new social media medium.

The current research focuses on a cross-section of urban and rural community colleges within the United States to identify the sentiment score of their microblogging activity, thereby obtaining a greater understanding on their relationship-building capabilities. As part of the study, I propose a new framework titled Microblog Dialogic Communication, which extends Kent and Taylors' Dialogic Loop Theory to include modern day components and expected activity in today's social media driven environment. The study analyzed the Microblog Dialogic Communication exhibited by a cross-section of urban and rural community colleges within the United States through a

sentiment analysis conducted on the college's Twitter activity. The content analysis performed used a machine learning system to automatically evaluate the tweet sentiment using a classification model that consist of over 1.5 million tweets that have been tagged as positive or negative. Furthermore, the Twitter activity of the urban and rural colleges was analyzed for their use of social media components. The study permits colleges to assess if its Twitter messages are developed in a favorable or positive orientation on behalf of the college, to augment their relationship-building capabilities. The Twitter activity could have a significant impact on registered student participation, the college's retention and enrollment, and further future recruiting efforts.

Summary of the Study

The quantitative content analysis investigated the importance of effective communication, sending out or tweeting positive messages or messages that are perceived as positive, as a significant characteristic that colleges and universities need to adhere to in their microblogging activities to establish relationship-building capabilities. Although Dialogic Loop Theory has been applied to the relationship building capabilities of websites, there is no significant understanding of dialogic loop applied to microblogging. The current research provides a perspective on the Dialogic Loop Theory and its use to effectively generate relationship building capabilities through microblogging. Relationship building is accomplished through the principles developed for the Microblog Dialogic Communication framework.

Microblogging communication is viewed as being reciprocal, with participants providing retweets and comments, thus continuously engaged in the sending and

receiving of messages (Velentzas & Broni, 2014, p. 127; West & Turner, 2010, p. 11). Therefore a dialogue is created through the tweets from the colleges and its public.

Since the main goal of colleges is to maintain beneficial relationships with its public, and microblogging or Twitter tweets are used to directly interact with a college's followers, it is essential that they implement social media strategies that foster those relationship-building opportunities. As part of the social media strategies, it was necessary to augment Kent and Taylor's model for relationship building through the World Wide Web to the microblogging activity on Twitter, establishing specific guidelines and rules that guide the Microblog Dialogic Communication framework that is defined in the present study. It is the intention of the research to provide a training mechanism on the ways a college or an organization can structure its Microblog Dialogic Communication in a way that fosters greater positive sentiment and amplifies the relationship-building capabilities of its social media activities. The research questions that were answered as the result of the present study are:

1. What is the microblogging sentiment of an urban college that interacts with a diverse ethnic student population? (Hypothesis 1)
2. What is the microblogging sentiment of a rural college that interacts with a homogeneous student population? (Hypothesis 2)
3. How does the microblogging activity of urban colleges differ from rural colleges with regard to their microblogging sentiment? (Hypothesis 3)
4. How does the microblogging activities of a college instigate positive outcomes with regard to the sentiment of retweets and comments? (Hypothesis 4 and 5)

5. How well does the Microblog Dialogic Communication framework explain differences that exist on relationship building capabilities in the Twitter activity between urban colleges and rural colleges? (Hypothesis 6)

Discussion of the Results

The following section provides a brief discussion for each research question and summarizes the results of the present study.

Summary and Discussion of Results for Research Question # 1

The urban colleges demonstrated a diverse ethnic composition with a majority of Hispanic demographic at 45.2%. The student population consisted of a greater percentage of female students (56.47%) than male students (43.47%). The microblogging activity for the urban colleges was composed of 69.42% positive sentiment and 30.58% negative sentiment from a total of 9,425 tweets.

The results were consistent with the studies expected outcomes. It was predicted that urban colleges would primarily have a predominant diverse ethnic student population. In regard to the microblogging sentiment, it is essential that tweets are crafted with positive sentiment to enhance the relationship-building capabilities of the social media environment (Agozzino, 2015; Kent & Taylor, 2002; McAllister, 2013; Muckensturm, 2013; Watkins, 2017). Urban colleges propagated positive sentiment tweets at a much higher percentage than negative sentiment tweets, thus satisfying the microblogging sentiment expectations.

Summary and Discussion of Results for Research Question # 2

The rural colleges demonstrated a less inclusive ethnic composition that was significantly more homogeneous than the urban colleges with a majority of White Non-

Hispanic demographic at 55.0%. The student population had a slightly higher percentage of male students (50.80%) than female students (49.20%). The microblogging activity for the rural colleges consisted of 81.42% positive sentiment and 18.58% negative sentiment from a total of 9,560 tweets.

The results mirrored the studies expected outcomes in regard to ethnicity and tweet sentiment for rural colleges. It was forecasted that rural colleges would be limited in regard to an ethnic diverse student population. The descriptive statistics confirmed the predicted outcome. With regard to the microblogging sentiment, rural colleges highly exceeded the expected sentiment outcomes, propagating significantly higher positive sentiment tweets and significantly lowers negative sentiment, firmly meeting the microblogging sentiment expectations.

Summary and Discussion of Results for Research Question # 3

With regard to the microblogging sentiment of the colleges tweet activity, rural colleges averaged 81.40% positive sentiment tweets, while urban colleges averaged 69.50% positive sentiment tweets. Additionally, rural colleges averaged 18.60% negative sentiment tweets and the urban colleges averaged 30.50% negative sentiment tweets. The statistical analysis revealed that the rural colleges over performed in regard to the positive sentiment tweets, while underperformed in the negative sentiment tweets. The urban colleges underperformed in regard to positive sentiment tweets and over performed in the negative sentiment tweets.

The quantitative data analysis demonstrated that on average, rural colleges propagated higher positive sentiment tweets and lowers negative sentiment tweets than

urban colleges. Therefore, rural colleges exceeded the microblogging sentiment expectations when compared to urban colleges.

Summary and Discussion of Results for Research Question # 4

The discussion for research question number four centered on the dialogic practices of colleges, with positive interactions seen as the essential component for relationship-building. With regard to microblogging and Twitter, the dialogic loop is ascertained by the retweet count that occurs on the positive sentiment tweets. Urban colleges obtained 19,962 retweets of their positive sentiment tweets, resulting in 2.98 retweets for each positive tweet, while rural colleges experienced 30,210 retweets of their positive sentiment tweets, thus each positive tweet was retweeted 3.99 times. The data analysis demonstrated that rural colleges, on average, had significantly higher positive sentiment retweets and considerably lower negative sentiment retweets than the urban colleges. Therefore, rural colleges significantly outperformed urban colleges and demonstrated greater capability at establishing better communication feedback than urban colleges.

Summary and Discussion of Results for Research Question # 5

One of the primary outcomes of the current research is the microblogging analysis of the college's tweets to identify if they are incorporating specific social media components that enhance relationship-building capabilities. Specifically, the present study aimed to augment Kent and Taylor's Dialogic Loop Theory to include these social media components and formulating a new Microblog Dialogic Communication framework. Therefore, an analysis was conducted on specific social media components that included URL links, at signs (@), and hashtags (#) used within the tweets.

The urban college tweets contained 8,711 URL links, 9,513 "at signs" (@), and 7,029 hashtags (#), while the rural college tweets contained 7,872 URL links, 10,315 "at signs" (@), and 5,271 hashtags (#). The urban colleges had a total of 38,428 followers with 1,012 friends and the rural colleges had a total of 89,010 followers with 7,184 friends. Considering the Microblog Dialogic Communication framework, urban colleges had higher use of social media communication elements than the rural colleges, although rural colleges had significantly higher followers and friends. Overall, urban colleges outperformed rural colleges in applying proper use of the social media components to increase relationship-building.

Practical Implications

The current study has practical implications for colleges and organizations that use microblogging or Twitter to interact with their public. When considering the structure of microblog messages, the positive image or communication propagated by a college when tweeting can increase student engagement, augment student retention rates and possibly increase student enrollment through the dialogic loop created by the college's tweets. Of particular importance is the use of the Microblog Dialogic Communication framework to extend the relationship-building capabilities of the college's tweet activity.

Considering the quantitative results and the statistics observed, there exist a statistical significance between a college's tweet activity and the tweet sentiment. Thus, it is important to note that tweets must contain the appropriate social media components to create a sense of interaction, relationship-building, and continued support or following. A significant contribution of the present work is the modification of the original Kent and Taylor's (1998) Dialogic Loop Theory to include these ever so important social media

components. The enhancement of the theory, by incorporating social media components into the microblogging activity of an organization, is labeled in the current body of work as Microblog Dialogic Communication framework. Below is Figure 26 with a visual representation of the Microblog Dialogic Communication framework:

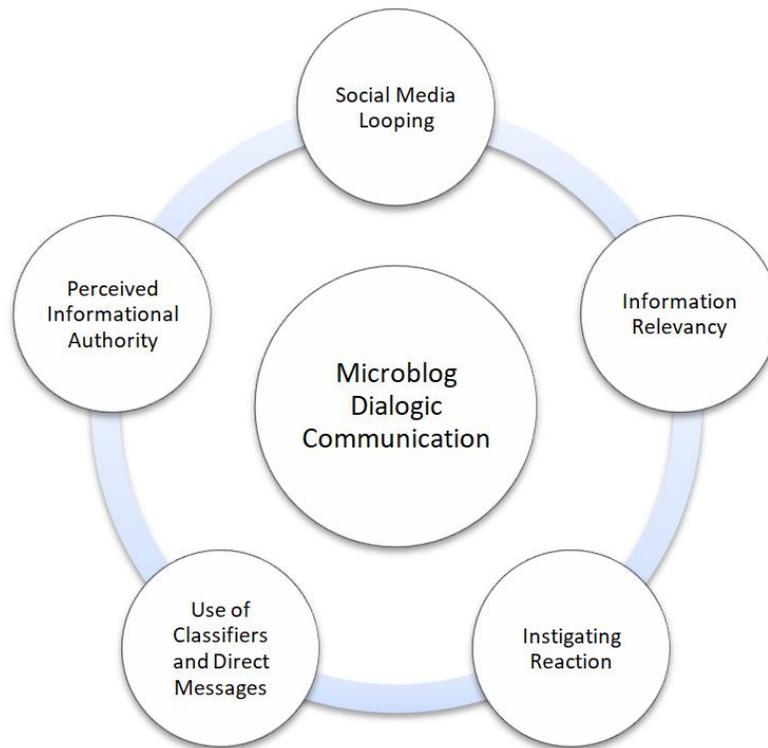


Figure 26. Microblog Dialogic Communication Framework

The Microblog Dialogic Communication framework requires that microblogging messages include 1) a way for the receiver of the message to provide feedback in the form of a comment or obtain "likes" or approval of the message, 2) will provide new, relevant, constructive, trustworthy, and engaging content, 3) that messages are created with the intent that it will be re-posted, shared, or re-tweeted in order to extend the messages reach beyond its planned target, 4) that messages include social media elements such as hashtags (#) and at signs (@) to foster social media interactions, and 5) that

messages include images and/or video links that extend the message's information and provide additional resources or facts. The following is an example of a microblogging tweet that is successfully using the aforementioned Microblog Dialogic Communication framework:

New @NatureGeosci study led by @jsmankin @dartmouth finds increased CO2 & warmer temps means plants consume more water, leaving less water for people in N.Amer, Eur & Asia via @natgeo by @StephenLeahy. #climatecrisis #climatechange #waterstress #drought



Thirsty future ahead as climate change explodes plant growth
Rising CO2 levels and a warmer earth means plants will grow bigger and have longer to suck the land dry. That's bad news for human water suppl...
nationalgeographic.com

Figure 277. Microblog Dialogic Communication Tweet Example

Finally, all colleges and organizations should be provided with guidance regarding their microblogging and social media activity, along with suggestions on how to formulate positive oriented messages with the necessary social media components to assist in increasing relationship-building functionality. These suggestions, as identified in the Microblog Dialogic Communication framework, serve as a guideline to help structure proper messages when considering the social media atmosphere that dominates the interactions among senders of the message such as a college or an organization, and the receivers of the message such as students, parents, or stakeholders. The practical

implication is that prospective colleges and organizations should be provided with guidance regarding online interactions and properly posting messages. The next section contains the limitations of the study.

Limitations

The purpose of the present content analysis study was to analyze the sentiment score of the microblogging activity of a cross-section of urban and rural community colleges within the United States. The study compared the sentiment score among these urban and rural college groups independently of each other, comparing to evaluate if the size of the college had an impact on their Twitter activity. Furthermore, a new model is defined, stating the necessary requirements for creating microblogging activity that increases the opportunity for relationship-building. However, there are a few factors that may have limited the study.

1. The study was limited to only colleges. Further analysis can be accomplished by expanding the test sample to include other industries such as non-for-profit organizations or private businesses that frequently use microblogging to interact with their public to determine if proper Microblog Dialogic Communication is being applied.
2. The study was limited in the scope of colleges being considered. Since only three urban and three rural colleges were selected for the research, further analysis could include a larger test sample of colleges from both spectrums to expand on the differences associated with the urban/rural categorization and microblogging activities.

3. Future studies could include analyzing tweets sent by colleges whose student bodies are more culturally diverse.
4. The study included a juxtaposition between the diverse demographics of urban colleges and the homogenous demographic of rural colleges. An analysis that expands the college test data using demographic type such as diverse or homogenous and the sentiment analysis could further yield a greater understanding on a college's microblogging activity given its targeted demographic audience.
5. A limitation of the present study is that it provided quantitative research. A future study could benefit from a mixed method study that includes qualitative research.

Even though the study contained some limitations, the present content analysis study contributed to the knowledge base of a college's microblogging activity and the field of sentiment analysis. Future research should be conducted on the topic.

Recommendations for Future Research

Future research should be conducted on other areas beyond the academic realm such as non-for-profit organizations or private businesses. Since microblogging or Twitter is a main component of the social media movement, many individuals and organizations engage in constant microblogging activity. Therefore, future studies could focus on other types of entities that frequently use microblogging to interact with their public to determine the sentiment being expressed and if the proper Microblog Dialogic Communication is being applied.

Additional research could extend the quantity and sizes of the colleges being considered. The current research was focused on three urban and three rural colleges,

further study could be done on a larger data set of only urban colleges to specifically note the sentiment being expressed in colleges with diverse demographics. Conversely, similar studies should be conducted on a greater quantity of rural college's, taking into consideration the aforementioned demographic.

Furthermore, since colleges are expected to play an increase role in civic engagement and social justice issues, future research can be conducted to determine if the tweets sent out by colleges focus on these topics. A main area of interest could be the civic engagement being expressed through the college's microblogging and how they are informing or motivating their students along the lines of civic responsibilities, inclusive language, and social justice matters.

Another future study could specifically be done on a retweet sentiment analysis of a college's or an organization's microblogging activity. The concept could consider a test group incorporating the Microblog Dialogic Communication framework for an extended period of time, to determine the quantity of followers and likes obtained through the test period. The intended focus for the study would be founded under the Microblog Dialogic Communication framework in considering the use of social media components as defined by the framework. The study could yield a greater understanding on the relationship-building capabilities that the Microblog Dialogic Communication framework offers when consistently applied.

Furthermore, the Microblog Dialogic Communication framework could be extended to other types of social media platforms such as Facebook, YouTube, Instagram, or Pinterest to name a few. Further study could be performed on the use of the

framework to generate additional followers and extend relationships with users within these mediums.

Other research can be conducted on analyzing the expansion of Twitters character limitation from 140 characters to 280 characters to note the implications, if any, on the sentiment analysis process.

Additionally, a future study could include a mixed method approach where not only quantitative data is collected but also qualitative data to obtain a better understanding directly from the stakeholders involved.

Finally, the current study focused on the microblogging sentiment expressed by urban and rural colleges, a future study could focus on extending the content analysis within similar groups such as comparing urban colleges to urban colleges, or expanding the research among urban and rural colleges, such as obtaining a larger corpus, or expanding the scope of the groups to obtain a better sample data for the sentiment analysis in general.

Conclusion

The present content analysis study attempted to conduct a sentiment analysis on the microblogging activity of urban and rural colleges to identify if colleges are using appropriate language that fosters relationship-building outcomes. Furthermore, other microblogging aspects were analyzed to determine if colleges are conforming to the use of several social media components beyond Kent and Taylor's (1998) Dialogic Loop Theory. The current research extends the theory to a new framework called Microblog Dialogic Communication, updating the theory to encompass the necessary social media components.

The current study found that rural colleges tweeted more positive sentiment tweets and less negative sentiment tweets when compared to the urban colleges tweet. Additionally, the rural positive sentiment tweets received more retweets than the urban colleges. Rural colleges also had considerably lower negative sentiment retweets that were retweeted in comparison to the urban college negative sentiment tweets that received significantly higher retweets. The two factors mentioned, positive sentiment tweets and the quantity of retweets obtained from the positive sentiment tweets and the negative sentiment tweets, are key factors influencing the colleges' overall image and relationship-building capabilities. The results suggest that rural college's Twitter activity obtain a higher impact and reach more of its public than the Twitter activity by urban colleges.

Additionally, the study considered the use of social media communication elements such as URLs, at signs (@), and hashtags (#) as a form of incorporating the Microblog Dialogic Communication framework to augment the relationship-building capabilities of Twitter. The research concluded that urban colleges used higher elements of social media communication than the rural colleges. Regardless of the social media components used by urban colleges, they still obtained lower positive tweets and lower retweets of the positive tweets, while registering higher negative tweets and higher retweets of those negative tweets. These results opposed the expected outcome for when a college utilizes the Microblog Dialogic Communication framework. It was also determined that the rural colleges incorporated the Microblog Dialogic Communication framework with tweets that contained many of the social media communication elements. The rural colleges obtained favorable results in regard to their positive and negative

sentiment tweet activity, and the volume of retweet activity on those tweet sentiments, respectively.

In conclusion, colleges should be provided specific guidance and structure in integrating the particular elements mentioned as part of the Microblog Dialogic Communication framework to enhance the relationship-building capabilities of their microblogging activities. Also, colleges should be mindful on the word choices used in formulating their microblogging communication to foster more positive sentiment and inclusive communication.

REFERENCES

- Abdelrazeq, A., Janssen, D., Tummel, C., Jeschke, S., & Richert, A., (2015). *Sentiment analysis of social media for evaluating universities*. Proceedings of Second International Conference on Digital Information Processing, Data Mining, and Wireless Communications, Dubai, UAE. Retrieved from http://www.exact.rwth-aachen.de/fileadmin/user_upload/Sentiment_Anas.pdf
- Adams, A., & McCorindale, T. (2013). Dialogue and transparency: A content analysis of how the 2012 presidential candidates used twitter. *Public Relations Review*, 39, 357–359.
- Agozzino, A. (2015). Dialogic communication through “pinning”: An analysis of top 10 most - followed organizations’ pinterest profiles. *Public Relations Journal*, 9(3). Retrieved from <http://prjournal.instituteforpr.org/wp-content/uploads/2015v09n03Agozzino.pdf>
- Alvidrez, Salvador & Franco-Rodríguez, Oziel. (2016). Powerful communication style on twitter: Effects on credibility and civic participation. *Comunicar*, 24(47). Retrieved from https://www.researchgate.net/publication/294276716_Powerful_Communication_Style_on_Twitter_Effects_on_Credibility_and_Civic_Participation
- Austin, L., & Jin, Y. (2015). Approaching ethical crisis communication with accuracy and sensitivity: Exploring common ground and gaps between journalism and public relations. *Public Relations Journal*, 9(1). Retrieved from <http://www.prsa.org/Intelligence/PRJournal/Vol9/No1/>
- Balahur, A. (2013). *Sentiment analysis in social media texts*. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, (pp. 120–128). Atlanta, Georgia. Retrieved from <http://www.aclweb.org/anthology/W/W13/W13-1617.pdf>
- Bayesian. (2018). Short history of Bayes’ theorem. <http://bayesian.org/>. Retrieved from <http://www.oulu.fi/biology/PopGenSchool/part1.pdf>
- Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. *Studies in Computational Intelligence*, 639, 313-340). Retrieved from <http://www.public.asu.edu/~gbeigi/files/BeigiSentimentChapter.pdf>
- Berelson, B. (1952), *Content analysis in communication research*, New York: The Free Press.

- Berger, J. & Milkman, K. L., (2011). What makes online content viral? *Journal of Marketing Research*, Retrieved from <http://jonahberger.com/wp-content/uploads/2013/02/ViralityB.pdf>
- Berkson, J. (1930). Bayes' Theorem. *The Annals of Mathematical Statistics*, Vol. 1, No. 1, pp. 42-56. Retrieved from <https://www.jstor.org/stable/2957673>
- Berners-Lee, T., Fielding, R., & Masinter, L. (1998). Uniform Resource Identifiers (URI): Generic Syntax. Internet Engineering Task Force. Retrieved from <https://tools.ietf.org/html/rfc2396>.
- BestColleges. (2018). The Stars of College Social Media. Retrieved from <https://www.bestcolleges.com/features/best-college-social-media/>
- Bordum, A. (2002). The Categorical Imperative: Analyzing Immanuel Kant's Grounding for A Metaphysics of Morals. Department of Management, Politics and Philosophy. Retrieved from <https://pdfs.semanticscholar.org/7dec/d3d33d7af24ae9b1902b5034f8181998d91e.pdf>
- Bortree, D. S., & Seltzer, T. (2009). Dialogic strategies and outcomes: an analysis of environmental advocacy groups' facebook profiles. *Public Relations Review*, 35, 317-319. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0363811109000691>
- Boyd, D. M. & Ellison, N.B., (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/full>
- Chaffee, S. H., & Metzger, M. J. (2001). The end of mass communication? *Mass Communication & Society*, 4, 365-379. Retrieved from <http://courses.washington.edu/com201/COM%20201%20readings/Chaffee%20and%20Metzger-The%20end%20of%20mass%20communication.pdf>
- College Data. (2018). Retrieved from <https://www.collegedata.com/>
- College Factual. (2018). Retrieved from <https://www.collegefactual.com>
- Curran, J. (2002). *Media and power*. London: Routledge.

- Danso, S., Atwell, E.S. & Johnson, O. (2013). A comparative study of machine learning methods for verbal autopsy text classification. *International Journal of Computer Science Issues*, 10(6). Retrieved from https://www.researchgate.net/publication/260250671_A_Comparative_Study_of_Machine_Learning_Methods_for_Verbal_Autopsy_Text_Classification
- Dave, K., Lawrence, S., & Pennock, D. M. (2003). *Mining the peanut gallery: opinion extraction and semantic classification of product reviews*. In Proceedings of the 12th international conference on World Wide Web, pages 519–528. Retrieved from <http://www.kushaldave.com/p451-dave.pdf>
- Davidov, D.; Tsur, O.; and Rappoport, A. (2010). *Enhanced sentiment learning using Twitter hashtags and smileys*. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, 241–249. Retrieved from <http://www.aclweb.org/anthology/C10-2028>
- Developer agreement and policy (2017). Retrieved from <https://developer.twitter.com/en/developer-terms/agreement-and-policy>
- DeVoe, K.M. (2009). Bursts of information: Microblogging. *The Reference Librarian*, 50(2), 212-214. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/02763870902762086?journalCode=wref20>
- Dodd, J. (2015). *Twitter sentiment analysis*. Diploma thesis. National College of Ireland. Retrieved from <http://trap.ncirl.ie/view/creators/Dodd=3AJohn=3A=3A.html>
- Dugan, L., (2011). Twitter 101: Why use hashtags? *AdWeek*. Retrieved from <http://www.adweek.com/digital/twitter-101-why-use-hashtags/>
- Eckler, P. and Rodgers, S. 2010. Viral marketing on the internet. *Wiley International Encyclopedia of Marketing*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781444316568.wiem04009/pdf>
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107-115.
- Gartner, M. (2012). How did Twitter decide on 140 characters? Was there a formula or extensive research behind that decision? Quora. Retrieved from <https://www.quora.com/Why-does-Twitter-limit-the-message-length-to-140-characters>
- Gauntlett, D. (2002). *Media, gender & identity*. London: Routledge.

- Go, A., Bhayani, R. & Huan, L. (2009). Twitter sentiment classification using distant supervision. *Processing*, 150, Retrieved from <https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>
- Gottfried, J. & Shearer, E., (2016). News use across social media platforms 2016. *Pew Research Center*. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/13/2016/05/PJ_2016.05.26_social-media-and-news_FINAL-1.pdf
- Greenwood, S., Perrin, A. & Duggan, M., (2016). Social media update 2016. *Pew Research Center*. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/14/2016/11/10132827/PI_2016.11.11_Social-Media-Update_FINAL.pdf
- Grosse, K., Chesnevar, C.I., & Maguitman, A.G. (2012). *An argument-based approach to mining opinions from Twitter*. In AT, 408–422. Retrieved from <http://ceur-ws.org/Vol-918/111110408.pdf>
- Grove, T. (2009). History bytes: To tweet or not to tweet. American Association for State and Local History. *History News*, 64(4), 5-6. Retrieved from <http://www.jstor.org/stable/42654232>
- Grove, T. (2009). History bytes: To tweet or not to tweet. *History News*, 64(4), 5-6. American Association for State and Local History. Retrieved from <http://www.jstor.org/stable/42654232>
- Hand, D. J. and Yu, K. (2001). Idiot's Bayes-Not So Stupid After All? *International Statistical Review*, p. 385-398. Retrieved from <https://pdfs.semanticscholar.org/bd6a/9d35dabbba8132f48835f636cf7c3b3e9c80.pdf>
- Hennessy, A. (2014). *Sentiment analysis of twitter: Using knowledge based and machine learning techniques*. National College of Ireland. Higher Diploma in Science in Data Analytics. Retrieved from http://www.academia.edu/28754910/Sentiment_Analysis_of_Twitter_Using_Knowledge_based_and_Machine_Learning_Techniques_2014
- Holsti, O. R. (1960). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/000169937001300209>

- Horne, C.F. (1917). *Sacred books and early literature of the east: Egypt*. New York, New York, and London, England: Parke, Austin and Lipscomb, Inc. Retrieved from http://billheidrick.com/Orpd/Sacr1917/Sacred_Books_2.pdf
- Jacobson, D. (2003). J.S. Mill and the Diversity of Utilitarianism. *Philosophers' Imprint*. Vol. 3. No. 2. Retrieved from <https://quod.lib.umich.edu/cgi/p/pod/dod-idx/j-s-mill-and-the-diversity-of-utilitarianism.pdf?c=phimp;idno=3521354.0003.002;format=pdf>
- Jhaveri, D., Chaudhari, A. & Kurup, L. (2015). Twitter sentiment analysis on e-commerce websites in India. *International Journal of Computer Applications*, 127(18). Retrieved from <http://www.ijcaonline.org/research/volume127/number18/jhaveri-2015-ijca-906730.pdf>
- Jhaveri, D., Chaudhari, A., & Kurup, L. (2015). Twitter sentiment analysis on e-commerce websites in India. *International Journal of Computer Applications*, 127(18). Retrieved from https://www.researchgate.net/publication/283244805_Twitter_Sentiment_Analysis_on_E-commerce_Websites_in_India
- Joachims, T. (1998). *Text categorization with support vector machines: Learning with many relevant features*. In Proc. of the European Conference on Machine Learning (ECML), pages 137–142. Retrieved from https://www.cs.cornell.edu/people/tj/publications/joachims_98a.pdf
- Kent, M. L., & Taylor, M. (1998). Building a dialogic relationship through the World Wide Web. *Public Relations Review*, 24(3), 321–340. Retrieved from <https://www.sciencedirect.com/science/article/pii/S036381119980143X>
- Kent, M. L., & Taylor, M. (2002). Toward a dialogic theory of public relations. *Public Relations Review*, 28(1), 21–37. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.577.1950&rep=rep1&type=pdf>
- Kent, M. L., & Taylor, M. (2003). Maximizing media relations: A Web site checklist. *Public Relations Quarterly*, 48(1), 14–18. Retrieved from http://www.academia.edu/2845555/Maximizing_media_relations_A_Web_site_checklist
- Kent, M. L., Taylor, M., & White, W. (2003). The relationship between Web site design and organizational responsiveness to stakeholders. *Public Relations Review*,

- 29(1), 66–77. Retrieved from
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.322.7141&rep=rep1&type=pdf>
- Kerstetter, K. A. (2008). Mapping and measuring Twitter engagement: Fortune 500 companies make good on a dialogic promise. *The University of Alabama*. Retrieved from https://getd.libs.uga.edu/pdfs/kerstetter_kerie_a_201405_ma.pdf
- Kiritchenko, S. Zhu, Xiaodan., & Mohammad, S. M. (2014). *Journal of Artificial Intelligence Research*. National Research Council Canada, pp. 723-762. Retrieved from <https://pdfs.semanticscholar.org/e2ae/833ec0bf4a0435d7feea94a48c027e06a0e1.pdf>
- Kouloumpis, E.; Wilson, T.; and Moore, J. (2011). *Twitter sentiment analysis: The good the bad and the omg!* In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 538–541. Retrieved from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2857/3251>.
- Kumar, S. H. (2014). *Twitter sentiment analysis*. University of California, Santa Cruz Computer Science. Retrieved from <https://shachihkumar.files.wordpress.com/2014/02/twitter-sentiment-analysis-project-report.pdf>
- Kwak, Haewoon & Lee, Changhyun & Park, Hosung & Moon, Sue. (2010). *What is twitter, a social network or a news media?* Proceedings of the 19th International Conference on World Wide Web, Retrieved from <https://an.kaist.ac.kr/~haewoon/papers/2010-www-twitter.pdf>
- Lasswell, H. D., Lerner, D., & Pool, I. de Sola. (1952). *The comparative study of symbol: An introduction*. California: Stanford University Press.
- Laws, S. (2014). *Corporate communication in the twenty-first century*. Other Graduate Scholarship.Paper 1. Retrieved from http://digitalcommons.liberty.edu/busi_grad_schol/1
- Ledingham, J.A. (2003). Explicating relationship management as a general theory of public relations. *Journal of Public Relations*, 15, 181-198. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S1532754XJPRR1502_4
- Lee, A.M. & Gil de Zúñiga, Homero & Coleman, Renita & Johnson, Thomas. (2014). The dialogic potential of social media: assessing the ethical reasoning of companies' public relations on facebook and twitter. *Ethical Practice of Social*

- Media in Public Relations*, 157-175. Retrieved from https://www.researchgate.net/publication/280304799_The_Dialogic_Potential_of_Social_Media_Assessing_the_Ethical_Reasoning_of_Companies%27_Public_Relations_on_Facebook_and_Twitter
- Lee, K., Oh, W., & Kim, K. (2013). Social media for socially responsible firms: Analysis of fortune 500's Twitter profiles and their CSR/CSIR ratings. *Journal of Business Ethics*, 118(4), 791-806. Retrieved from <http://www.jstor.org/stable/42921266>
- Lesswrong. (2011). A History of Bayes' Theorem. Lesswrong.com. Retrieved from <https://www.lesswrong.com/posts/RTt59BtFLqQbsSiqd/a-history-of-bayes-theorem>
- Lim, Y. & Lee-Won, R. J. (2017). When retweets persuade: The persuasive effects of dialogic retweeting and the role of social presence in organizations' Twitter-based communication. *Telematics and Informatics*, 34, 422–433. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0736585316301186>
- Lovejoy, K., Waters, R. D. & Saxton, G. D. (2012). Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. *Public Relations Review*, 38, 313– 318. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1106/1106.1852.pdf>
- Mackey, T. P.& Jacobson, T. E., (2011). Reframing information literacy as a metaliteracy. *College & Research Libraries*,72(1), 62-78. Retrieved from <https://pdfs.semanticscholar.org/6d77/5e0711644fad14b8abb22314fb19b9c79bca.pdf>
- Macnamara, J. R. (2003). Media content analysis: Its uses, benefits & best practice methodology. *Asia Pacific Public Relations Journal*, 6(1), 1–34. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.477.4814&rep=rep1&type=pdf>
- Madden, M. (2012). Privacy management on social media sites. *Pew Research Center*. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Privacy_management_on_social_media_sites_022412.pdf
- McAllister, S. M. (2013). Toward a dialogic theory of fundraising, Community College. *Journal of Research and Practice*, 37(4), 262-277. Retrieved from http://bluehawk.monmouth.edu/~smcallis/McAllister_13_CCFR.pdf

- McAllister-Spooner, S. M. (2009). Fulfilling the dialogic promise: A ten-year reflective survey on dialogic Internet principles. *Public Relations Review*, 35, 320–322. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0363811109000587>
- McAllister-Spooner, S. M., (2008). User perceptions of dialogic public relations tactics via the internet. *Public Relations Journal*, 2(1). Retrieved from <http://prjournal.instituteforpr.org/wp-content/uploads/User-Perceptions.pdf>
- McCornack, S. & Ortiz, J. (2016). *Choices & connections: An introduction to communication*. 2nd Edition. Bedford: St. Martin's
- McCoy, C.G., Nelson, M. L. and Weigle, M. C. (2017). University Twitter Engagement: Using Twitter Followers to Rank Universities. Retrieved from <https://arxiv.org/pdf/1708.05790.pdf>
- McNally, J. R., (1970). Toward a definition of rhetoric. *Philosophy & Rhetoric*, 3(2), 71-81. Retrieved from <http://www.jstor.org/stable/40236708>
- Miller, D., Costa, E., Haynes, N., McDonald, T., Nicolescu, R., Sinanan, J., Spyer, J., Venkatraman, S. & Wang, X. (2016). *How the world changed social media: What is social media?* UCL Press: London, UK. Retrieved from <http://www.jstor.org/stable/j.ctt1g69z35.8>
- Mills, K., & Chandra, V. (2011). Microblogging as a literacy practice for educational communities. *Journal of Adolescent & Adult Literacy*, 55(1), 35–45. Retrieved from http://www.jstor.org/stable/41309645?seq=1#page_scan_tab_contents
- Mitsis, P.(1988). *Epicurus' Ethical Theory: The Pleasures of Invulnerability*. Cornell University Press. Retrieved from <http://www.jstor.org/stable/10.7591/j.cttq45fk>
- Morgan, G. Leech, N., Gloeckner, G. Barrett, K. (2007). *SPSS for introductory statistic, use and interpretation (3rd ed.)* Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Moss, C. (2013). Our updated guide to Twitter slang, lingo, abbreviations and acronyms. *Business Insider*. Retrieved from <http://www.businessinsider.com/a-guide-to-twitter-slang-lingo-abbreviations-and-acronyms-2013-9>
- Muckensturm, E. (2013). Using dialogic principles on facebook: How the accommodation sector is communicating with its' consumers. *Clemson University Tiger Prints, All Theses Theses 5*. Retrieved from

https://tigerprints.clemson.edu/cgi/viewcontent.cgi?article=2657&context=all_theses

- Murthy, D. (2012). Towards a sociological understanding of social media: Theorizing twitter. *Sociology*, 46(6), 1059-1073. Retrieved from <http://www.jstor.org/stable/43497339>
- Nasukawa, T. & Yi, J. (2003). *Sentiment analysis: Capturing favorability using natural language processing*. K-CAP '03 Proceedings of the 2nd international conference on Knowledge capture. 70-77. Retrieved from <https://dl.acm.org/citation.cfm?id=945658>
- Neuman, W.L. (1997). *Social research methods: qualitative and quantitative approaches*. Needham Heights, MA: Allyn & Bacon.
- Newbold, C., Boyd-Barrett, O. & Van Den Bulck, H. (2002). *The media book*. London: Arnold.
- Nigam, K., Lafferty, J. & McCallum, A. (1999). *Using maximum entropy for text classification*. In Proc. of the IJCAI-99 Workshop on Machine Learning for Information Filtering, 61-67, Retrieved from <https://www.cc.gatech.edu/~isbell/reading/papers/maxenttext.pdf>
- Ntaflos, L. (2013). *Spatiotemporal mining in microblogs*. Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, 1-33. Retrieved from <https://www.cse.ust.hk/~entaflos/PQE-Ntaflos.pdf>
- Number of social media users worldwide from 2010 to 2021 (in billions) (2017). Retrieved from <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
- Number of social network users in selected countries in 2017 and 2022 (in millions) (2017). Retrieved from <https://www.statista.com/statistics/278341/number-of-social-network-users-in-selected-countries/>
- O'Neil, O. (2009). Ethics for Communication?. *European Journal of Philosophy* 17:2. pp. 167–180. Blackwell Publishing Ltd. Retrieved from <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0378.2009.00346.x>
- Pak, A. & Paroubek, P.(2010). *Twitter as a corpus for sentiment analysis and opinion mining*. In Proceedings of the Seventh Conference on International Language Resources and Evaluation,1320–1326, Valletta, Malta. Retrieved from

<https://pdfs.semanticscholar.org/ad8a/7f620a57478ff70045f97abc7aec9687ccbd.pdf>

- Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2),1–135. Retrieved from <http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>
- Pang, B., Lee, L. & Vaithyanathan, S. (2002). *Thumbs up? Sentiment classification using machine learning techniques*. Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79--86. Retrieved from <http://www.cs.cornell.edu/home/llee/papers/sentiment.pdf>
- Park, H., & Reber, B. H. (2008). Relationship building and the use of web sites: How Fortune 500 corporations use their web sites to build relationships. *Public Relations Review*, 34(4), 409-411. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0363811108000921>
- Paulus, M. J. (2007). Blogging for the record: A study of blogging from an archival point of view. *Journal of Archiving Organization*, 4(3-4), 31-41. Retrieved from http://dx.doi.org/10.1300/J201v04n03_03
- Pearson, R. (1989). *A theory of public relations ethics*. (Unpublished doctoral dissertation). Ohio University, Athens, OH.
- Rate limiting (2017). Retrieved from <https://developer.twitter.com/en/docs/basics/rate-limiting>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. *Encyclopedia of Database Systems*. (532-538). Retrieved from https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_565
- Rose, S., Spinks, N., & Canhoto, A.S., (2015). *Management research: Applying the principles*. New York: Routledge.
- Rosen, A. (2017). Tweeting Made Easier. Retrieved from https://blog.twitter.com/official/en_us/topics/product/2017/tweetingmadeeasier.html
- Ruhi, U. (2014). *Social media analytics as a business intelligence practice: Current landscape & future prospects*. *Journal of Internet Social Networking & Virtual Communities*, Vol. 2014 (2014), Article ID 920553. Retrieved from <http://ibimapublishing.com/articles/JISNVC/2014/920553/920553.pdf>

- Rybalko, S., & Seltzer, T. (2010). Dialogic communication in 140 characters or less: How Fortune 500 companies engage stakeholders using Twitter. *Public Relations Review*, 36, 336-341. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0363811110000792>
- Sentiment140. (2018). Retrieved from <http://www.Sentiment140.com/>
- Smits, M., & Mogos, S. (2013). *The impact of social media on business performance*. Proceedings of the 21st European Conference on Information Systems. Retrieved from https://www.researchgate.net/publication/273138667_The_impact_of_social_media_on_business_performance
- Stigler, S.M. (1983). Who Discovered Bayes's Theorem? *The American Statistician*, Vol. 37, No. 4, Part 1 (Nov., 1983), pp. 290-296. Retrieved from <https://www.jstor.org/stable/2682766>
- Stigler, S. M. (1986), "Laplace's 1774 Memoir on Inverse Probability", *Statistical Science* 1(3):359–363. Retrieved from <https://www.jstor.org/stable/2245475>
- Suchdev, R., Kotkar, P., Ravindran, R., & Swamy, S. (2014). Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach, *International Journal of Computer Applications* (0975– 8887), Vol. 103, No.4. Retrieved from <https://pdfs.semanticscholar.org/d5b7/264fc852e65bfaecdde1d42e7af42d9deb06.pdf>
- Taylor, L. (2013). Interactive communication. *2013 Cambridge English Centenary, Symposium on Speaking Assessment*. Retrieved from <http://events.cambridgeenglish.org/cambridge-english-centenary-symposium/files/taylor-interactive-communication.pdf>
- Turney, P.D. (2002). *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*. Institute for Information Technology National Research Council of Canada, Ottawa, Ontario. Retrieved from <https://arxiv.org/ftp/cs/papers/0212/0212032.pdf>
- Tweepy: Twitter for python. (2018). Retrieved from <https://github.com/tweepy/tweepy>
- Tweet data dictionaries. (2018). *Twitter.com*. Retrieved from <https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object>
- uniRank. (2018). Retrieved from <https://www.4icu.org/twitter/>

- Velentzas, J. & Broni, G. (2014). *Communication cycle: definition, process, models and examples*. Proceeding of the 5th International Conference on Finance, Accounting and Law. 17, 117-131. Retrieved from <http://www.wseas.us/e-library/conferences/2014/Istanbul/FINANCE/FINANCE-17.pdf>
- Velentzas, J. & Broni, G., (2014). *Communication cycle: Definition, process, models and examples*. Recent Advances in Financial Planning and Product Development, Technological Institute of Western Macedonia. Retrieved from <http://www.wseas.us/e-library/conferences/2014/Istanbul/FINANCE/FINANCE-17.pdf>
- Vidhya, K.A. & Aghila, G. (2010). A survey of naïve bayes machine learning approach in text document classification. *International Journal of Computer Science and Information Security*, 7(2), 206-211. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1003/1003.1795.pdf>
- Waller, S., Tietjen-Smith, T., Davis, J. & Copeland, M. (2008). *Urban versus Rural Community Colleges: A National Study of Student Gender and Ethnicity*. Academic Leadership: The Online Journal, Vol. 6, Article 17, Issue 4 Fall 2008. Retrieved from <https://scholars.fhsu.edu/cgi/viewcontent.cgi?article=1243&context=alj>
- Wardy, R. (1998). *The birth of rhetoric: gorgias, plato and their successors*. Series: Issues in Ancient Philosophy. Routledge; 1 edition. Retrieved from <http://www.mercaba.org/SANLUIS/Filosofia/autores/Grecia%20y%20Roma/Estudios%20sobre%20filosof%C3%ADa%20griega/Wardy,%20Robert%20-%20The%20Birth%20of%20Rhetoric.pdf>
- Waters, R. D., & Jamal, J., Y. (2011). Tweet, tweet, tweet: A content analysis of nonprofit organizations' Twitter updates. *Public Relations Review*, 37, 321– 324. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0363811111000361>
- Watkins, B. A. (2017). Experimenting with dialogue on Twitter: An examination of the influence of the dialogic principles on engagement. *Public Relations Review*, 43, 163–171. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0363811115300138>
- Weber, R. P.(1985). *Basic content analysis*. New Delhi: Sage.
- Weller, K., Bruns, A., Burgess, J., Puschmann, C., & Mahrt, M. (2014). *Twitter and society*. New York: Peter Lang

- West, R. & Turner, L. H. (2010). *Understanding interpersonal communication: Making choices in changing times*. Wadsworth Publishing: US edition. Retrieved from <https://www.ffri.hr/~ibrdar/komunikacija/seminari/West,Turner%20-%20Introduction%20to%20interpersonal%20communication.pdf>
- Zhang, D. & Lee, W. S. (2003). *Question classification using support vector machines*. In proceedings of SI-GIR'03, ACM Press. Retrieved from <https://pdfs.semanticscholar.org/860c/f3a1ad32f3c78cf0e36a9d3ca56807ddca78.pdf>
- Zhang, L., Ghosh, R., Dekhil, M., Hsu, M. & Liu, B. (2011). *Combining lexicon-based and learning-based methods for Twitter sentiment analysis*. Technical Report Hewlett-Packard Development-2011-89. Retrieved from <http://www.hpl.hp.com/techreports/2011/HPL-2011-89.pdf>

VITA

EUGENE H. PONS

Born: Long Island, New York

2001-2009	Institute for Cuban and Cuban-American Studies University of Miami Marketing Director
2002-2006	University of Miami, School of Business Masters in Business Administration
2006-2008	University of Miami, School of Communication Masters in Communication Studies
2007-2014	University of Miami, School of Communication Adjunct Faculty
2012-2014	Miami Dade College, English and Communication Department Adjunct Faculty
2012-2019	Florida International University PhD, Curriculum and Instruction Specializing in Learning Technologies
2014-2019	Miami Dade College, English and Communication Department Faculty

PUBLICATIONS AND PRESENTATIONS

Pons, A., De La Rosa, A., Vidaurre, S., Vargas, L., & Pons, E. (2017). Security and privacy implications of 'do not track', *International Journal of Information Privacy, Security and Integrity (IJIPSI)*, Vol. 3, No. 2, 2017

Pons, A. & Pons, E. (2015). Social Learning Theory and Ethical Hacking: Student Perspectives on a Hacking Curriculum, *Proceedings of the Information Systems Education Conference*, Orlando, Florida USA. Nov. 7, 2015

Pons, E. (2001). *Castro and Terrorism A Chronology 1959-1967*, Institute for Cuban & Cuban-American Studies, Occasional Paper Series, September 2001

Pons, E. (2011). *Aggressive and Passive Propaganda: Cuba and the United States: A Content Analysis in Communication*, LAP LAMBERT Academic Publishing, February 15, 2011

Thirunarayanan, M.O. & Pons, E. (2016). *Preliminary Results of a Survey on Cheating Behavior*, 27th International Conference on College Teaching and Learning (ICCTL), Jacksonville, FL April 4-7, 2016