EXPLORING SYMBOLIC REPRESENTATIONS OF IDENTITY AND COLLECTIVE BEHAVIOR IN SOCIAL MEDIA: EMOJI USE IN TWITTER

by

Melanie Swartz
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Computational Social Science

Committee:

____	Dr. Andrew Crooks, Committee Chair
____	Dr. William G. Kennedy, Committee Member
____	Dr. Dieter Pfoser, Committee Member
____	Dr. Anthony Stefanidis, Committee Member
____	Dr. Jason Kinser, Department Chair
____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
____	Dr. Fernando Miralles-Wilhelm, Dean, College of Science
Date: ________________________________	Summer Semester 2020 George Mason University Fairfax, VA

Exploring Symbolic Representation of Identity and Collective Behavior in Social Media: Emoji Use in Twitter

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Melanie Swartz
Master of Arts
George Mason University, 2015
Bachelor of Science
The Pennsylvania State University, 1999

Director: Andrew Crooks, Professor
Department of Computational and Data Sciences

Summer Semester 2020
George Mason University
Fairfax, VA

DEDICATION

To my family and friends.

ACKNOWLEDGEMENTS

First and foremost, I must thank my advisor and chair of my dissertation committee, Dr. Andrew Crooks. His scholarship and teaching style were the reason I sought out this Computational Social Science program. The tireless support and dedication to research and raising the intellectual bar of his students are what kept me in the program. I can't thank him enough and words don't express my gratitude for the time spent reviewing and guiding me through the process of both the master's and dissertation research.

To the members of my committee, Dr. William G. Kennedy, Dr. Pfoser, and Dr. Stefanidis, thank you for the thought-provoking discussions and the feedback which inspired the themes covered in this research. In addition, thank you to the entire CDS staff and CSS faculty for your support to the program over the years and for helping us students make it through the intellectual and administrative process.

For the students and alumni of the CSS program, whether it was working on class projects, Friday seminar, talks over coffee, conversations after class and sometimes during, thank you for the camaraderie and encouragement along the way. I would also like to thank the many friends, family members, and supporters who have helped make this happen. You were my continual inspiration and strength to take on and finish this endeavor.

And finally, a special thank you to Judd for your patience and support throughout this process, without which this journey would not have been possible to complete.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

Application Programming Interface	API
Computational Social Science	CSS
Eastern Standard Time	EST
Hierarchical Density-based Spatial Clustering of Applications with Noise	HDBSCAN
Natural Language Processing	NLP
Non-retweet	Non-RT
Point or Place of Interest	POI
Retweet	RT
United States	U.S.
Coordinated Universal Time	UTC

ABSTRACT

EXPLORING SYMBOLIC REPRESENTATION OF IDENTITY AND COLLECTIVE BEHAVIOR IN SOCIAL MEDIA: EMOJI USE IN TWITTER

Melanie Swartz, Ph.D

George Mason University, 2020

Dissertation Director: Dr. Andrew Crooks

Social media is ubiquitous in the world today; however, analysis of social media typically focuses on content, activity patterns of users, and online communities arising from networks of interactions. With the development of computer mediated communication, emojis are now a part of the digital language of social media, yet emojis are often overlooked in social media analysis. Building on existing computational social science research, this dissertation adds new knowledge by focusing more directly on the behavior of emoji use in social media. This analysis reveals how emojis provide intrinsic cues about individual identity and collective behavior, and contributes novel methodologies that can be used to analyze and compare emoji use across users and in documents regardless of the language of accompanying text.

1 INTRODUCTION

The following chapter provides the context for this dissertation and is organized as follows. I begin with an overview, in Section 1.1 of the research area and knowledge gap related to emoji use in social media addressed by this dissertation. In Section 1.2, I explain my motivation to study the behavior of emoji use from a computational social science perspective. I summarize contributions of this research in Section 1.3. The topic areas and research questions guiding this dissertation are presented in Section 1.4. Section 1.5 introduces the remaining chapters.

1.1 Overview

Social media use is ubiquitous in the world today and analyzing it offers an unprecedented peak into the digital fabric of society, revealing insights into the patterns and dynamics of human behavior arising from the interactions of users communicating and engaging within online digital communities. These online communities form around a variety of topics of interest ranging from current events, natural disasters, politics, sports, and tourism (e.g., Crooks et al., 2013; Padilla et al., 2018; Stier et al., 2018; Vraga et al., 2018; Wood et al., 2013; Yuan & Crooks, 2018; Zhao et al., 2011). These communities emerge out of the interactions of users from a variety of backgrounds (Baym, 1998; Duggan & Brenner, 2013). While engaging on social media via posts, reading, liking, sharing, and replying, users leave behind digital traces about themselves and their communities (Hepp et al., 2018).

Collectively these traces can be analyzed to provide situational awareness for current events and social context to places (e.g., Croitoru et al., 2013; Gazaz et al., 2016; Jenkins et al., 2016; Stefanidis et al., 2013). They can be utilized for behavior modeling to create a profile of a user's activities, locations, and connections (e.g., Kavak et al., 2018; Kim et al., 2020; Rajabi et al., 2019). Even the text contained in the content posted reveals information about a user's personality and emotional state based on the linguistic style of the words used, and interests from the topics discussed (Azucar et al., 2018; Pennebaker et al., 2003). The visual media in posts, such as memes, videos, and images, have been studied to identify how they influence behavior dynamics and diffusion across the social network (e.g., Guadagno et al., 2013; Highfield & Leaver, 2016; Kaneko & Yanai, 2016). While much study of social media has focused on user activity patterns (e.g., Preoţiuc-Pietro & Cohn, 2013), network analysis of the interactions forming online communities (e.g., Himelboim et al., 2017), and topic modeling from analysis of the textual content (e.g., Hong & Davison, 2010), the analysis of emojis within social media content is often overlooked.

Emojis are glyph representations of various symbols, objects, face gestures, food, animals, and more. Social media research which takes emojis into account typically focuses on the emojis that are digital representations of face- or body-gestures (e.g., Gawne & McCulloch, 2019), treating them as indicators for emotion, sentiment, or sarcasm (e.g., Felbo et al., 2017), which only considers a small subset of emojis and does not take into account the remaining non-anthropomorphic emojis. From existing research, differences in which emojis are used has been associated with sub-groups of users based on gender and skin tone (Barbieri & Camacho-Collados, 2018), but no one has compared emoji use with other social factors such as religion, sexual orientation, and political ideology. And while there have been broad brush studies to identify the emoji most often used by large groups of users even at the country level, the patterns of emoji use and variation of which emojis are used has not been compared at the individual level. Similarly, it is not known if users are consistent in emoji user and if emoji use changes based on where a user is posting. There has been a lot of focus on the meaning and interpretation of individual emojis (e.g., Barbieri et al., 2016) and even the impact on meaning from nearby words (e.g., Miller et al., 2017). However, what has not been studied is differences in meaning associated with the structure of emoji use, such as the placement, order, and repetition of emojis within a social media post.

There is a major gap in existing social media and emoji research for the analysis of the behavior of emoji use in social media. While most of the research on emojis has focused on individual emojis, there is still much more to be understood about the behavior and communication styles of emoji users and what these behaviors and patterns reveal about the individual user and the online community. Analysis of the behavior of emoji use can reveal insights about both individual and group patterns of communication styles which provide cues for social identity and the collective behavior and processes shaping emoji use. This dissertation addresses this gap with a computational social science (CSS) perspective.

1.2 Motivation

This dissertation is within the domain of CSS which analyzes social behavior giving rise to complex systems utilizing computational and data science techniques. Building upon existing CSS research related to social media (summarized in Chapter 2), this dissertation adds to this body of knowledge and addresses the gap in research described in Section 1.1. This research analyzes the behavior of emoji use in social media from a computational social science perspective by viewing emojis as part of the complex adaptive system of language. It also considers the social complexity of individual and collective emoji use in social media as from interactions of users and feedback loops from social norms. In addition, this research also takes a multidisciplinary approach in combining social science theories with computational and data science techniques.

CSS is the interdisciplinary study of social and behavioral relationships and interactions of social phenomena using social theory and computational techniques in order to advance the understanding of society and social dynamics (Cioffi-Revilla, 2014). A key aspect of CSS is the theory of complex adaptive systems which emerge out of the non-linear interactions of individuals with limited knowledge but are also adapting based on the result of these interactions and feedback loops (Holland, 2003; Miller & Page, 2007; Simon, 1952). This complex adaptive system of language evolves from the interactions of heterogenous individuals which results in the emergence of linguistic styles at the social group level (Ellis et al., 2009; Levin, 1999; Saetre & Browning, 2008). Through these interactions forms the identity and norms of a social group (Axelrod, 1984; Simon, 1952) as the members of the groups adapt their linguist styles and behaviors to accommodate to being part of the group (Baym, 1998; Revay & Cioffi-Revilla, 2017), thus resulting in the communication styles and behaviors unique to that group (Khalid & Srinivasan, 2020). Applying this perspective to the analysis of the communication style on social media with respect to the development and adaptation of social norms within the online community is a needed contribution (Geber & Hefner, 2019).

Emoji use is associated with social complexity and the complex adaptive systems of language arising from social media use and interactions with online communities. From this perspective, this dissertation is about identity, social groups, and the patterns of behaviors arising from the interactions and feedback loops between them as users engage in social media by using linguistic styles incorporating emojis, which together represents the language of users on a social media platform as a complex adaptive system. With this dissertation I propose that emojis provide cues about individual and collective identity and behavior based on the way emojis are used as part of communication style. These styles are influenced by social norms, technology, and online interactions. These interactions also can result in the emergence of shared styles and patterns of emoji use specific to online communities of user groups. Studying these communication patterns with a CSS perspective enables insight in to these dynamic social processes and enables greater understanding of the behavior of emoji use in computer mediated communication.

This dissertation also draws from multiple social science disciplines in order to understand and represent individual and collective behavior of emoji use arising from the social media interactions. Within the realm of CSS this research draws on social science theories for the processes contributing to communication patterns on social media and uses computational methods to analyze social media data in order to understand the behavior of emoji use. The analysis of the behavior of emoji use combines social science theory, communication studies, and uses computational and data science techniques to present novel contributions to field of CSS, studies on social media, and emoji analysis, as shown in Figure 1.1.

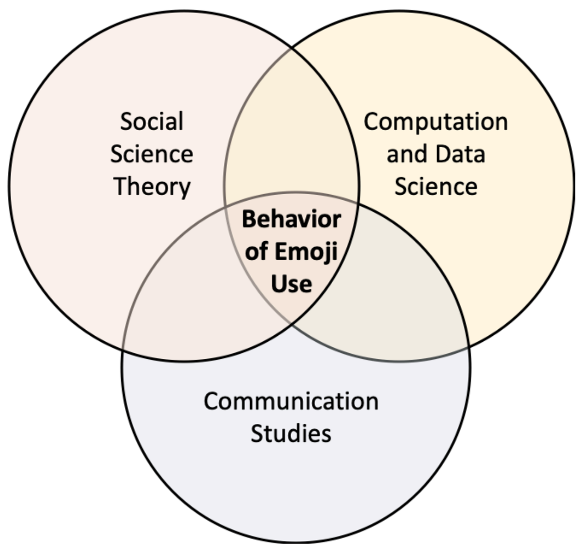


Figure 1.1 Research domain areas of this dissertation.

From the social science perspective this research draws from social-identity theory (Stets & Burke, 2000) and symbolic interactionism (Blumer, 1962), which are further explained in Chapter 2. This research also applies perspectives of communication studies from socio- and psycho-linguistics regarding the linguistic style and accommodation (Giles, 1973; Levelt & Kelter, 1982; Niederhoffer & Pennebaker, 2002; Pennebaker et al., 2003). In addition, the computational and data science techniques for data mining, natural language processing (NLP), geo-temporal analysis, and visualization enable the analysis of the emojis in social media data across millions of users (e.g., Croitoru et al., 2017; Crooks et al., 2015; Gazaz et al., 2016; Stefanidis et al., 2013).

1.3 Contributions of this Dissertation

This dissertation adds to the body of knowledge by researching the less studied area of the behavior of emoji use in social media by combining theories and techniques from multiple disciplines, and applying them with a perspective of CSS. Specifically, this research examines the individual and collective properties, behaviors, and interactions associated with the patterns evolving from the behavior of emoji use in a way that has not been explored before. Contributions of this dissertation include that this research demonstrates the importance of including emojis in analysis as more than just a barometer for sentiment or sarcasm. In addition, this is the one of the first studies examining how the use of emojis provides cues about the identity, behavior, and communication styles of individuals. This dissertation also shows how despite individual preferences and behavior of emoji use, the influence of social norms and interactions result in feedback loops that collectively shape and result in distinct behavior patterns across user groups and document types.

In addition, the computational methodologies developed to enable this research are made available via GitHub at https://github.com/msemoji and have been documented in recent publications so that they can be used to analyze and compare emoji use within any type of document and regardless of the language of accompanying text. The research and methodologies of three of the chapters in this dissertation have already been published. Chapter 3 was presented at the Eighth IEEE International Workshop on Semantic Computing for Social Networks and Organization Sciences: From user information to social knowledge in 2020. It is published as part of the proceedings of 2020 IEEE 14th International Conference on Semantic Computing (ICSC ‘20). The content of Chapter 4 was published in the proceedings for the International Conference on Social Media and Society (SMSociety ‘20): Diverse Voices: Promises and Perils of Social Media for Diversity. Chapter 5 will be presented and published in the proceedings for the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRIMS ‘20). The citations for these three papers are:

- Chapter 3: Swartz, M., & Crooks, A. (2020, February). Comparison of Emoji Use in Names, Profiles, and Tweets. In *2020 IEEE 14th International Conference on Semantic Computing (ICSC ‘20)* (pp. 375-380). IEEE.
- Chapter 4: Swartz, M., Crooks, A., & Kennedy, W. (2020, July). Diversity from Emojis and Keywords in Social Media. In *International Conference on Social Media and Society* (pp. 92-100).
- Chapter 5: Swartz, M., Crooks, A., & Croitoru, A. (2020, October). Beyond Words: Comparing Structure, Emoji Use, and Consistency Across Social Media Posts. In *2020 International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRIMS ‘20)*. Springer.

1.4 Research Questions

This dissertation analyzes the individual and collective behaviors and patterns of emoji use in social media, specifically using Twitter as a representative use case, from a CSS perspective. In order to accomplish this research, there are four key objectives which are guided by three overarching research questions. The four research objectives which enable the analysis for this dissertation are: (1) formation of a baseline of emoji use in a social media data sample, (2) develop a framework to enable comparison of emoji use and structure in documents, (3) identify and describe emoji use relating to individual or collective identities in social media, (4) identify communication patterns and emojis related to particular events and types of places. Addressing these four objectives enables answering the following three research questions.

RQ1: What are the differences in emoji use across users and documents in social media?

The first research question significantly extends the analysis of emojis in social media beyond the existing state of the art by focusing on the behavior of emoji use through comparison and identification of differences in emoji use across users, type of documents, and user roles. To address this question, I analyze emojis used in Twitter user names, user profile descriptions, tweets, and retweets from over 3 million users and for a variety of user roles such as news organizations, marketers, students, bots and others. In order to compare emoji use, I developed a methodology that describe emojis based on attributes, such as the Unicode group, sub-group, type, color, direction; and another methodology to represent the way emojis are used in a document based on their position, order, and repetition. This research lays the foundation for comparing which emojis are used and in what way and establishes a baseline identifying differences in emoji use across users and document types.

RQ2: In what way do emojis reveal cues about social identity and individual communication style preferences?

Building on existing research that indicates differences in emoji use based on gender and skin tone (Barbieri & Camacho-Collados, 2018), this research explores emoji use associated with additional diversity characteristics such as religion, sexual orientation, and political ideology. It also extends the field of social psychology and linguistics by identifying patterns of emoji use as indicators of a signature of behavior. To address this research question, I develop a diversity language model that compares emojis and keywords for characteristics associated with gender, skin tone, religion, sexual orientation, and political ideology. This enables analysis of reported user-identity and compares differences in emoji use versus traditional text mining approaches using only keywords. In addition, I developed a methodology to perform structural content analysis on social media posts to identify user or group signatures of behavior for emoji use. I also developed a metric to measure consistency of emoji use for an individual or user group as a way to compare and group users based on the structures of their communication style. The focus on structural content analysis of social media posts and emoji use is a new contribution to the field of computational linguistics and builds on social linguistics theory and social media topic modeling that focus on just the words used by an individual.

RQ3: What are the collective patterns and behaviors that arise from individual emoji use and what do they reveal about social norms?

Building on the methodologies and insights gained by addressing the previous two research questions, this one focuses on identifying the collective patterns of emoji use rising from the social cues and norms impacting individual emoji use behaviors. To address this research question, I examine emoji use related to specific events such as observing an eclipse, participating in International Women's Day, symbolism of solidarity and national pride, and role of geography and shared place. This requires collection of datasets specific to these events as well as temporal and geographic analysis of emoji use. In addressing this question, bridges the gap from individual emoji use to analysis of the collective behavior of emoji use and identifying evidence of the impact of social norms.

1.5 Dissertation Overview

The structure of this dissertation is as a collection of chapters that address different aspects of the research areas and questions raised in Section 1.4. Next, Chapter 2 provides the background of the social science theories and communication studies framing this research. It also summarizes the history of emojis and provides an overview of the current state of the art of emoji related research. A baseline analysis and comparison of emoji use across users and document types associated with the social media platform Twitter is presented in Chapter 3. Chapter 4 shows how emojis provide cues about social identity and how their use compares with keywords associated with diversity characteristics for gender, race, skin-tone, religion, sexual orientation, and political ideology. Then, Chapter 5 provides a framework for measuring consistency of user behavior for structure of social media content and emoji use represented as emoji attributes, position, order, and repetition within a document. This is followed by, Chapter 6 which demonstrates the collective use of emojis as first hand observations of a lunar eclipse event. Chapter 7 examines the use of emojis as symbols and icons during national, cultural, religious, and social events. The relationship between emoji use and place, such as types of places of interest, are examined in Chapter 8. Finally, Chapter 9 concludes the dissertation and synthesizes the previous chapters not only to summarize the results, but to show how together they answer the research questions.

2 BACKGROUND AND RELATED WORK

This chapter provides the background knowledge and summary of related work which frames the analysis in the remaining chapters. It begins with an overview of the recent research related to social media, Section 2.1. This is followed by a brief review of identity of self on social media in Section 2.2 and an overview of communication studies with an emphasis on the psycho-linguistics of computer mediated communication in Section 2.3.Then the social science theory related to symbolic interactionism and how meaning are derived from symbols and sings in communication via interactions are presented in Section 2.4. A brief history of emojis and the current state of the art of research related to emojis is presented in Sections 2.5, and 2.6, respectively.

2.1 Recent CSS Research on Social Media.

Recent research related to CSS demonstrates the power of using social media to provide context for human behavior in relation to current events, place, and online communities, and can be used in generating datasets for social simulation. For instance, Crooks et al. (2013), Stefanidis et al. (2013), and Croitoru et al. (2013) demonstrate that social media provides valuable context as a social sensor for situational awareness about places and current events, such as earthquakes. Social media can also be used to provide social context to places in order to understand where people go, the types of place, and the form and function of cities (e.g., Crooks et al., 2016; Jenkins et al., 2016; Stefanidis et al., 2016). Social media data can be used to generate signatures of behavior, such as using user locations to identify residences (e.g., Kavak et al., 2018), and using topics discussed in social media as fingerprint for places (e.g., Gazaz et al., 2016). Combining user activity and content reveals online communities around topics such as vaccines (e.g., Yuan & Crooks, 2018), cancer (e.g., Vraga et al., 2018), Zika (e.g., Stefanidis et al., 2017), and can even be used to compare bots (e.g., Schuchard et al., 2018; Yuan et al., 2019). These insights from social media analysis can then be used inform behavior modeling such as how people cognitively respond to misinformation (e.g., Tulk et al., 2018) and for generating data to be used for social simulations (e.g., Kavak et al., 2019; Kim et al., 2020) such as for disaster modeling (e.g., Burger et al., 2019; Burger et al., 2017).

2.2 Social Identity and Communication

As the research presented in this dissertation is based on the study of individual and group identity and patterns of communication arising from interactions on social media. The challenge of many social science theories is that many were formed based on in person face-to-face interaction via communication (Hier, 2005). With the proliferation of digital communication, many of these theories have been adapted to the study of online social identity, communities, and their interactions (Walther, 2012). Marwick and Boyd (2011) interviewed users of Twitter how they manage the presentation of self (Goffman, 1990) which is established by interactions on social media while users also balance authenticity versus the identity they want to project. Their results reveal that users adapt a strategy to determine what and how to communicate based on perceptions of who their audience is and goals for tweeting. Some of these goals may be to attract followers, to promote a brand or message, to share personal information with friends, to provide news, share ideas and information about things of interest.

2.3 Communication and Linguistic Style

Communication is the processing of information exchange during interactions (Berea, 2019). As an individual communicates, the order of words, specific aspects of language such as structural markers, and non-verbal features such as face gestures can be considered as part of an individual's linguistic style (Maynard & Peräkylä, 2006). Linguistic style forms unconsciously and as a result of this interactions (Levelt & Kelter, 1982) and can even be a signature for an individual (Pennebaker et al., 2003). As people engage in conversation, individuals may adapt their words, non-verbal cues such as semiotics (Morris, 1946). In social groups, these accommodations result in a convergence of communication style (Giles, 1973) as members of the group use linguist style matching such as the use of shared words, cues, and patterns of speech (Niederhoffer & Pennebaker, 2002). While much research has focused on face-to-face interactions, Walther (2012) argues that technology and the use of computer mediated communication impacts linguistic style accommodation as users have to adapt based on technology such character limits or availability of emojis. Danescu-Niculescu-Mizil et al. (2011) results show that there is linguist style matching of users in Twitter. Khalid & Srinivasan (2020) revealed that online communities based on a social media platform and even topic have a distinct linguistic style. Most recently, Barach et al. (2020) analysis of linguistic style accommodation in tweets based on combination of pronoun usage and emojis with a face, gesture, or objects indicates differences in users based on communication patterns of tweet style with emojis. While these communication studies typically focus on the types of words such as and number of words used, this dissertation extends these theories to include emojis and also structural aspects of communication style such as the order and types of content incorporated into social media posts.

2.4 Symbolic Interactions and Emojis

This dissertation argues that in addition to the meaning attributed to emojis, the way in which emojis are used within computer mediated communication such as social media are also shaped as a result of these interactions and social processes. These interactions shape not only shape social identity (Stets & Burke, 2000), they also give meaning to shared symbols and behaviors which come to define a social group or society through a process called symbolic interactionism (Blumer, 1962). Symbolic interactionism is the way in which non-verbal cues and signs are attributed with a value that can be interpreted by the receiver. Often this takes place through typically communication, but also from experience, and then the receiver chooses whether to accept the meaning as described, rejects it, or adapts the meaning in some way. I extend this theory to emojis as the way in which meaning is attributed to emojis. While each user may have their own interpretation of an emoji meaning, this meaning is shaped based on their previous experiences including any previous definitions stated for what the emoji means. However, emojis can be polysemic and take on multiple meanings, changing based on nearby text, which is already well researched (Tigwell & Flatla, 2016). With this dissertation, I offer that the way in which emojis are used in terms of where they are placed, the order, and repetition of nearby emoji with similar attributes also may impact meaning.

2.5 History of Emojis

Emojis are glyph representations available on a digital device that supports the rendering of various symbols and common objects as pictographs. Emojis are intended to be a visual form of communication to be used along with text or on their own in an application that supports typed characters. Emojis originated in 1997 to enable mobile device communication to include more than just text to express emotion with anthropomorphic face-like symbols and select everyday objects like food items (Blagdon, 2013). Figure 2.1 shows the original color emojis released on the NTT DOCOMO mobile device in 1999 (Kurita, 1999). Since the adoption of the emoji character set by Unicode in 2010 (Unicode, 2010), emojis have become available on nearly all mobile devices such as the iPhone in 2011 (OSX Daily, 2011) and social media platforms like Twitter.

Figure 2.1 Original 176 color emojis from 1999.

Emojis have a unique role in communication as they can represent common objects, and expressions to “anthropomorphize” written text in a universal visual way that otherwise would be lost or not make sense with just the text alone. This is evident by the use of the emoji “Face with Tears of Joy”, 😂, as the Oxford 2015 word of the year (Oxford Dictionaries Blog, 2015). Use of emoji in Twitter is very common and has been tracked in real time since 2013 by http://emojitracker.com/ which shows a tally of how many times an emoji was used and clicking on the emoji will display the most recent tweets using that emoji and even in 2020 it shows that “Face with Tears of Joy” is the most popular emoji, Figure 2.2. In addition, new emojis are released each year by Unicode thus further adding to the interests of emoji users.

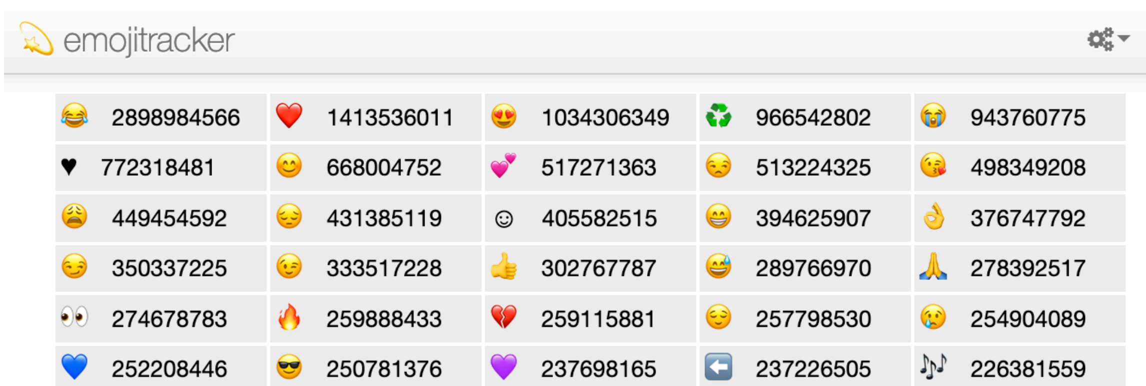


Figure 2.2 Most used emojis in Twitter as shown by emojitracker.com on 3 July 2020.

2.6 Current State of the Art of Emoji Related Research

The topics covered in academic literature on emojis primarily falls into two broad areas which are discussed below: 1) emoji interpretation and language analysis focused on semantics, sentiment, impact of device platform emoji rendering on interpretation, and disambiguation; and 2) emoji use and human factors such as demographics, personality, and psychology. Additional topics related to emoji research include comparison of emoticon and emoji usage (e.g., Pavalanathan & Eisenstein, 2015), finding street gang members (e.g., Balasuriya et al., 2016; Wijeratne et al., 2016) and applying machine learning to classify emojis and prediction of emoji use (e.g., Barbieri et al., 2018; Cappallo et al., 2018).

Emoji interpretation and language. Although there is a Unicode standard for emoji definitions and digital character representation, these are only a suggestion for the depiction of the emoji (Davis, 2018). As a result, each device platform and application can render emoji as a modification of the Unicode specification. These various emoji renderings are shown on website Emojipedia.org, Figure 2.3.

Figure 2.3 Differences in emoji rendering of Grinning Face emoji as of July 2020.

This variation in renderings can result in variations in how and which emojis may be used (Lu et al., 2016; Tigwell & Flatla, 2016) and different interpretations of what the same emoji may mean across different applications (Miller et al., 2016). The majority of language related research on emojis has focused on meaning and sentiment. Some emojis may have multiple meanings (Ai et al., 2017; Miller et al., 2017; Riordan, 2017) or can take on different meanings based the text near the emoji (Barbieri et al., 2016; Donato & Paggio, 2017; Tigwell & Flatla, 2016) or the Unicode definition of the emoji in comparison to the nearby text (Eisner et al., 2016; Wijeratne et al., 2016). Many of the face or people based emojis that appear to relay an emotion have been a focus of study for the sentiment of emojis (e.g., Felbo et al., 2017; Kimura & Katsurai, 2017; Kralj Novak et al., 2015), sentiment inferred from the text based on the emoji (e.g., Hernandez et al., 2012; Rodrigues et al., 2017; I. Wood & Ruder, 2016), and how emojis make people feel (e.g., Barbieri et al., 2018; Hu et al., 2017).

Emoji use and human factors. A handful of articles have explored human factors related to emoji use and its impact on interpretation. For example, Chen et al. (2018) explored the gender differences in use of emojis on Android devices, while Herring & Dainas (2018) revealed how the different sexes interpret emojis one sees in a message. Some studies have also explored the role of personality in emoji use and how it compares to the big 5 personality traits (e.g., Li et al., 2018; Marengo et al., 2017). Kaye et al. (2016) delved more in to the use of emojis as a means for psychological assessment and communication. Ljubešić and Fišer (2016) and the report on the blog of the keyboard manufacturer Swiftkey (2015) provide a snapshot of top used emojis by country. Zhou et al. (2017) focused on emoji use in China while Cramer et al. (2016) focused on the U.S.. Ljubešić and Fišer (2016) also compared emojis to the wealth index of countries. With the availability of skin-toned emojis in 2016, users are more likely to use emojis that look similar to the their own (Barbieri & Camacho-Collados, 2018; Robertson et al., 2018).

Variations in use of emoji by gender and age have also been described (Herring & Dainas, 2018; Medlock & McCulloch, 2016; Na’aman et al., 2017a).

Even though emojis pre-date social media analysis of emoji use is relatively recent to the past few years. There are still a number of topics and factors that can be explored to provide more insight into the how the usage and meaning of emoji varies and the implications of individual emoji use to provide additional insight about collective social behavior. Analysis of social media for computational social science is extended by this dissertation to examine the behavior of emoji use.

3 COMPARISON OF EMOJI USE IN NAMES, PROFILES, AND TWEETS[1]

Online social networking applications are popular venues for self-expression, communication, and building connections between users. One method of expression is that of emojis, which is becoming more prevalent in online social networking platforms. As emoji use has grown over the last decade, differences in emoji usage by individuals and the way they are used in communication is still relatively unknown. This chapter fills this gap by comparing emoji use across users and collectively in user names, profiles, and in original and re-shared content. It presents a methodology that enables comparison of semantically similar emojis based on Unicode emoji categories and subcategories. This methodology is applied to a corpus of over 44 million tweets and associated user names and profiles to establish a baseline which reveals differences in emoji use in user names, profile descriptions, non-retweets, and retweets. In addition, the results of this analysis reveal emoji super users who have a significantly higher proportion and diversity of emoji use. This chapter provides a baseline analysis and methodology for summarizing emoji use and enables systematic comparison of emojis across individual user profiles and communication patterns, thus expanding opportunities for semantic analysis of social media data beyond just text.

[1] This chapter is based on: Swartz, M., and Crooks, A. (2020). Comparison of Emoji Use in Names, profiles, and Tweets. In *2020 IEEE 14th International Conference on Semantic Computing (ICSC),* pp. 375-380. IEEE. https://doi.org/10.1109/ICSC.2020.00075

3.1 Introduction

Social media data enables researchers the opportunity to analyze public discourse, social norms, and trends often centered around current events. Social networking sites such as Twitter enable users to communicate by sharing information about themselves in the user profile and to engage with users and content in the form of posting, replying, or tagging other users in content that includes text, links, video, and images. It is not uncommon for users to include emojis alongside or in place of textual characters as a popular form of self-expression and communication on social networking platforms. Emoji use in social media is especially popular with some users because an emoji can be an effective way to express sentiment, sarcasm or feelings which are not easily conveyed as text (Novak et al., 2015). In addition, some social networking sites, such as Twitter, limit the size of the content or number of characters allowed and emojis can be more efficient than their textual equivalent (Rodrigues et al., 2017).

Although emojis originated in the late 1990s, their use only recently become popular on social networking sites. The choice and ability to incorporate emojis into social media content have become more prevalent since the adoption of Unicode standards for emojis in 2010 (Unicode, 2010), combined with the availability of emoji keyboards on mobile devices and emoji rendering on social media platforms. In addition, several new emojis are approved by Unicode each year, further adding to the variety of emojis available for users (Unicode, 2019). Despite the popularity of emoji use in social media, limited research has focused on analysis of the behavior of emoji use or how to compare emoji use across users or documents. These are the two main areas of research I focus on in this chapter.

The overall contributions of this chapter are: (1) a methodology to extract, aggregate, and compare emoji use across a collection of documents based on Unicode emoji category and subcategories, (2) a baseline of statistics of emoji use in user names, profile descriptions, and tweets, and (3) comparison of emoji use as categories and subcategories between users and content a user shares in the user name, profile description, retweets and non-retweets.

The remainder of this chapter is organized as follows. In Section 3.2, briefly highlights related work which is followed by the methodology to aggregate and analyze emojis (Section 3.3). The results and comparison of emoji use applied to a corpus of tweets and user profiles related to the 2018 U.S. midterm elections are presented in Section 3.4. Section 3.5 concludes this chapter and identifies areas for further research.

3.2 Related Work

This section provides a brief review of how emojis have been studied with respect to behavior of emoji use, methods for analyzing emojis in text, and comparison of emojis.

3.2.1 Behavior of Emoji Use

Most research on the behavior of emoji use to date has focused on summarizing the most frequently used emoji at broad aggregate scales of analysis. Such research has revealed differences in emoji use by cultural (Guntuku et al., 2019), gender (Chen et al., 2018), and at the country level (Ljubešić & Fišer, 2016). At the individual level, research has correlated emoji use to social identity (Ge, 2019). Other researchers have identified personal preferences on emoji use related to marketing and how people respond to them (Dogan & Collins, 2019). While previous research indicates differences in emoji use, there is limited research that focuses on individual behavior of emoji use such as how many and consistency of emojis used as well as differences of emoji use based on document types. These are the behaviors of emoji use I explore and present in this chapter.

3.2.2 Content Analysis

Content analysis on documents containing emojis often focuses on sentiment, which typically utilizes the subset of emojis representing faces or gestures. These emojis are used as a barometer to assess the magnitude of positive or negative sentiment which is then applied to the whole document or words in proximity to the emojis (Kaye et al., 2016; Morstatter et al., 2017). Other content analysis approaches with respect to emojis is to perform text analysis and convert the emoji representation to the emoji Unicode full or short name (Miller et al., 2017) or to omit emojis all together (Yuan et al., 2019b). A limitation of content analysis that applies only sentiment to face and gesture emojis is that these anthropomorphic emojis account for only 17% of all emojis, thus ignoring the remaining emojis. Content analysis that converts emojis into textual names or other labels may misrepresent the intended meaning of the emoji. These approaches do not fully utilize the value of all emojis for content analysis.

In regards to analysis of emojis related to semantics, research has focused mainly on identifying differences in meaning of individual emojis that can arise from varying interpretations based on culture and emoji rendering by device (Hillberg et al., 2018;

Morstatter et al., 2017). Not yet fully explored by current semantics research is examination of how emoji use differs across individuals and based on document type. The methodology and results presented in this chapter addresses this gap by considering emojis grouped by Unicode categories and subcategories as semantically similar. Further, I use these groupings to enable comparison of emoji use.

3.2.3 Comparing Emoji Use

Current approaches comparing emojis have focused on individual emoji or the most frequently used emojis. Research has compared the appearance of individual emoji to actual human gestures, actions, and facial expressions (McCulloch & Gawne, 2018; Pohl et al., 2017). At the document or user group level such as gender or country, emoji use is often compared based on the most frequently used emojis (Chen et al., 2018; Guntuku et al., 2019; Ljubešić & Fišer, 2016). Comparing the most frequent or distinctive emojis by user group or occurring within a document is useful to visualize differences of specific emojis. However, this approach can be challenging when there is a large variety of emojis used. I feel that in addition to existing approaches, aggregating emojis based on semantically similar Unicode emoji categories and subcategories provides a useful summarization of emoji use and enables semantic comparison across documents and users.

3.3 Methodology

To summarize and compare emoji use in documents and per user, this methodology consists of four parts: 1) collect data, 2) extract emoji, Unicode category, and subcategory, 3) summarize emojis per document and user with aggregation, and 4) compare emoji use across documents, user groups, and document types. Figure 3.1 shows the four parts of this workflow.

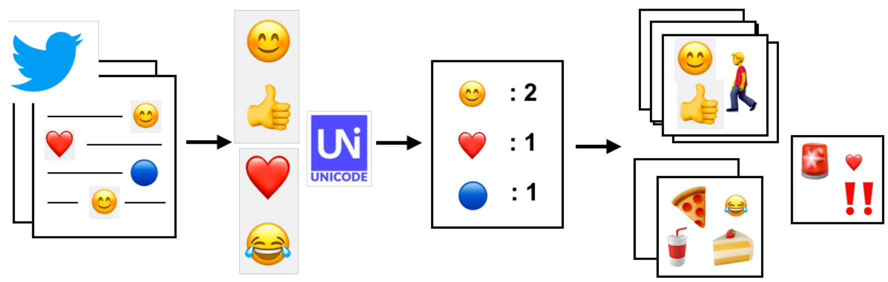


Figure 3.1 Workflow to assign Unicode emoji category and subcategory.

3.3.1 Data Collection

For this analysis, I analyzed emoji use in tweets collected from the free public Twitter streaming application programming interface (API). I then used the Twitter standard search API to collect user profiles of all the users who authored tweets or were retweeted in this corpus. The streaming API enables researchers to query for tweets containing keywords or based on geographic coordinates. The streaming API returns a portion of all tweets based on Twitter's filtering process (Morstatter et al., 2013). After this step, I created a dataset containing selected fields from the collected tweets (i.e., tweet text, tweet author user screen name, and retweeted user screen name, if present) and from the user profile (user screen name, user name, and user profile description).

3.3.2 Extract Emoji, Categories, and Subcategories

To extract a unique set of emojis contained in each tweet, user profile description, and user name, the following automated steps were undertaken using Python: 1) used regular expressions library (regex) to return the majority of emojis; 2) filtered out non-emoji characters returned from step 1; 3) extract additional emojis not included in regex, such as keycaps; and 4) rebuilt compound emoji sequences and flags. A compound emoji sequence typically is rendered as a single character but is comprised of multiple emojis, for example 👨‍👩‍👧‍👧 is made up of the individual emojis 👨 👩 👧 👧. Emojis with gender or skin-tone are another; this emoji 👩🏾‍💻 is composed of 👩, skin-tone modifier, and 💻. In addition, country flag emojis are made up of two emojis, such as 🇺 and 🇸 for the flag of the United States of America 🇺🇸. Finally, 5) to generate the unique list of emojis, I removed duplicate occurrences and sorted the unique emojis. this rationale for doing this was to enable comparison and counting of emojis used per user and tweet during aggregation.

For each unique emoji extracted, I identified its corresponding Unicode emoji category and subcategory. Each emoji belongs to one of 95 Unicode emoji subcategories. Each subcategory belongs to one of the nine Unicode emoji categories: Activities, Animals & Nature, Flags, Food & Drink, Objects, People & Body, Smileys & Emotion, Symbols, Travel & Places. These categories are similar to the emoji groups displayed on an emoji keyboard but not identical. To illustrate how emojis correspond to a category and subcategory, Figure 3.2 shows the Unicode category Smileys & Emotion and corresponding subcategory of face-smiling for two emojis. I used the Unicode data files version 12.1 (Davis & Edberg, 2019) to generate the full mapping of each of the 3000 plus emojis to their respective category and subcategory.

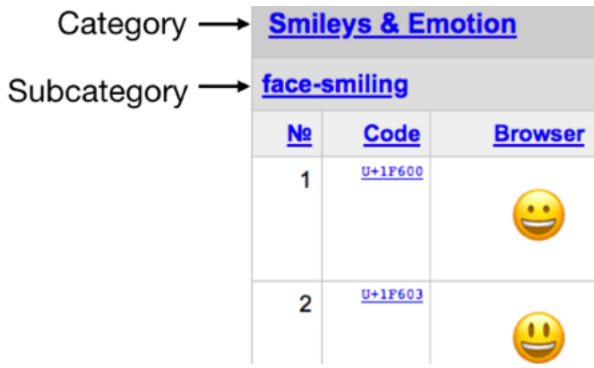


Figure 3.2 Unicode emoji category and subcategory example.

3.3.3 Aggregate by Unit of Analysis

As this chapter summarizes and compares emoji use by user and document type, I consider user and document type as this units of analysis for aggregation. The document types analyzed are the user name, profile description, retweet content, and non-retweet content. For each user, I generated a list of their unique emojis, categories, and subcategories they used per document type. Next I created a sorted list in descending order based on the count of user retweets and non-retweets per emoji, categories, and subcategories. Emoji use was also summarized across all users per each document type. For emoji, category, and subcategory, I created a list for each that included counts and percent of users containing each per respective document type and sorted in descending order based on counts. It should be noted, however, that this aggregation does not yield a sum of 100 percent as some users use emojis from multiple categories and subcategories in the same tweet.

3.3.4 Comparison of Emoji Use

After the emojis for each user and document type have been aggregated, the final step is to compare their use. I summarized and compared emoji use by categories and subcategories using various methods including visualization and summary statistics. Furthermore, to assess the similarity of emoji use, I measured similarity of emojis and subcategories that were used by at least 50% of the user population. To do this, I used the Jaccard similarity coefficient, Equation 3.1, to measure similarity of the most frequently used emojis and also subcategories between user names, user profiles, retweets, and non-retweets. This metric measures the amount of overlap between two sets of values, A, B. It is calculated by dividing the number of values in common (intersection, represented as ∩) by the total number of unique values combined from the two sets (union, represented as ∪). It returns a ratio between 0 (no overlap) and 1 (both sets contain the same values). I report Jaccard similarity measures between document types in the results.

Equation 3.1 Jaccard similarity

$$J_{A,B} = \frac{|A \cap B|}{|A \cup B|}$$

3.4 Results and Analysis

Utilizing the methodology from Section 3.3, I compared emoji use across a set of tweets and associated user profiles related to the 2018 U.S. midterm elections to discern differences in emoji use across users and document types. I analyzed behavior of emoji use at the user level by comparing percent of emoji tweets and number of unique emoji and subcategories per tweet. In addition, I also compared differences of emojis, categories, and subcategories across all users collectively for four document types: user names, user profile descriptions, retweets, and non-retweets. In this section, I present the results measuring behavior of emoji use and comparison of emojis used in names, profiles, and tweets.

3.4.1 Dataset and Percent of Emoji Use

For this analysis I used a corpus of over 44 million tweets collected between October 5 and November 6, 2018, based on keywords related to the U.S. midterm elections. Tweets were collected based on keywords such as republican, democrat, MAGA, and several Twitter user screen names of candidates running for office. The percent of tweets that contained an emoji was 8.28%. I divided the emoji tweets into retweets and non-retweets. Retweets accounted for 83% of this collected tweets. Non-retweets had a slightly higher percent of tweets containing emoji (9.54%) compared to retweets (8.03%). Table 3.1 summarizes the dataset as tweet counts and percent utilizing emojis.

Table 3.1 Dataset overview of tweets with emoji counts and percent

	Count	***Count with Emoji***	***Percent with Emoji***
tweets	44,388,440	3,675,589	8.28
retweets (RT)	36,933,494	2,964,519	8.03
non-retweets (non-RT)	7,454,946	711,070	9.54

The dataset contained tweets from 3.3 million unique users, of which 19.29% sent a tweet containing an emoji, 21.98% used an emoji in their profile description, and 13.24% used an emoji in their user name. There was a slightly higher percent of users using emojis in retweets (18.49%) compared to users with emojis in non-retweets

(17.5%). For the users that used emojis in tweets, a quarter of them (24.82%) also used emojis in their user profile description, and a little over a third (35.42%) used emojis in their user name. In this dataset, there was not a strong correlation between behavior of emoji use in a user name, profile description, and tweets, which indicates unique behaviors of emoji use associated with these document types. Table 3.2 summarizes the dataset as author counts and percent utilizing emojis.

Table 3.2 Dataset overview of authors using emoji, counts and percent

	Count	***Count with Emoji***	***Percent with Emoji***
authors of tweets	3,300,373	636,707	19.29
authors sending RT	2,673,696	494,495	18.49
authors of non-RT	1,291,726	226,091	17.5
authors with profile descriptions	2,237,222	491,907	21.98
authors with user name	2,830,888	374,905	13.24

3.4.2 Number of Unique Emojis within a Tweet

Turning to emoji use within tweets, of the 3.6 million tweets that contained emojis, just over 1 million were unique. These unique tweet texts predominantly contained one or two unique emojis, although a few contained many more, Figure 3.3. This was the same pattern in retweets and non-retweets. However, retweets had much greater diversity of emojis used across all retweets compared to non-retweets. This variation becomes more apparent in the following subsections as I show emoji use by category and subcategory.

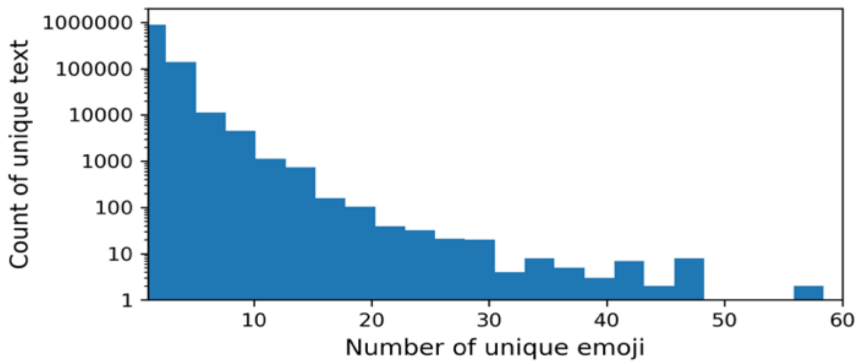


Figure 3.3 Histogram of the number of unique emojis per unique text.

3.4.3 Behavior of Emoji Use and Emoji Super Users

During the one-month timeframe of this data collection, I found that it was more common for emoji users to send both non-retweets and retweets containing emojis. With this in mind, I combined them to calculate percent of emoji tweets per user. The average percent of tweets sent by a user (including both retweets and non-retweets combined) that included an emoji was 38%. Figure 3.4 shows each user and the number of tweets they sent and the percent of their tweets that contained emojis. Out of a total of 636,000 users who used emojis in tweets, 334 (0.05%) had both a high volume of tweets (over 100 tweets sent) and a high percentage of emoji tweets (over 60%). I consider these users to be emoji super users. These super users also stood out based on the number of unique emojis used across all their tweets averaging 33.2 unique emojis compared to 4.7 unique emojis for all users.

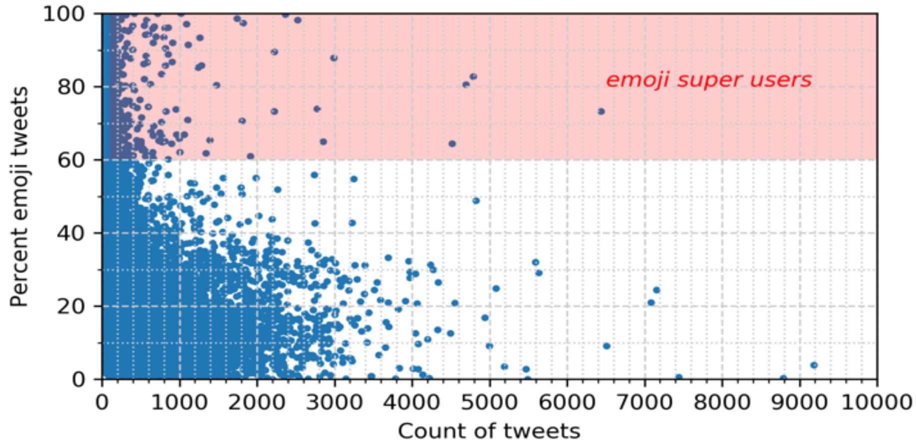


Figure 3.4 Plot of count of tweets and percent emoji tweets per user.

I then compared the number of unique emojis used between a user's set of retweets and non-retweets. The average number of unique emojis used across each user's non-retweets was only 2.2 and 5.1 unique emojis for their retweets. Emoji super users on the other hand had much greater use of emojis on average using 15.3 unique emojis across non-retweets and 25 across their retweets. This indicates most users are consistent in emoji use by using only a few unique emojis while emoji super users use more emojis.

Next I compared emojis used in names and profile descriptions and diversity of emojis used. Earlier I noted that for users of emojis in tweets, 25% used emojis in profiles and 35% used emojis in user names. However, nearly half of emoji super users used emojis in their profile description (47%) and slightly less than a third used emojis in their user name (28%). In addition, emoji super users on average also had significantly more diversity of emojis used compared to others. For non-retweets, on average emoji super users used emojis from 9.7 unique subcategories compared to only 1.7 for other users. For retweets, the super users used emojis from 13.9 unique subcategories while other users used only 2.2. This result indicates that compared to emoji super users, most users only used a few emojis and from the same few subcategories. Next I compare emoji categories and subcategories used collectively across all users for the document types of user names, profile descriptions, retweets, and non-retweets.

3.4.4 Emoji Categories

As noted in Section 3.3, there are nine overarching categories that encompass the full set of emojis. In this section I present the results of this analysis as the total proportion of users of emojis per emoji category for the four document types (i.e., user names, user profile descriptions, retweets, and non-retweets). In Figures 3.5 and 3.6, emoji categories are labeled with text, and the proportion of emoji use is displayed for retweets (orange), non-retweets (blue), user names (yellow), and profile descriptions (green). Figure 3.5 shows the relative proportion of emoji use by Unicode category per document type. It shows that there are differences in which emoji categories are likely to be associated with use in user names, descriptions, retweets, and non-retweets. Figure 3.6 shows the proportion of use for the 250 most-used emojis across all document types. The emojis are sorted by Unicode category which enables comparison between emojis within the same category and reveals that there are small groups of emojis within the same category have similar patterns of use.

In this dataset, there was a similar proportion of emoji use across document types in three categories Activities, Animals & Nature, and Flags. The other categories showed prevalence of use for a specific document type. For example, emojis from the categories Objects, Symbols, People & Body, and Travel & Places were more likely to be used in retweets. Meanwhile, categories of Smileys & Emotion and People & Body were greatest with non-retweets and were not widely used in user names and profile descriptions. Similarly, emojis from the Food & Drink category were more likely to be used in a user name, but not in non-retweets. Next, I compare groupings of emojis by subcategories within the Unicode emoji categories.

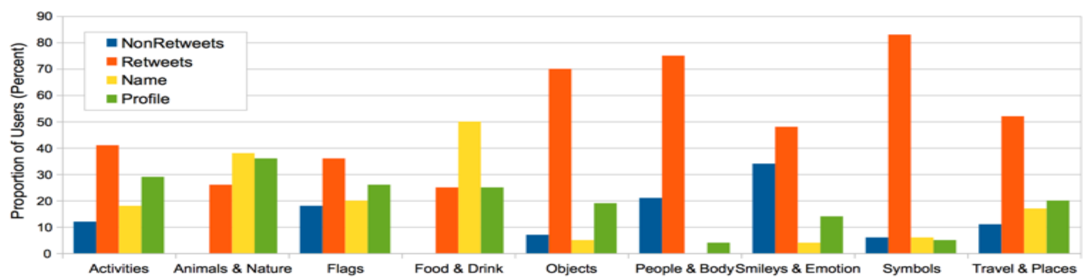


Figure 3.5 Comparison of use by emoji category per document type.

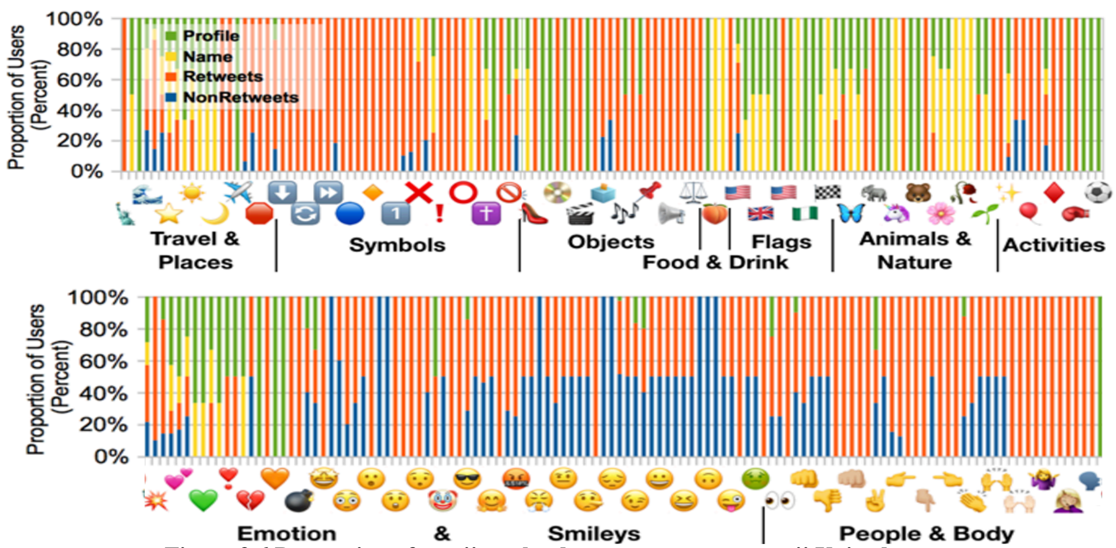


Figure 3.6 Proportion of emoji use by document type per emoji Unicode group.

The proportion of emoji use between the four document types of user profile descriptions, user names, retweets, and non-retweets were compared across the 95 Unicode emoji subcategories. I consider emojis within the same subcategory to be semantically similar. Figure 3.7 shows the proportion of emoji use per document type for the 63 subcategories that were used by at least one percent of users. The figure is organized in decreasing order of use going counter-clockwise starting at the 12 o'clock position. The subcategory name is labeled with text and shown with the emoji from that subcategory which was used by the greatest number of users. Proportion of emoji use is indicated by color for each document type with profile descriptions (green), user names (yellow), retweets (orange), and non-retweets (blue).

For example, the most-used emoji subcategory in this dataset was emotion and the most-used emoji in that subcategory was the red heart emoji "❤️". For this subcategory, 38% of use of emojis in this subcategory were retweets, 34% user profile descriptions, 15% non-retweets, and 13% were user names. Other subcategories of note included the second most popular subcategory country-flag with the U.S. flag emoji "🇺🇸" used the most, which is expected as this data was collected on keywords related to the 2018 U.S. midterm elections. The fourth most popular subcategory, was sky & weather with the water wave emoji "🌊" having greatest use which was associated with the Democratic party campaign slogan, blue wave. The mail subcategory with the ballot-box with ballot emoji "🗳️", associated with voting, was also in the top subcategories. The proportion of use for these subcategories was nearly equally distributed compared to other subcategories that showed higher proportion of use for one or two document types.

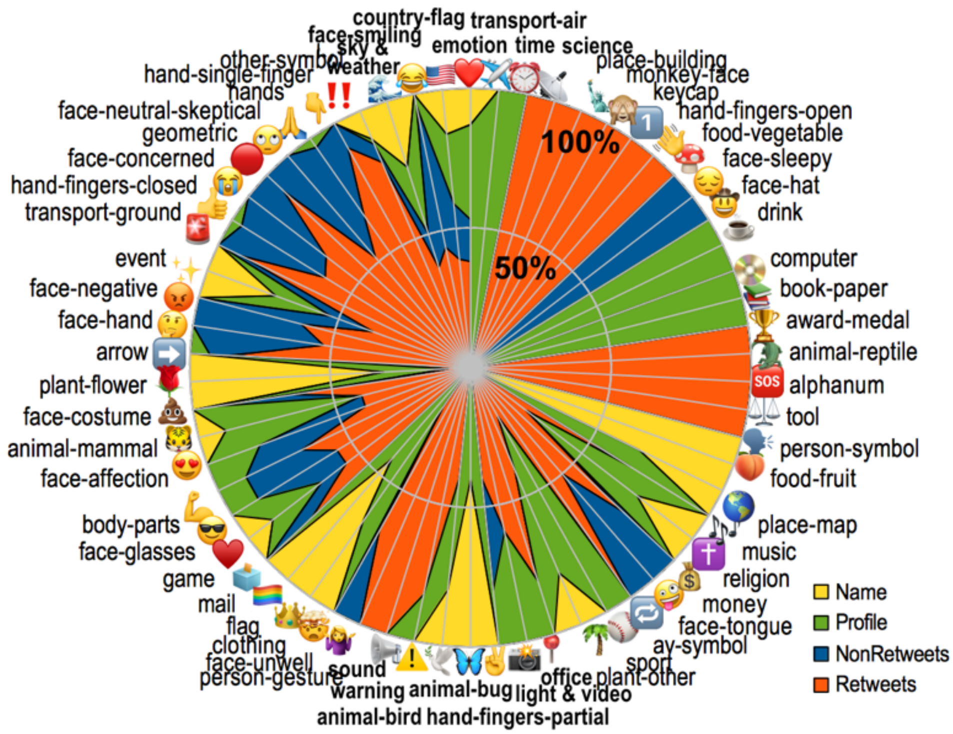


Figure 3.7 Proportion of emoji use by subcategory.

Many subcategories had a higher proportion of use associated with one or two document types. For example, retweets accounted for nearly 100% of use of emojis in subcategories: geometric, alpha-num, and keycaps. Non-retweets were associated with several face-related subcategories: face-neutral-skeptical, face-sleepy, and face-hat. The subcategories most associated with user profile descriptions were related to themes associated with hobbies, interests, and activities: sports, drink, book-paper, and transport-air. Meanwhile, subcategories most associated with emoji use in user names included:

plant-flower, food-fruit, and animal-bug. Next I measured similarity of the top emojis and subcategories per document type.

3.4.5 Emoji and Subcategory Similarity

I measured similarity of the top emojis and top subcategories associated with user names, profile descriptions, retweets, and non-retweets using the Jaccard similarity coefficient as described in Section 3.3. As nearly all possible emojis were used at least once in each document type, comparing all emojis and all subcategories would yield a Jaccard similarity coefficient near to 1, indicating identical use of emojis with respect to document types. However, as shown in the results from comparison of emoji categories and subcategories, proportion of emoji use across document types is not identical. Therefore, I chose the number of emojis and subcategories for comparison to minimize the number of unique values to be compared, maximize variation, and represent over 50% of users.

For this analysis, based on the above rationale, I compared the top 250 most-used emojis for each document type, which corresponds to 77%, 78%, 62%, and 65% of all users who used these emojis in non-retweets, retweets, user names, and user profiles respectively. For reference, the top 15 emojis for each are shown in Table 3.3. The Jaccard similarity coefficients between the top 250 emojis for each document type (Table 3.4) indicates retweets and non-retweets have similarity in individual emoji use, while profile descriptions had moderate similarity with names, and top emojis in user names had low similarity with top emojis in retweets and non-retweets.

Turning to subcategories, I compared the most-used 35 subcategories which accounted for 58%, 63%, 72%, and 64% of users of emojis in non-retweets, retweets, user names, and profile descriptions, respectively. Table 3.5 shows the top 15 emoji subcategories per document type. The Jaccard similarity of top 35 subcategories between each document type is shown in Table 3.6. The lower Jaccard similarity coefficients for top subcategories compared to top emojis indicates that the document types are more distinct semantically when considering emoji use by subcategories, even though there may be overlap of a few individual emojis.

Table 3.3 Top emoji for each communication type

	Top emoji with greatest number of users
non-retweet	😂,🇺🇸,🤣,❤️,🤔,🙄,👍,🌊,😭,💙,😉,😊,😡,🗳️,🙏
retweet	🇺🇸,😂,🚨,👇,❤️,🔥,🔴,💥,‼️,🤣,🗳️,🤔,👉,✅,🙄
name	🇺🇸,✨,🌊,🏳️‍🌈,🌻,🌹,🦋,❤️,⭐,❌,💜,👑,🌸,💙,🌈
profile	🇺🇸,❤️,✨,💙,🏳️‍🌈,💜,💕,🌊,💛,🌈,📚,🌎,✈️,⚽,💖

Table 3.4 Jaccard similarity for the top 250 emojis

	Non-retweets	*Retweets*	*Names*	*Profiles*
Non-retweets	1	.71	.3	.46
Retweets	.71	1	.27	.4
Names	.3	.27	1	.56
Profiles	.46	.4	.56	1

Table 3.5 Top emoji subcategories per communication type

	Top Subcategories Ranked by Use
Non-retweets	face-smiling, country-flag, emotion, face-concerned, face-neutral-skeptical, hand-fingers-closed, hands, sky & weather, face-hand, person-gesture, face-negative, hand-single-finger, other-symbol, face-affection, hand-fingers-partial
Retweets	country-flag, face-smiling, hand-single-finger, emotion, other-symbol, transport-ground, sky & weather, geometric, hands, face-concerned, hand-fingers-closed, face-neutral-skeptical, arrow, face-negative, face-hand
Names	country-flag, emotion, sky & weather, plant-flower, event, other-symbol, animal-mammal, animal-bug, flag, plant-other, clothing, food-fruit, game, animal-bird, hand-fingers-partial
Profiles	emotion, country-flag, sky & weather, animal-mammal, event, plant-flower, flag, clothing, hand-fingers-partial, zodiac, sport, face-smiling, plant-other, other-symbol, game

Table 3.6 Jaccard similarity for top 35 subcategories

	Non-retweets	*Retweets*	*Names*	*Profiles*
Non-retweets	1	.58	.21	.3
Retweets	.58	1	.23	.3
Names	.21	.23	1	.55
Profiles	.3	.3	.55	1

3.5 Conclusion

Analysis of the role of emoji use in online communication is still a growing area of research. This chapter contributes to this research by presenting a baseline of analysis and methodology to enable summarization and comparison of emoji use by aggregating emojis based on Unicode emoji categories and subcategories per user and document. By considering this semantic grouping of emojis, I move the research on emojis beyond just comparing individual emojis and broad aggregations. In applying this methodology to a set of 44 million tweets and over 3 million user profiles relating to the 2018 U.S. midterm elections, I find that differences in emoji use emerged based on document type (i.e., user names, profile descriptions, retweets, and non-retweets) and for emoji super users. In addition, this analysis shows that while individual authors can choose from over 3000 emojis, users consistently choose a few unique emojis from one or two subcategories while emoji super users, in contrast, use a greater variety from several subcategories. Further, comparing emoji use across users reveals a collective preference of emojis from select emoji categories and subcategories for specific document types. For example, retweets had higher proportion of symbol emojis while non-retweets had a greater proportion of face-gesture emojis.

However, this work is not without its limitations. One such limitation is that analysis of emoji categories and subcategories adds additional dimensions of complexity compared to just examining the most frequently used emojis. Another limitation is: this work only looks at one use case and thus a question is how representative are these findings? To answer this question, more case-studies are needed to be carried out along with exploring how emojis are used on other social networking platforms using the methodology presented in this chapter.

While these limitations exist, this work shows that emojis are more than just text and the methodology in this chapter supports semantic content analysis of documents containing emojis. this approach of emoji groupings by categories and subcategories provides a descriptive summary which enables comparison of emoji use in a way that has not been done before. As such, this work offers a new lens to study and compare forms of self-expression across a variety of digital media content types. Further, the analysis of individual and collective emoji use can enrich this understanding of the methods and styles of digital communication in online social networks.

4 DIVERSITY FROM EMOJIS AND KEYWORDS IN SOCIAL MEDIA[2]

Social media is a popular source for political communication and user engagement around social and political issues. While the diversity of the population participating in social and political events in person are often considered for social science research, measuring the diversity representation within online communities is not a common part of social media analysis. This chapter attempts to fill that gap and presents a methodology for labeling and analyzing diversity in a social media sample based on emojis and keywords associated with gender, skin tone, sexual orientation, religion, and political ideology. I analyze the trends of diversity related themes and the diversity of users engaging in the online political community during the leadup to the 2018 U.S. midterm elections. These results reveal patterns along diversity themes that otherwise would have been lost in the volume of content. Further, the diversity composition of this sample of online users rallying around political campaigns was similar to those measured in exit polls on election day. The diversity language model and methodology for diversity analysis presented in this chapter can be adapted to other languages and applied to other research domains to provide social media researchers a valuable lens to identify the diversity of voices and topics of interest for the less-represented populations participating in an online social community.

---

[2] This chapter is based on: Swartz, M., Crooks, A., and Kennedy, W. G.. (2020). Diversity from Emojis and Keywords in Social Media. In International Conference on Social Media and Society (SMSociety '20), July 22–24, 2020, Toronto, ON, Canada. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3400806.3400818

4.1 Introduction

Social media studies provide insights on themes contained within social media content and user interactions across a variety of topics including, for example, natural disasters (Crooks et al., 2013), vaccinations (Yuan & Crooks, 2018), and politics (Soares et al., 2019). While social media analysis has been used to study a variety of social and political issues, there has been less attention given to measuring the diversity represented by the users and content within the social media sample for these various studies. Applying a diversity lens to social media analysis enables researchers to better understand the diversity representation of the population being studied as well as to identify diversity related themes within the social media content. This is particularly important with respect to social media and politics. In an era when news and political leaders are using social media to deliver their messages (Barberá & Zeitzoff, 2018; Shearer & Gottfried, 2017) and political groups use social media to rally support or engagement (Stier et al., 2018; Vergeer et al., 2013), it has never been more important to ensure that the diverse population of a nation is being reached and the voices of less represented populations in online social-political communities are not lost in the noise (Hargittai, 2020; Hodson & Petersen, 2019).

To understand the political landscape of a country, including the concerns of the population and composition of political parties, traditional research methods are popular because they are designed to be rigorous, targeted, statistically valid, and typically representative of the diverse populations interviewed or surveyed (Halperin & Heath, 2020; Preoţiuc-Pietro & Ungar, 2018). With social media now comprising a large part of political activity and campaigning (Bode & Dalrymple, 2016; Sterling et al., 2020; Vergeer et al., 2013), these formal survey methods may not adequately capture or account for the topics and concerns expressed in less formal styles and behaviors of communication (e.g., slang, emotion, sarcasm, gestures) in social media (Felbo et al., 2017; Gawne & McCulloch, 2019). Studying social media presents its own set of opportunities and challenges (Giorgi et al., 2019; Stieglitz et al., 2018). Many approaches that study the diversity and demographics of social media users rely on location-based information associated with where content is posted (Duggan & Brenner, 2013; Sloan et al., 2013). Often researchers will infer demographics and diversity attributes of users based on the location of the user's profile or content and compare it with other locational datasets for the same geographic area such as census or voter statistics aggregated at varying scales of geography (Barbera, 2016; Giorgi et al., 2019; Preoţiuc-Pietro & Ungar, 2018). Using location from social media content relies on the provider of the platform as well as the individual user settings. Accuracy and precision of this location information varies greatly and currently ranges from precise coordinates to broad geographic areas such as a city or country (Na'aman et al., 2017). However, as the availability of precise geolocation information varies substantially across platforms and is becoming less available due to privacy concerns (Culliford, 2019), alternative approaches are needed to explore diversity within social media communities and datasets.

To fill this gap, this chapter presents a novel method using a diversity language model to associate diversity related attributes to social media user accounts and content by analyzing the emojis and keywords used. I apply this model to publicly available tweets that contain keywords related to American politics, specifically the 2018 U.S. midterm. this methodology for diversity analysis is then applied to identify the groups of social media users and trends in content with similar diversity attributes. These results reveal patterns of social media engagement across political lines among the diverse populations that otherwise would not have been apparent if I only analyzed the content for key political terms without taking diversity of the users into account.

The three main contributions of this chapter are: (1) the development of a diversity language model based on the use of emojis and keywords, (2) the development of a methodology for diversity analysis to label and analyze diversity attributes within a social media sample, and (3) applying the methodology to analyze the diversity of the online community using political party campaign slogans associated with American politics during the lead-up to the 2018 U.S. midterm elections. The remainder of this chapter presents a review of related research with respect to diversity and the use of emojis in social media. This is followed by a description of the datasets collected for this research and the methodology to develop a diversity language model and conduct diversity analysis. Then I present and discuss the results of this analysis of the 2018 U.S. midterm elections and conclude the chapter with areas for further research.

4.2 Background

The diversity of political party membership and engagement are often measured and analyzed by methods such as interviews, surveys, voter registration, exit polling, and attendance of (e.g., (Doherty et al., 2018; Schuldt & Pearson, 2016; Tyson, 2018)). However, while much political activity takes place in an online setting, such as social media (Bode & Dalrymple, 2016; Vergeer et al., 2013), there has been limited research or methods developed to measure and compare the diversity of online political userbase and engagement. While there are challenges with working with social media data with regards to studying politics (Deb et al., 2019; Hodson & Petersen, 2019), it is important to try to observe the diversity of a social media sample (Giorgi et al., 2019; Preoţiuc-Pietro & Ungar, 2018). As bots and trolls try to influence social and political outcomes with the creation of accounts and the use of language and characteristics similar to the groups they are trying to influence (Badawy et al., 2019; Kosmajac & Keselj, 2019), it can be challenging to accurately measure the true identity and diversity of social media user presence (Deb et al., 2019). Identifying diversity attributes associated with these accounts and content may reveal which groups are being targeted and in what way.

In addition, viral content, retweets, influencers, and organized campaigns may drown out unique and relevant content of less represented populations (Hargittai, 2020; Hughes & Asheer, 2019). Analyzing the diversity of a social media sample of users can potentially provide cues about the groups engaging with content and enable these voices to be heard even within the massive volumes of social media content. This analysis can also serve as a baseline measurement of diversity associated with topics and users in social media which can then be compared over time to identify behaviors and accounts associated with bots, trolls, or proliferators of fake or viral content (Kosmajac & Keselj, 2019).

For this research I compare the diversity of the user base sharing content containing political campaign slogans and election-related themes pertaining to the U.S.

Democratic and Republican party activities. The goal of this research is to characterize the diversity of the population of users that utilized election related or political slogans or phrases in their social media content, specifically related to the U.S. 2018 midterm elections. While the diversity and demographics of the political base at rallies and events is measured directly via more formal approaches (e.g.(Doherty et al., 2018)), here I attempt to measure the diversity of the user base engaging in online political communities solely with digital online social media data.

There are many concerns with studying social and political issues using social media data. A common one is bias arising from the differences in social media users compared to the real-life population (Blank & Lutz, 2017; Hargittai, 2020). While identifying the demographic makeup of the user base of a social media platform is already an area of research (Kalimeri et al., 2019; Mislove et al., 2011; Sloan et al., 2013), it should also be noted that the users represented in a social media sample may not even reflect the composition of the user base of that same social media platform (Chakraborty et al., 2017). Further, there are differences in the styles, language, ways that people engage, and even how individuals represent themselves in real-life compared to social media (Sterling et al., 2020). Diversity analysis can help to identify these differences and also quantify the bias represented in the sample along diversity related characteristics and themes.

Computer mediated communication styles on social media are a unique set of linguist patterns that users have evolved to adapt to the social trends and norms of an online community using a social media application. Often these adaptations arise based on the availability of the platform's features as well as to work within limitations such as the amount or type of content that can be posted (e.g., (Bode & Dalrymple, 2016; Herring & Dainas, 2018)). For example, stickers, website url shorteners, and emojis are popular within social media and text messaging applications (Gawne & McCulloch, 2019). Thus, the digital linguistic styles and patterns of users should also be considered during social media analysis. In this research, I focus on the use of emojis and keywords as part of this social media analysis.

There are three common approaches to conducting social media analysis with content containing emojis. Many studies will remove or sometimes replace emojis with words, (e.g., (Yuan & Crooks, 2018)). Another popular approach is content analysis that assigns a score based on the presence of specific emojis as indicators of emotion, sentiment, or sarcasm (Felbo et al., 2017). However, these approaches typically only analyze the few emojis which are anthropomorphic, such as face- and body-gestures and do not consider the thousands of other emojis that may exist, such as emojis depicting symbols, animals, food, and objects. The other approach is semantic analysis to assign meaning to emojis (Barbieri et al., 2016; Tigwell & Flatla, 2016). Although the Unicode Consortium provides a standard for emoji codepoints and names, the definition and digital character representation of emojis are only suggestions (Davis, 2018). As a result, emoji presentation, specifically the color, shape, or details of an emoji, may vary across devices and even social media platforms. These differences can impact how emojis are used and perceived (Gawne & McCulloch, 2019; Herring & Dainas, 2018; Tigwell & Flatla, 2016). The interpretation of emojis within social media content will also be impacted by the socio-cultural context of the users (Barbieri et al., 2016), nearby text (Donato & Paggio, 2017; H. Miller et al., 2017), or even the type of document such as a username, tweet, or user profile (Swartz & Crooks, 2020).

I chose to examine the use of emojis as cues for diversity, inspired by the handful of studies which show differences in the most common emojis used per various user populations. For example, trends in emoji use were described based on culture and geography at the country level (e.g., (Ljubešić & Fišer, 2016)). Others have described emoji use based on age (Na'aman et al., 2017) or differences based on gender (Chen et al., 2018; Herring & Dainas, 2018). And since 2016, with the availability of skin-toned emojis, some researchers have shown that people typically use emojis with a similar appearance as their own, such as skin-tone (Barbieri & Camacho-Collados, 2018; Robertson et al., 2018). However, these studies all describe the aggregated emoji use across populations when the diversity attributes are already known. Nonetheless, the current state of emoji related research demonstrates the utility of including emojis in social media analysis and provides a useful starting point for examining diversity of users based on keywords and emoji use.

4.3 Data Collection

For this research, I conducted social media analysis on publicly available tweets and user profiles collected from Twitter. I describe the datasets in this section.

4.3.1 Tweets

The analysis of this chapter is based on a dataset of over 44 million publicly available tweets posted during the timeframe of October 1, 2018 to November 7, 2018, to coincide with the timing of the one month prior to and including the day of the 2018 U.S. midterm elections. The tweets were collected using Twitter's free streaming application programming interface (API), which provides only a sampling of all available tweets. The keywords and account names used to collect tweets related to the 2018 U.S. midterm elections, campaign slogans, and specific user accounts associated with the Democratic and Republican political parties and candidates, Table 4.1.

Table 4.1 Subset of keywords for 2018 U.S. midterm election

Election related	***Campaign slogans***	***Accounts***
election2018	BlueWave	@TedCruz
midterms2018	FlipTheSenate	@BetoORourke
democrat	FlipTheHouse	@FLGovScott
DNC	VoteThemOut	@SenateDems
republican	MAGA	@HouseDemocrats
RNC	KAG	@SenateGOP
GOP	TakeItBack	@HouseGOP

Twitter returns the tweet content and metadata about the tweet. One attribute in the metadata indicates whether the tweet is a retweet. For this analysis I divided the tweet collection into retweets (RT) and non-retweets (Non-RT). Retweets accounted for 84% of all the tweets I collected. Of the over 3 million unique authors in this collection, 61% only sent retweets, 19% never sent retweets, and 20% sent both retweets and non-retweets.

4.3.2 User Profiles

For the 3 million authors of tweets in this collection, I used the free Twitter search API to collect the user profiles. The user profile information returned from Twitter includes an attribute for the user profile description, a free-form narrative that users can fill in if they wish. Most of the user profiles descriptions I collected contained information and only a handful were blank. While the amount of detail and length varied greatly, the type of information in the user profiles included hobbies, interests, political beliefs, religious beliefs, race, language, national or cultural identity, relationship status, educational status, gender, sexual orientation, employment status, history, and more. Although it is difficult to validate the authenticity of the information contained in user profile descriptions, this volunteered narrative provides a wealth of user information. Next I describe how I assign diversity categories and subcategories based on diversity-related keywords and emojis contained in user profiles and tweets.

4.4 Methodology

In this section I present this methodology for conducting diversity analysis of social media. I discuss the creation of a diversity language model to associate emojis and keywords with diversity-related characteristics. The model is used to label social media content with diversity categories and subcategories. The diversity-labeled content is then analyzed to identify themes and patterns of behavior across users with similar diversity characteristics. Figure 4.1 provides a graphical summary of this diversity analysis workflow.

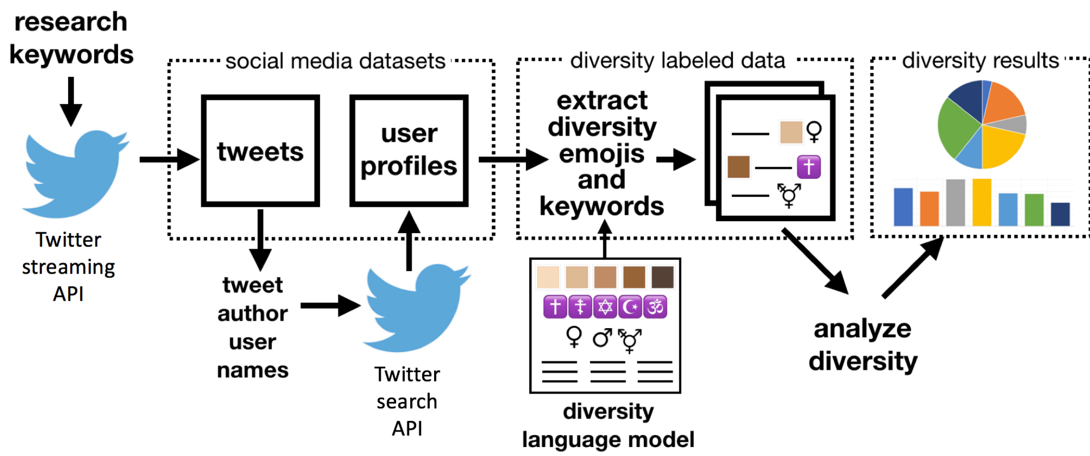


Figure 4.1 Workflow for diversity analysis of social media content.

4.4.1 Diversity Language Model

I developed a diversity language model based on specific keywords and emojis that may be associated with the following five diversity categories: gender, religion, race/skin tone, sexual orientation, and political ideology. Categories were divided into subcategories, e.g., the gender category contains the following subcategories: female, male, and transgender. Figure 4.2 lists the diversity categories and the main subcategories of this diversity language model. It also includes the emojis and a few of the Spanish and English keywords associated with a particular category and subcategory.

While the diversity language model displayed is not the exhaustive list of keywords I used, it is representative of how the final model was constructed. I acknowledge these are not fully representative lists and may not be endorsed by everyone; however, the keywords and emojis were used by more than 100 users in this collection in a way as to indicate diversity characteristics. I developed the model through an iterative process and refined the keywords and emojis included with the goal of having a robust model that describes the diversity characteristics of many users while also minimizing the number of terms that could lead to misclassifications. I also acknowledge that the presence of any of these terms or emojis from the diversity model that appear within a tweet or user profile does not always mean the user is self-identifying, and that the use of the term could be about someone else, discussion of a diversity topic, or in some cases may not be related to diversity at all.

Category	**Subcategories**	**Emojis**	**Keywords examples (English/Spanish)**
Skintone	Light Light-medium Medium Medium-dark Dark	🏻 🏼 🏽 🏾 🏿	white, blanco moreno black, negro, negra
Gender	Female Male Transgender	♀ ♂ ⚧	woman, she/her, man, he/him transgender
Sexual Orientation	Indicated	🏳️‍🌈	gay, lesbian, bisexual, straight, hetero, polyamor
Religion	Christian Jewish Islamic Hindu Atheist	✝️,☦️ ✡️,🔯 ☪️ 🕉️,☸️	Catholic, Catolic, Jewish, Judaism Muslim, Islam Hindu, Sikh atheist, ateo
Political	Conservative Liberal		Republican Democrat

Figure 4.2 Diversity Language Model.

The subcategories used in this model in some cases were limited based on the emojis I selected as part of the diversity language model. For the skin tone category, the five subcategories are based on the five Fitzpatrick skin tone emoji modifiers. I chose these emoji modifiers for use as diversity cues because a user must specifically select an emoji of a person or body-part that contains a skin tone because most default emoji presentations do not have skin tone modifiers. Users typically prefer to use a skin tone emoji with a similar appearance as their own (Barbieri & Camacho-Collados, 2018; Robertson et al., 2018). For gender, I chose the subcategories of female, male, and transgender based on the availability of corresponding emoji symbols and modifiers. For religion, I created subcategories for Christian, Jewish, Orthodox, Islam, and Hindu, based on emoji symbol availability. However, I also added the category for Atheist due to the large number of profiles using this term despite the lack of an emoji. For sexual orientation, I decided to keep this as a general category because of the lack of emoji differentiation for subcategories. And finally, I added political ideology as the fifth diversity category and chose only keywords based on their presence and context of use within user profiles in this collection.

As I tuned the diversity language model based on analysis of keywords and emojis in user profiles more heavily than tweets, there were a few cases where diversity associated terms used in the profile to indicate a diversity attribute occasionally had different meanings in tweets. For example, the keyword “white” for the skin tone subcategory for light worked well for labeling user profiles, but for tweets resulted in content containing a phrase such as “white house” to be wrongly labeled as a diversity attribute for skin tone. I mitigated this by only using the skin tone emojis during this analysis of tweets for the skin tone diversity category.

4.4.2 Assign Diversity Label

Once the diversity language model, Figure 4.2, is created and verified, I then use the model to assign diversity category and subcategory labels to tweets and user profiles that contained one of the terms or emojis in the model. In some cases, a single user profile or tweet may be assigned multiple subcategories within the same category. When this happened, which was rare, I would associate the content with the diversity subcategory as “mixed”.

4.4.3 Diversity Analysis

Using the social media content labeled with diversity categories and subcategories, I then review the composition of the dataset collected and identify trends based on the diversity characteristics. As part of the initial exploratory analysis I measure the prevalence of diversity characteristics as the percent of user’s profiles and tweets containing specific diversity emojis and keywords. The composition of the datasets is then summarized as the proportion or percent of user accounts, retweets, and non-retweets for each of the diversity categories and subcategories. To understand differences in the way diversity emojis and terms are used, I also compared if a user includes the same diversity emojis and keywords in both their profile and tweets.

After reviewing the overall summary metrics of diversity for the collections, I analyze specific trends per set of users or tweets with the same diversity characteristics. I also examine temporal trends in tweet and author volume for each diversity characteristic. With this baseline understanding of the datasets along the diversity measures, I then focus on the patterns from combining multiple categories, e.g., political ideology and religion.

4.5 Results and Discussion

As the aim of this chapter is to explore diversity characteristics in a social media sample based on use of keywords and emojis in tweets and user profiles, building upon this methodology, I now turn to the results of this diversity analysis of social media content related to the 2018 U.S. midterm elections and discuss these findings. In what follows, I first gauge how prevalent the use of diversity emojis and keywords are within tweets and user profiles. I then establish a baseline by analyzing the patterns and trends within the dataset collections associated with diversity characteristics. Finally, I summarize the diversity of the user base in relation to election-related themes and campaign slogans.

4.5.1 Diversity in Tweets and Profiles

To understand how prevalent diversity characteristics are in these datasets, I measured what percent of tweets and user profiles contained diversity emojis or keywords. I also compared if these proportions were different between retweets and non-retweets. I present a summary of the results and key findings below.

Presence of diversity-related emojis and keywords in tweets. Across the corpus of 44 million tweets, approximately 15% (6.6 million tweets) contained either a diversity keyword or emoji, which I refer to as diversity tweets. Diversity tweets were sent by 36% of the authors in this collection. Of diversity tweets, 95% contained at least one of the diversity keywords and 5% contained at least one diversity-associated emoji, regardless of retweet or non-retweet status. I then examined the value of including diversity emojis as part of the language model by comparing the overlap of tweets with diversity emojis and diversity keywords. Of diversity tweets containing diversity emojis, only 15% also had diversity keywords. In terms of emoji use in general, only 9% of all tweets containing emojis had a diversity-related emoji for skin-tone, gender, or religion.

Presence of diversity-related emojis or keywords in user profiles. Next I examined the prevalence of diversity characteristics within user profiles. In the full set of 3.3 million user profiles in this collection, 15% of them (approximately 500,000) contained either a diversity keyword or diversity emoji, which I refer to as the diversity profiles. For authors sending both retweets and non-retweets, 22% had diversity characteristics in their profile. Authors sending only retweets had 18% and authors of only non-retweets had 12% of profiles containing diversity keywords or diversity emojis. For comparison, 19% of authors of diversity tweets had a diversity profile. In regards to the use of both diversity keywords and diversity emojis in the same profile, this was much more likely with authors that sent both retweets and non-retweets (44%) compared to authors of only retweets (27%) and authors of only non-retweets (25%).

Key findings for presence of diversity characteristics in tweets and profiles. While only 15% of tweets and user profiles in this collection contained diversity characteristics, over a third of the authors sent at least one tweet containing a diversity emoji or keyword. Users that included diversity characteristics in their profile were often as a way to self-identify diversity characteristics. Authors with diversity profiles were not any more likely to use diversity characteristics in their tweets, and vice versa. This indicates that both user profiles and tweets should be considered when taking diversity into account as they each may reveal different insights about the diversity of the social media sample. The use of diversity emojis and keywords within the same document occurred for a small percentage of tweets and user profiles. This indicates emojis provide diversity cues not fully covered by keywords alone, thus, both are valuable to include in the diversity language model.

4.5.2 Analysis of Diversity

Next I established a baseline of patterns and trends of diversity characteristics in this collection by analyzing tweets and profiles along each diversity category and subcategory. I summarize noteworthy results below.

Proportion of profiles and tweets by diversity category. I compared the proportion of diversity keywords and diversity emojis used in profiles and tweets across the diversity categories, Figure 4.3. This analysis reveals that user profiles were slightly more likely than tweets to contain diversity cues associated with skin tone, gender, and sexual orientation. However, tweets were more likely to contain cues associated with religion. The use of skin tone emojis was more prevalent in user profiles than tweets.

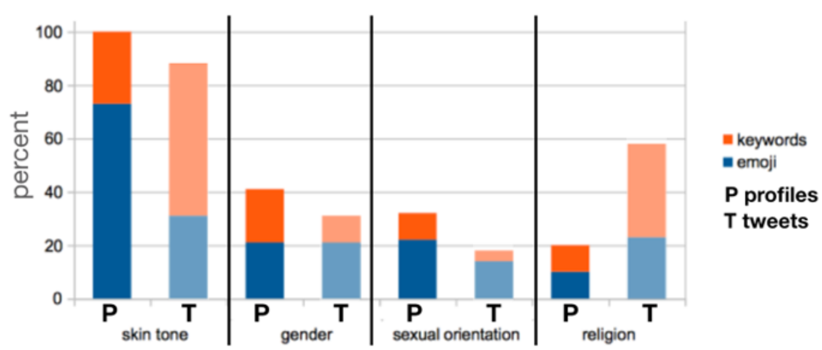


Figure 4.3 Percent of diversity profiles and tweets with diversity emojis and keywords.

Diversity composition of user profiles per diversity category. I then analyzed the diversity composition of the collection of user profiles by measuring the percent of profiles labeled with a subcategory for each diversity category, Figure 4.4. Although this analysis only focuses on users with diversity profiles, it does provide new insight about the diversity characteristics represented within this collection of user profile descriptions. Of these diversity profiles in this collection, there was a greater percent with characteristics associated with: medium or medium-light skin tone emojis, female gender, and Christian religion. In addition, profiles that contained the rainbow flag also predominantly included keywords for sexual orientation.

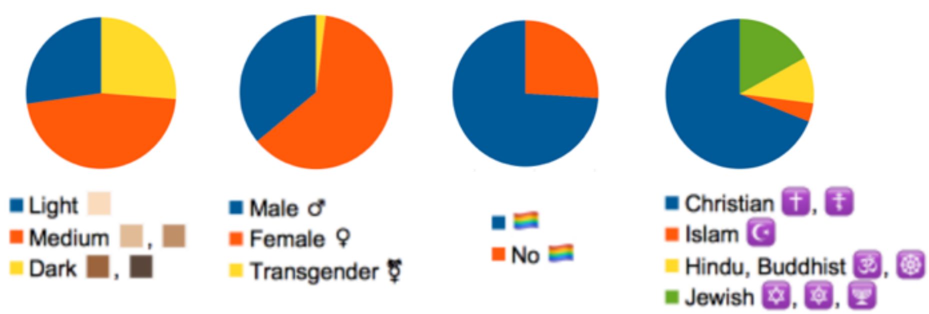


Figure 4.4 Proportion of user profiles by diversity subcategory.

Temporal analysis of diversity tweets. Analyzing the volume of tweets by diversity category per day provides a baseline to observe changes over time and identify peaks and valleys of activity. There were two spikes of activity in the collection of tweets containing emojis with skin tone, one in mid-October and the other the day before the elections. Plotting the volume of non-retweets and retweets separately reveals different patterns of activity volume, Figure 4.5. Analysis of the retweets and non-tweets with skin tone shows that the day before the elections came from several non-retweets encouraging voting. The spike mid-October was retweet of a single tweet by a celebrity that included a political message, included an emoji with medium skin tone, and asked to be retweeted.

Content analysis of tweets based on diversity in author profiles. Themes and topics were identified in tweets by authors whose profiles were labeled by diversity subcategory. Tweet text was processed using natural language processing methods to remove punctuation and common words (e.g., the, and, as). The remaining words were stemmed for aggregation, e.g., voting, voter, votes, and vote were stemmed to the term vote. For each tweet, unique terms and hashtags were identified in retweets and non-retweets per diversity subcategory. The results did not yield distinct differences based on diversity due to the large amount of terms and keywords I used to collect the data. However, I identified that two hashtags in particular “#maga” and “#bluewave” were among the top used hashtags in tweets across users of all diversity subcategories. These hashtags were associated with political campaigns for the Republican and Democratic political parties and I analyze the diversity representation of these users next.

Key findings from analysis of diversity characteristics. Diversity analysis enabled baseline of diversity compositions of the collection of tweets and user profiles. With this baseline, top hashtags and content trending was associated with specific diversity subcategory and otherwise would not have been easily found in the volume of tweets. This insight into the composition and behavior of engagement is especially useful to understand the approaches and dynamics of online political campaigns.

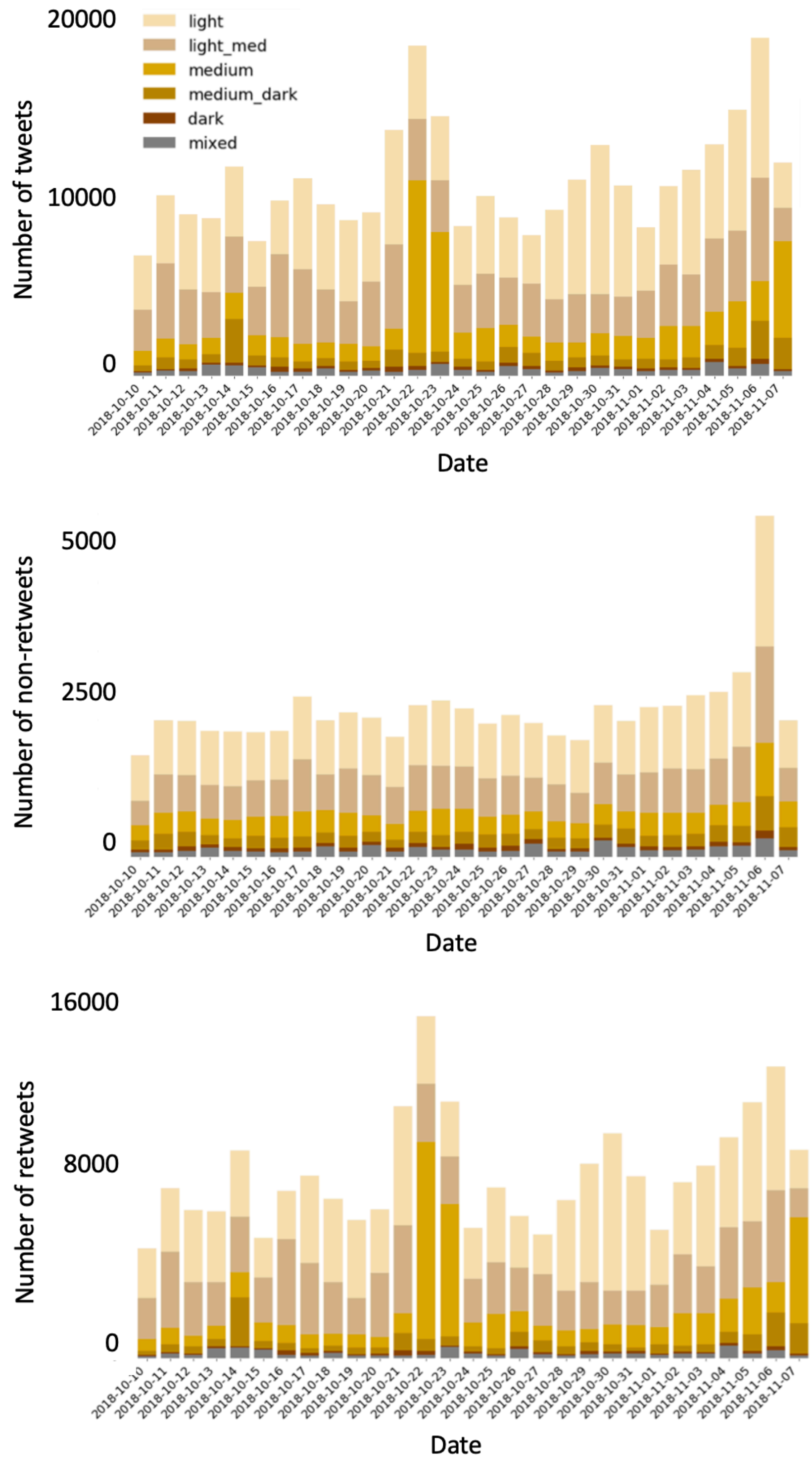


Figure 4.5 Volume of tweets, non-retweets, and retweets with skin tone emoji by date.

4.5.3 Diversity and Political Ideology

Most tweets in this collection contained terms related to political ideology, however their use often did not reflect political party affiliation of the user and it was common to see keywords for both political parties included in the same tweet. For user profiles, only 5 percent contained political ideology keywords. From earlier content analysis I did find that tweet authors predominantly used campaign slogans across their tweets for only one of the political parties and similarly within user profiles. So rather than using political ideology, I analyzed the diversity represented in user profiles for authors of tweets that contained election related phrases and political party campaign slogans associated with the U.S. 2018 midterm elections. One of the top political campaigns for the Democratic party was the “Bluewave”, which included terms such as “Blue Wave”, “#bluewave”, and the wave emoji. There were 22,051 users in this collection that had authored tweets including terms associated with Bluewave and also had diversity keywords or emoji in their user profile description. For the Republican party, the “MAGA” slogan, which stands for “Make America Great Again”, and the hashtag “#maga” were used by 65,695 users who had diversity profiles. I compared the diversity composition of users based on their use of these political campaign slogans. I found users of Bluewave campaign phrases had a greater proportion of user profiles attributed with female gender; skin tones for medium-light, medium, and medium-dark; and atheist and Jewish for religion. In contrast, MAGA users had a higher percent of diversity profiles containing keywords and emojis representing the diversity subcategories of male, light skin tone, and Christian. Figure 4.6 shows the proportion of users in this collection per diversity category associated with MAGA and Blue wave.

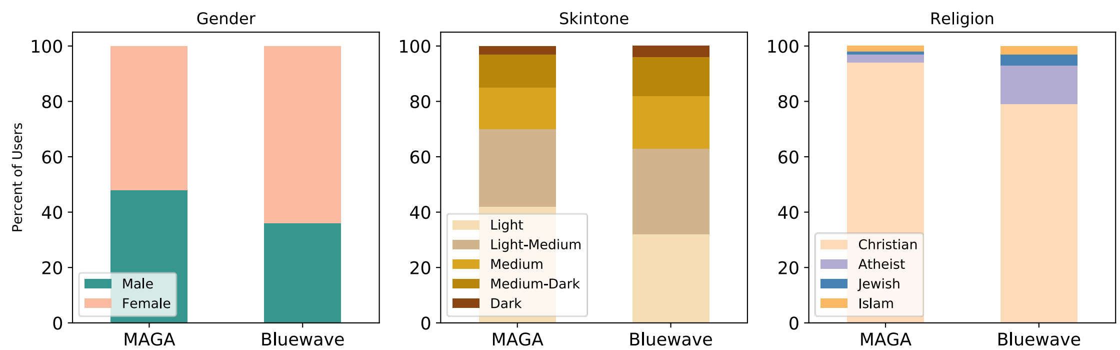


Figure 4.6 Composition of users in this collection for two political campaigns.

Key findings from diversity analysis of political ideology and campaigns. While I was not able to easily divide users based on political ideology, I was able to differentiate users and content based on use of political campaign slogans in tweets and profiles. I found users did not typically self-identify political party affiliation in their user profile or in tweets. The use of political terms such as party names in tweets were more often used in banter or occasionally in discussion of social issues and policies associated with a political party. Alternatively, I was able to conduct diversity analysis along use of political campaign slogans. Campaign slogans were used across the diversity categories and were polarizing in that users typically only used campaign slogans for one political party. In addition, the diversity analysis of political campaigns yielded proportions across diversity categories similar to those reported in exit polls on election day for the 2018 U.S. midterm elections (Tyson, 2018). This indicates that the diversity analysis of social media users expressing support for a political party online through the use of campaign slogans may indicate diversity composition associated with voting outcomes. Further, diversity analysis provided additional insight about the content and user base represented in this social media sample and is a valuable addition to social media analysis.

4.6 Conclusions

Diversity analysis of social media data, as presented in this chapter, provides an additional lens for studying diversity related themes in political discussion and the diversity of users engaging in online social communities. This chapter presents a novel methodology for labeling and analyzing diversity represented in a social media sample based on keywords and emojis associated with gender, skin tone, sexual orientation, religion, and political ideology. In applying this methodology on a social media dataset collected during the leadup to the 2018 U.S. midterm elections, I established a baseline and identified trends of diversity related content that would not have been found in the volume of tweets otherwise. Furthermore, the diversity composition of users of political party campaign slogans yielded proportions along political party lines similar as those measured in exit polls on election day.

The results indicate that both social media content and user profiles reveal different insights related to diversity and both should be considered when conducting a diversity analysis. Specifically, I observed that users were more likely to self-identify using diversity keywords and emojis in the user profile description rather than in their tweets. This means that for deriving a diversity composition of users, analysis of user profiles is preferred over aggregating social media posts by user. I also found semantic differences in the way diversity related emojis and keywords were used, which indicates that emojis are a useful addition to the analysis of diversity in social media.

This research is not without its challenges. The diversity keywords and emojis do not represent all diversity attributes and at times may also take on additional meanings not related to diversity. Further, the number of social media posts and user profiles containing diversity related keywords and emojis will vary based on how the data are collected and social norms of the social media platform. In addition, when conducting social media analysis, there is always inherent bias when comparing the number and composition of users represented in a social media sample to a real-life population. This is further impacted by the difficulty in validating the authenticity of user accounts and veracity of content, especially with the prevalence of fake accounts, bots, trolls, and misinformation on social media platforms.

While this research focused on baseline and trends of diversity representation in social media, there are several opportunities for future work. this approach for labeling diversity attributes in social media data can be compared to other methods using machine learning or manual tagging. Detailed content analysis, such as topic modeling or sentiment analysis, could be used to connect diversity with issues such as hate speech. Diversity analysis can be used to identify the representation of accounts associated with bots, misinformation efforts, and political influence campaigns which may reveal insights about their intent and targeted audience. Social network analysis can be used to examine diversity networks of users with similar diversity characteristics to measure how connected diverse groups are on social media related to political issues. In addition, the diversity language model presented in this chapter can be further adapted for other languages or topics. It can be applied beyond just tweets to assess the extent to which diversity has been discussed as well as identifying diversity represented in online community engagement associated with other social and political topics.

5 BEYOND WORDS: COMPARING STRUCTURE, EMOJI USE, AND CONSISTENCY ACROSS SOCIAL MEDIA POSTS[3]

Social media content analysis often focuses on just the words used in documents or by users and often overlooks the structural components of document composition and linguistic style. I propose that document structure and emoji use are also important to consider as they are impacted by individual communication style preferences and social norms associated with user role and intent, topic domain, and dissemination platform. In this chapter I introduce and demonstrate a novel methodology to conduct structural content analysis and measure user consistency of document structures and emoji use. Document structure is represented as the order of content types and number of features per document and emoji use is characterized by the attributes, position, order, and repetition of emojis within a document. With these structures I identified user signatures of behavior, clustered users based on consistency of structures utilized, and identified users with similar document structures and emoji use such as those associated with bots, news organizations, and other user types. This research compliments existing text mining and behavior modeling approaches by offering a language agnostic methodology with lower dimensionality than topic modeling, and focuses on three features often overlooked: document structure, emoji use, and consistency of behavior.

[3] This chapter is based on: Swartz, M., Crooks, A., and Croitoru, A.. (2020). Beyond Words: Comparing Structure, Emoji Use, and Consistency Across Social Media Posts. In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRIMS ’20), Washington D.C., 10 pages. Springer.

5.1 Introduction

As social media users engage with online conversations and form virtual communities, social media analysis often focuses on the topics discussed (Wirth et al., 2019), user activity patterns (Rajabi et al., 2019), and networks arising from interactions of users (Schuchard et al., 2019; Yuan et al., 2019). Often overlooked is the communication style associated with a user's social media posts. For instance, analysis of the specific words used can provide a fingerprint of the individual posting the content (Pennebaker et al., 2003) and reveal shared linguistic styles of the online community (Khalid & Srinivasan, 2020). Users also adapt their language to address limitations and norms associated with technology (Walther, 2012) (e.g., character limits, availability of emojis). Within this chapter, I propose to move beyond just words, as the structural components of a document's composition and the way in which emojis are used within a document, such as a tweet, also provide cues about the individual and social norms for online communication styles and preferences.

In this chapter I introduce and demonstrate a language-agnostic methodology to characterize structures of content and emoji use within a document, measure consistency of structures across a set of documents, and cluster documents and users with similar patterns and behavior. By comparing these patterns and behaviors across users and user roles such as journalists, bots, and others, I can generate baselines and gain insights into the unique or shared structures of communication styles and emoji use.

Three main contributions of this research are: 1) a novel methodology for structural content analysis; 2) analysis of the structure of emoji use as the attributes, position, order, and repetition of emojis within a document; and 3) user behavior modeling with regards to consistency of structure of document and emoji use. Benefits of this approach include it is language-agnostic, requires less dimensions than traditional topic modeling, and yields additional measures that can be combined with other text and user metrics. Further, this chapter addresses a gap of current social media analysis by focusing on the structural components of communication style, enables comparison of emoji use, and models consistency of user behavior based on social media content. In what follows, Section 5.2 provides an overview of current approaches to content analysis and analysis of emoji use, followed by this methodology in Section 5.3. I then present and discuss these results in Section 5.4, and conclude with areas for further work in Section 5.5.

5.2 Background

5.2.1 Content Analysis of Social Media

Content analysis of social media has mainly focused on the words contained in posts to identify discussion topics or associate groups of users based on their use of specific terms, hashtags, or group of words identified via topic modeling (Wirth et al., 2019; Yuan et al., 2019b). Recently, content analysis combined with other metrics for user activity and network connections, has been applied in order to identify or categorize bots (Schuchard et al., 2019; Wirth et al., 2019). Analysis of the structure of social media is fairly nascent. Rajabi et al., (2019) considered number of words in text in addition to user activity metrics to identify user intent in spreading misinformation. In addition to content length, Zannettou et al., (2019) took into account content type such as presence of urls, and hashtags to describe activity associated with troll accounts. Comunello et al., (2015) examined the order of lexical properties of a tweet (such as place name, event date and time, event description) in order to improve the effectiveness of messaging during earthquakes. While Pederson, (2016) focused on the order of content within a tweet and impact on communication styles.

5.2.2 Analysis of Emoji Use

There is still much more to be learned about the way visual content, including emojis, are used in social media (Highfield & Leaver, 2016). However, most social media research pertaining to emojis has focused on the meaning of emojis or emojis as indicators of sentiment or sarcasm (e.g.,(Felbo et al., 2017)). Only recently has the emphasis shifted to the behavior and structure of emoji use. (Swartz & Crooks, 2020) revealed differences in the way emojis are used based on document types such as tweets, user names, and profile descriptions. (Varol et al., 2017) identified how emojis are used as structural markers based on where they are placed in text.

5.3 Measuring Document Structure, Emoji Use, and Consistency

In this section, I first describe how to represent structures of a document, Section 5.3.1. I characterize the structure of emoji use by describing the attributes of emojis used in Section 5.3.2 followed by the position, order, and repetition of emojis within a document in Section 5.3.3. Next, Section 5.3.4 explains how to measure consistency of structures across a set of documents. Finally, I describe how to cluster users based on structures in Section 5.3.5 and by consistency scores in Section 5.3.6.

5.3.1 Document Structure, Content Structure, and Emoji Spans

In order to define the structures of a document, first identify the types of content associated with documents in the collection. For the purpose of this chapter I use data from Twitter and view a single tweet as a document. For tweets, I identify content types: retweet indicator, text, emoji, punctuation, hashtag, mention, and url. Each document is divided into spans by content type, irrespective of spaces, and assigned a sequential number as span number. Span length is the number of features per span. Document structure is represented as a list with the content type and number of features for each span, in order of occurrence. Similarly, content structure is a list of span content types in order. Representing a document and content structure in this way enables comparison of documents based on the type or order of contents and enables grouping of documents with similar structural format and style.

For documents containing emojis, I identify which emojis are used as a sorted list of unique emojis. I use the emoji spans (i.e., the spans with content type of emoji) and document structure to describe the way that emojis are used in a document, which I refer to as the structure of emoji use. Figure 5.1 shows a sample tweet represented as document structure, content structure, emoji spans, and unique emojis. In the next two sub-sections, I demonstrate the structure of emoji use as the emoji attributes paired with the analysis of the position, order, and repetition of emojis within a document.

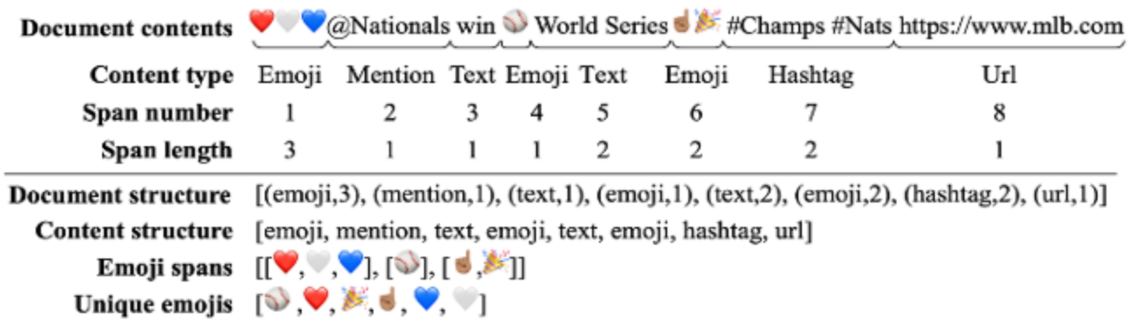


Figure 5.1 Structures of a sample tweet.

5.3.2 Attributes of Emojis in a Document

For each emoji in the emoji spans and in the unique emojis list I describe the emoji along eight attributes noted below. These eight attributes were chosen because each can be used alone or in combination to enable comparison of emojis. Additional attributes could be added such as sentiment or meaning. The first three attributes are from Unicode (Davis, 2020) and were chosen based on previous research showing the value of using emoji group and sub-group for comparison of emoji use (Swartz & Crooks, 2020). The other attributes are based on heuristics used to sort emojis.

1. **Unicode Group**: Unicode assigns each emoji to one broad category (e.g., Smileys & Emotion, Animals & Nature, Food & Drink, Travel & Places, Objects, Symbols, Flags, and People & Body which also includes Activity).
2. **Unicode Sub-Group**: Unicode assigns emoji to sub-category (e.g., face-smiling).
3. **Unicode Name**: The Unicode emoji name (e.g., “face with tears of joy”).

4. **Type**: A label assigned by mapping sub-group to another descriptive property based on a research topic. For this chapter, I use shape, anthropomorphic, and other.
5. **Anthro-type**: For anthropomorphic (human like) emojis, I map sub-groups to: face, face-gesture, hand-gesture, body-gesture, body-part, single person, multiple.
6. **Shape**: Indicated by emoji name: triangle, circle, square, star, heart.
7. **Color**: Indicated by emoji name: red, blue, yellow, pink, purple, orange, green, brown, white, black. I also include the five Fitzpatrick skin-tone colors used for emojis: light, light-medium, medium, medium-dark, and dark. I use the name because color appearance may vary across platforms.
8. **Direction**: Based on words in emoji name to indicate: up, down, left, or right.

To demonstrate how the above set of attributes enables descriptive comparison of similarities and differences of individual emojis, consider these two emojis, 🔺 and ☝🏽. There are differences in appearance and type with one a red triangle (type of shape) and the other a hand holding up an index finger (anthropomorphic hand-gesture). Yet they are similar in direction of pointing "up". Table 5.1 summarizes attributes for these emojis.

Table 5.1 Emoji attributes

Emoji	*Group*	*Sub-group*	*Name*	*Type*	*Anthro-type*	*Shape*	*Color*	*Direction*
🔺	Symbol	geometric	Up-pointing red triangle	Shape	None	Triangle	Red	Up
☝🏽	People & body	Hand-single-finger	Index pointing up:medium skin tone	Anthro	Hand-gesture	None	Medium	Up

5.3.3 Emoji Position, Order, and Repetition

Position of Emojis in a Document. I describe the general position of emojis in a document based on relative position of emoji as: first, beginning, middle, end, or last. I use document structure to derive relative position based on span number for the emoji content in relation to total number of spans in the document, divided into thirds, (e.g., the first third of spans is the beginning). Content in the very first and last spans are labeled as such. Documents with less than five spans are only first, middle, or last.

Emoji Order. The order of emojis and attributes are noted both within and across emoji spans. I take into account emoji order, as emoji color order within the same emoji span could result in a set of emojis taking on different meanings, based on context of text or user. For example, the set of heart emojis ❤️🤍💙 with color order red, white, blue could represent colors of a sports team (as in Figure 5.1) or a country flag (e.g., Netherlands or United States). The order of emojis or attributes can also indicate a pattern. For example, "💙🌟text🌟💙", represents a pattern I call emoji reversal which occurs when two consecutive emoji spans contain the same emojis or attributes, but the order in the second span is reversed.

Emoji Repetition. I categorize repetition of emojis or emoji attributes as three types: redundant, emphasis, and amplification. Redundant is the repetition of the same emoji or attribute within the same span, sometimes representing magnitude or quantity, (e.g., 😂😂). Emphasis is often used to draw attention and generally occurs when two emoji spans contain the same emojis regardless of order (e.g., "🚨Alert🚨", "🔵🌊Bluewave🌊🔵"). Amplification is repetition of an attribute across different emojis within the same span or across multiple spans (e.g., color red in: "🔺Vote🗳📕 all Red❗🔴").

5.3.4 Measure of Consistency

With the structures of document contents and emoji use represented, I can then measure consistency of these structures across a set of documents, such as a user's or group's tweets. This measure makes it is possible to highlight differences in behavior based on relative consistency in terms of document content, style, or emoji use.

To measure consistency, for a set of documents, iterate across the unique structures (U) (e.g., document structure, content structure, or emoji use). For each unique structure, divide the number of documents in the set with that structure (d_i) by the number of documents for the structure in the set with the most documents (max (Ud)), then square the results. Calculate the measure of consistency for the set of documents, (C), as 1 divided by the sum of theses squares of normalized proportions of documents per unique structure. The resulting measure of consistency ranges between 0 and 1 with larger values indicating greater consistency and smaller values approaching 0 representing greater variation. The measure of consistency is represented by equation:

Equation 5.1 Measure of Consistency

$$C = 1 / \sum_{i=1}^{U} \left(\frac{d_i}{\max(Ud)} \right)^2$$

I chose this approach compared to other measures (e.g., Shannon or Simpson's Index) to enable standardized comparison regardless of collection size and to support a variety of distributions for document counts per unique structures. In addition, I add weight for unique structures that comprise a greater proportion of a user's documents.

5.3.5 Clustering by Content, Structure, Emoji Use

Even though the text of individual documents varies greatly, users and documents can be clustered based on similarity of document structure, content style, or emoji use. I also identify common structures used across of users, as well as identifying the users of specific structures via aggregation. These approaches support analysis of communication patterns to identify common or unique structures used, structures associated with specific types of users or groups, as well as identifying documents or users that may be related based on similar style defined by the structures used.

5.3.6 Clustering by Consistency

In addition, I cluster users based on their consistency scores for structures of document, content, and emoji use. I use the unsupervised clustering algorithm HDBSCAN (McInnes et al., 2017) as it does not require defining the number of clusters, supports multiple dimension data, finds stable clusters within noisy data, and can handle clusters of varying density, size, and shape. For each cluster, I describe behavior traits as low, medium, or high consistency for each factor based on the greatest percent of users of that cluster falling within interquartile ranges for low (first quartile), medium (second and third quartiles), or high (fourth quartile). The composition of users in a cluster is then summarized based on additional information such as keywords in user profile descriptions or labeled data such as if account has previously been labeled as bot-like. Clustering users based on consistency enables comparison and grouping of users with similar behavior patterns associated with their communication style.

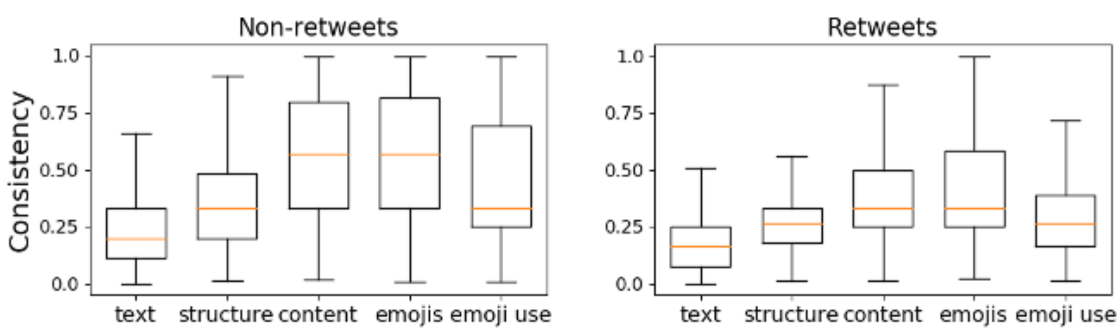
5.4 Experiment Results and Discussion

5.4.1 Experimental Setup

I apply this methodology to a corpus of 44 million tweets collected in October and November 2018 related to the 2018 U.S. midterm elections based on keywords, hashtags, and user accounts associated with candidates or political parties. For each of the 3.3 million unique users set of retweets and non-retweets I measure consistency of document structure and content structure. To improve consistency scores and reduce dimensionality, I modified the document structure by removing spans of punctuation and by not including the count of features for text spans. For the 30% of users of emojis in their tweets, I measure consistency of which unique emojis are used and also measure structure of emoji use represented as a vector of attributes, position, order, and repetition of emojis (eVAPOR). Using HDBSCAN I cluster users based on consistency scores for their retweets and non-retweets separately. I then describe the composition of each cluster and measure the percent of accounts labeled as bot-like based on Botometer scores. Next I present the results of this analysis.

5.4.2 Distributions of Consistency

I compared the distribution of consistency scores for users that sent more than two non-retweets or more than two retweets. Figure 5.2 shows the range of these scores for tweet text, document structure, content structure, unique emojis, and structure of emoji use. As expected, I found little user consistency in tweet text. In general, users were more consistent in their non-retweets than retweets, especially for content structure and which specific emojis were used. This indicates users tend to use the same order, format, and often the same emojis for their own tweets, whether knowingly or not. Users had less consistency with retweets likely a result of retweeting multiple users. Analysis of user behavior for document structure, content structure, and emojis in tweets reduces dimensionality and yields new information compared to traditional text analysis.



(a) (b)

Figure 5.2 Distribution of consistency scores for users sending non-retweets (a) and retweets (b) shown with interquartile ranges.

5.4.3 Analysis of Structures for Document, Content, Emoji Use

Using the methodology presented in Section 5.3, I identified common structures of non-retweets and retweets used by a large percent of tweets or users. Table 5.2 shows the most common non-retweet content structures with emojis. Analysis of these structures used by bot accounts led to identification of additional accounts likely to be bots not yet labeled. Tweets of these accounts exhibited identical content structure and structure of emoji use, although the tweets had different text, urls, emojis, and document structures (e.g., same content type order with variation in number of urls, emojis, and mentions). Given the similarity of the user profile descriptions and names for these users, it would not be surprising if these accounts are related. This is just one of many examples I found demonstrating structural content analysis can identify specific styles of communication that may be a signature for an individual or group of users.

Table 5.2 Most common content structures with emojis for non-retweets

Content Structure	*Percent of tweets*	*Percent of users*
[atmention, text, emoji]	18%	31%
[text, emoji]	6%	15%
[atmention, text, emoji, text]	4%	9%
[text, emoji, url]	2%	5%
[atmention, text, emoji, hashtag]	2%	4%
[atmention, emoji]	1%	4%

5.4.4 Clustering Users by Consistency

I compared consistency scores for user non-retweets and retweets across four dimensions: document structure, content structure, unique emojis used, and emoji use. Clustering users based on consistency scores in two dimensions reveals groups of users in the dataset with similar behaviors for document structure and emoji use, content structure and unique emojis used with colors representing cluster assignments, Figure 5.3. With the t-SNE algorithm I visualize the clusters of users with similar behavior across four dimensions and label the clusters with numbers, Figure 5.4.

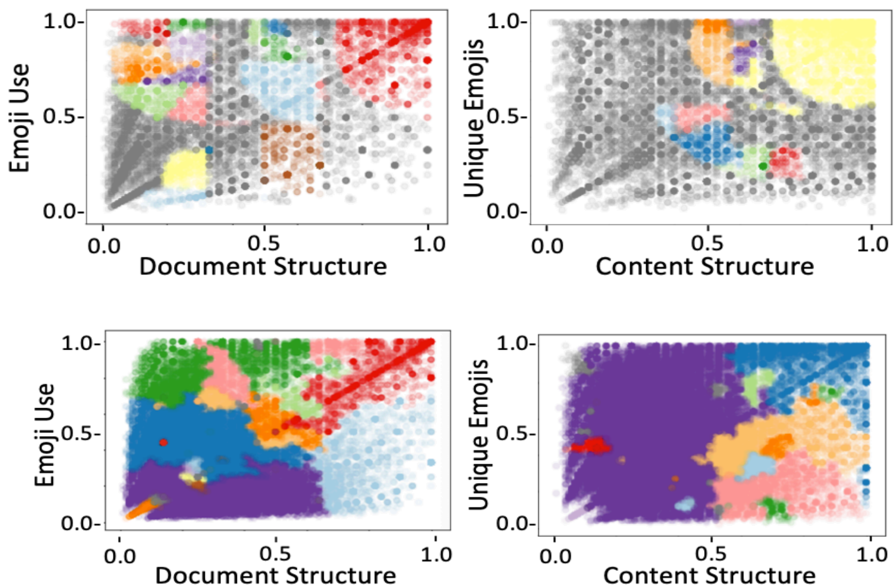


Figure 5.3 Clusters of users with similar behavior for two factors in non-retweets (top) and retweets (bottom).

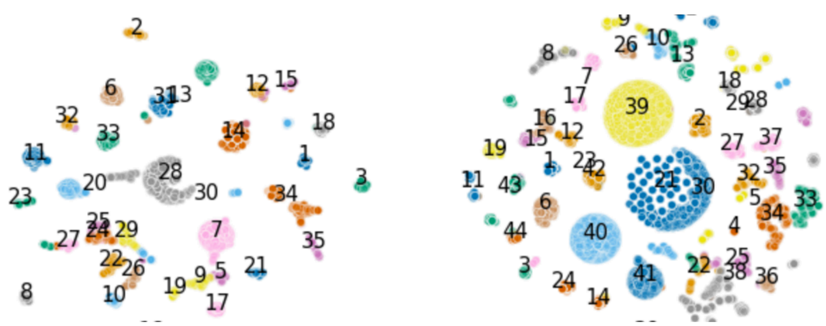


Figure 5.4 Clusters of users with similar behavior across four factors for non-retweets (left) and retweets (right).

5.4.5 Behavior Traits and Composition of Users in Clusters

For each cluster of users, I describe the behavior traits in terms of low, medium, or high consistency across each of the four dimensions. While most clusters had medium consistency for at least one dimension, 6.5% of non-retweet and 4.8% of retweet users were grouped into clusters that had high consistency across all four dimensions. I then calculated the percent of user accounts for each cluster that were likely bots based on Botometer scores (Varol et al., 2017). One of the non-retweet clusters had 45% bots, compared to the average 19% for other clusters. While not all bot-like users had high consistency scores, this particular cluster did for each of the four factors. This could indicate that additional users within this cluster may be related or also bots but not yet identified by existing bot detection algorithms.

Next, I analyzed the composition of user roles per cluster and by behavior. I define users by role based on keywords in their user profile (e.g., journalists and news organizations, marketers, businesses, celebrities, government, activists, veterans, students). Most users were clustered into groups with medium consistency scores across the four factors for non-retweets. However, the user group with verified user accounts and indicating the user is a journalist, reporter, or news organizations (which I label as 'News') had the greatest percent of users with high consistency across all four factors in non-retweets. Many of their tweets appear to be auto-generated using a template as tweets exhibited same structure but changing information such as top news and weather reports throughout the day. Similarly, 'Marketer' also had high percent of users with high consistency and tweets with similar structure and emoji use were indicating weekly or daily sales or specials. While most users had relatively low consistency in retweets, the user group of retired military veterans had the greatest percent of users with high consistency for which emojis were used and the way emojis were used in retweets. This could indicate that tweets with specific emojis and style of emoji use are more likely to be retweeted by this group. Table 5.3 summarizes the top user roles based on percent of users for categories indicating consistency of behavior associated with document structure and structure of emoji use in non-retweets.

Table 5.3 Percent of users by role per consistency category in non-retweets

		Consistency of Structure of Emoji Use		
		Low	***Medium***	***High***
Consistency of Document Structure	***Low***	9% Celebrity 8% Activist	11% Bot 10% Activist	5% Coach 4% Government
	Medium	9% Bot 7% Coach	66% Marketer 45% Coach	12% Government 11% Veteran
	High	2% Business 2% Artist	12% Student 10% Celebrity	20% News 16% Marketer

While it is not easy to verify authenticity of a user account or role, I demonstrate how to identify unique and common patterns and traits among a group of users with the same attributes in their user profile description. Overall, these results reveal how new insight can be gained by identifying and analyzing communication style patterns of individuals and groups of users with similar roles or behaviors for consistency across structures or emoji use in their documents.

5.5 Conclusions and Future Work

This chapter introduces and demonstrates a new language-agnostic approach for structural content analysis and user behavior modeling by characterizing the structure and emoji use of a document, and then measuring and clustering by user consistency. With this methodology I described signatures of communication styles and behaviors for individuals, user groups, and clusters of users. I also identified users with document structural properties and user consistency metrics similar to accounts already labeled as bot-like. Limitations of this study are that I focused on only one collection of tweets related to American politics and it is difficult to verify authenticity of user accounts. Areas for further research could compare tweet styles and author consistency for other topics and user roles such as sports, tourism, and health or message effectiveness. Structural content analysis and measuring consistency across documents, as presented in this chapter, compliments existing text mining techniques and provides a new perspective for social media analysis by linking document style and user behavior.

6 EMOJIS AS FIRST-HAND OBSERVATIONS FOR EVENT REPORTING

Social media has been shown to be a valuable resource for tracking events in real time. The use and accuracy of the text, images, and links included in posts on social media platforms with respect to natural disasters and current events are well researched. However, little research has been done with respect to the value of emojis as indicators of and as user observations of events as they unfold, which I address in this chapter. To demonstrate how emojis can be used for event reporting I use emojis in tweets during the January 2019 total lunar eclipse, a phenomenon that lasted several hours and was viewable simultaneously across large swaths of the globe. This analysis shows that the emojis used in most posts corresponded to the appearance of the moon in the user's geographic region at the timing of their post to social media. Further, information about a user's probable location can be inferred based on the emojis used in social media posts during the specific event timeframe. Based on this finding, this study moves the emoji research beyond just looking at sentiment and shows how emojis can be used for identifying the occurrence of events and for comparing first-hand observations in a way not always represented in just the text or images posted to social media.

6.1 Introduction

As a variety of events unfold around the world at any time, social media platforms are often an outlet for users to share experiences related to these events and a mechanism for journalists to obtain first-hand observations of them as they unfold. The value of social media reporting using text, images, hashtags, links, and retweets about events whether they be political, sporting, or natural disasters has been well studied (e.g., Chierichetti et al., 2014; Crooks et al., 2013; Kaneko & Yanai, 2016). However, the use of emojis for event detection in near real time is under-explored, and in this chapter, I address this issue in order to add insight about events and to complement or reveal new information not available in the frequently used medium of text, and images.

Research on the use of emojis is relativity recent. Existing research related to emojis has examined the role of emojis in communication (Ai et al., 2017) and the different meanings of an emoji that arises based on gender, device rendering, and culture (Ljubešić & Fišer, 2016). When it comes to the value of emojis in social media, often research focuses on emojis as an indicator of sentiment or sarcasm (Felbo et al., 2017). An area not as well researched is the value of analyzing emojis for detecting and reporting observations of an event as it happens.

In order to demonstrate how emojis can be used to detect an event I chose an event that was discrete, but also of interest to users who I thought were likely to share information about the event with others via social media. As the users of social media come from diverse linguistic and cultural backgrounds, I sought out an event with a large geographic coverage so I could compare emoji use across geographic regions. For this research, I analyzed emojis, keywords, and images in tweets collected during the 2019 total lunar eclipse.

The total lunar eclipse of 2019 January received a great deal of attention as it was considered a rare event. Although lunar eclipses happen a few times a year, most are only partial or a less obvious penumbral eclipse. This eclipse also was given the nickname “super blood wolf moon” owing to three conditions. A full moon in January is called the “wolf moon” in North American folklore. The color of the moon during a total lunar eclipse takes on a dark red appearance which resembles “blood”. In addition, the position of the moon on its orbit was at its nearest to Earth thus appearing slightly larger as a “super moon”. Figure 6.1 shows a representation of a total lunar eclipse showing the Earth obscuring the sun’s rays from fully reaching the moon. The moon takes on a reddish-orange appearance when fully within Earth’s inner umbra shadow. During a total lunar eclipse, the moon will pass between two of Earth’s shadows, the penumbra and the umbra. As the moon passes through a part of the umbra shadow and is partially obscured by Earth’s shadow, this eclipse phase is called a partial lunar eclipse. The phase when the moon is fully in umbra shadow is called total eclipse.

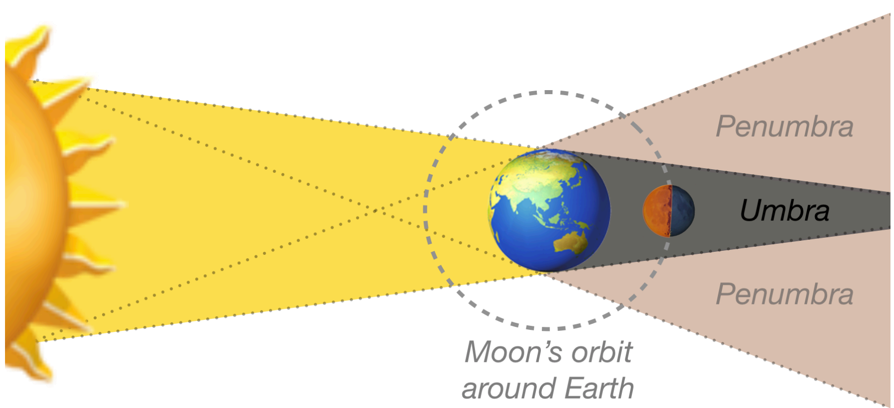


Figure 6.1 Representation of a total lunar eclipse.

The total lunar eclipse event of 2019 was ideal for this analysis as it was a discrete event lasting several hours and covered a large geographic area. The entire duration of this eclipse event was 5 hours and 12 minutes. From start to the end of the partial eclipse phases was 2 hours and 15 minutes. The duration of the total eclipse phase lasted 1 hour and 2 minutes. This eclipse was visible through-out the Western Hemisphere and parts of Europe and Africa. Figure 6.2 (Espenak, 2009) shows the geographic extent of the eclipse viewing area in white. The region shaded dark gray shows the areas where the eclipse was not visible as it was day time on this part of the globe. The eclipse event was observable by everyone at the same time in the night sky for the visible regions if the sky was clear of clouds.

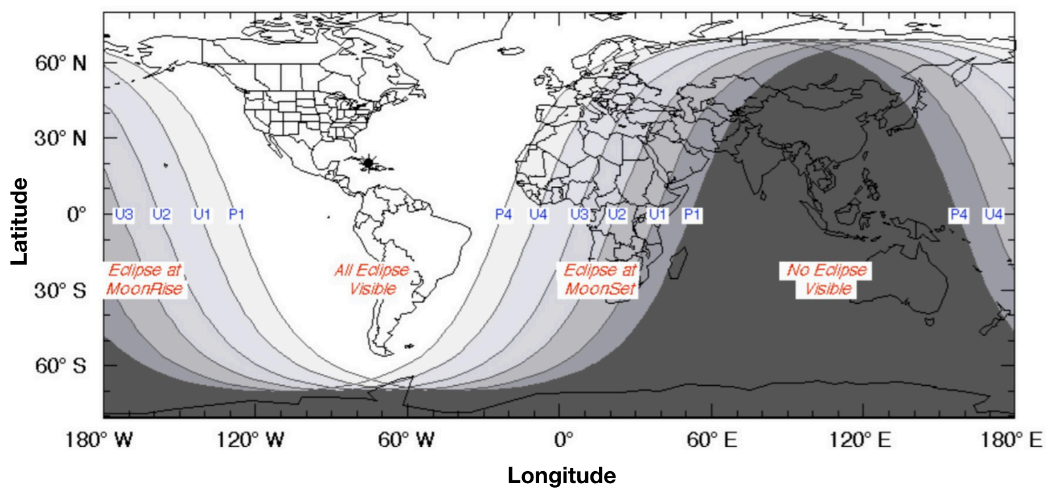


Figure 6.2 A map showing geographic extent of the 2019 January total lunar eclipse.

I examined the use of emojis in tweets posted during the 2019 total lunar eclipse as it was a single event lasting several hours and observable to a large area of the globe at the same time. In addition, the event had garnered a lot of interest for viewers around the world (Rice, 2019) which meant it was likely to be reported in social media with text, images, and as I show in this chapter, emojis too. Interest in eclipse events while not unique (Pang, 1993) is not largely studied form the perspective of social media. There has only been a limited number of research papers that discusses how social media reacts to such events, for example Srivastava (2018) and Clarkson et al. (2019) looked at the 2017 social eclipse and found nothing that focused on lunar eclipse events or emojis.

For this chapter, I examine the use of the Sky and Weather emojis resembling the different phases of the moon. I refer to these emojis as the "eclipse emojis" because collectively their appearances are similar to phases of an eclipse. The eclipse emojis also include a handful of emojis with faces that also resemble the shape of the moon in various phases. Figure 6.3 shows these emojis and the corresponding eclipse phase associated with the moon's appearance during the 2019 January total lunar eclipse, as viewed from the northern and western hemisphere in the Eastern Standard Timezone.

The motivations of this research are to determine if analysis of the collective social behavior of emoji use in social media can provide an indicator of the occurrence of events and be used as first-hand reporting of events in real time in a way that is different or complementary to the use of text or images. In addition, I consider the use emoji and the timestamp in a social media post associated with an event as a way to place the user in a particular part of the world, assuming the emoji chosen is a representation of the user's observation of the event at the time and location of posting. I compare results of inferring user locations to those provided by Twitter in the form of geo-referencing by coordinates, place, or location in the user profile. In addition to this, I also compare the emojis used to the keywords and images associated with the same post to see if there is similarity or overlap of information. The goal of this chapter, is to show that the value of emojis in social media goes beyond emojis for sentiment and demonstrates the informative value of emojis for event reporting.

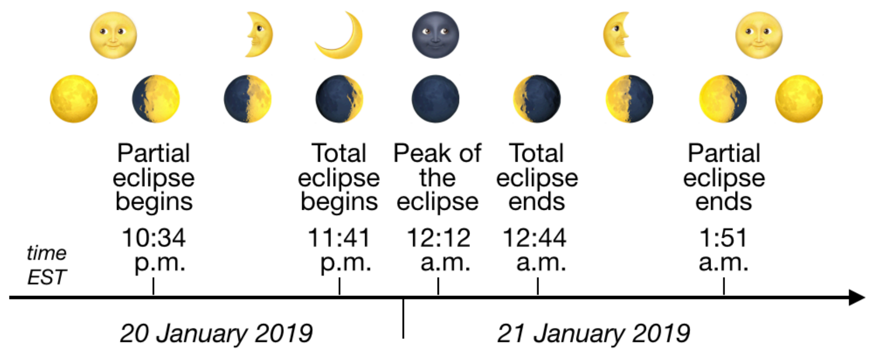


Figure 6.3 Timeline and emojis corresponding to the moon's appearance in the northern hemisphere during the 2019 January total lunar eclipse.

6.2 Methodology

In this section I describe how I collected the tweets, labeled them based on eclipse phase, and how I identified or inferred geographic location of users.

6.2.1 Social Media Data Collection

I collected and analyzed over 1 million unique tweets (i.e., I excluded non-retweets) during the time frame of the 2019 January total lunar eclipse. As the eclipse was viewable over many countries as shown in Figure 6.2, I wanted to ensure this collection would yield tweets about the event in a variety of languages. To this end, I collected tweets using keywords for eclipse, moon, and lunar eclipse translated into 29 different languages. In addition, I collected tweets based on pre-selected emojis that were most representative of the stages of the moon's appearance during the eclipse which I refer to as the "eclipse emojis" which were shown Figure 6.3. In total, 13% of the tweets I collected contained some sort of eclipse emoji. Of these tweets, only 14% contained the English word "eclipse", and 19% contained images.

Although less than 1% of the tweets had precise coordinates, the density of the geocoded tweets corresponded to the regions where the eclipse had the greatest visibility, primarily North and South America and Western Europe. This alignment with the viewing area of the eclipse was also noted in the 7% of tweets had place information at the country level, despite not having precise geolocation. I used both sets of georeferenced tweets as the basis to verify this analysis of eclipse emojis used per geographic region. I also compared these locations to the locations reported in the user profile

6.2.2 Labeling Eclipse Phase based on Timing

I used the time of the tweet creation date for the temporal analysis and labeling of the eclipse phase. I compared tweets by time at the level of detail of a minute, due to the differences in temporal detail provided by the twitter APIs. Tweets collected from the streaming API include milliseconds but tweets collected from the search API only at the minute level. I calculated epoch minute for each tweet and assigned the corresponding eclipse phase.

The eclipse phase was labeled for each tweet as taking place before, during, or after the lunar eclipse event. For tweets posted during the eclipse, for each minute increment of the eclipse, the phase of the lunar eclipse as penumbral start, partial start, total start to peak, peak to total end, partial end, penumbral end. During the partial eclipse phase, tweets posted during this time frame are given an added label coinciding with the ranges of 0-25%, 25-50%, 50-75%, and 75-100% of the moon in shadow as there are eclipse emojis that can be used to represent these more detailed eclipse phases.

6.2.3 Inferring User Location Based on Emoji

I also labeled each tweet with two probable locations based on a combination of the emoji used and the timing of the tweet posting, and the other generically based on the user profile location or language. this assumption for labeling based on emoji and timing is that users in the same geographic area will see the same view of the moon and position of the earth's shadow on the moon, especially during the partial eclipse phases. The user profile location was provided for nearly 20% of the tweets. While the location reported in the user profile may not coincide where a user is located at the time of the tweet posting, 17% of these tweets also had tweet location in the same geographic region as the user profile location and would thus have a similar lunar eclipse observation.

6.3 Results

6.3.1 Event Detection with Emojis

To identify the ability to detect events based on emojis, I compared the volume of tweets before, during, and after the event to the volume of tweets with eclipse related emojis. I found that the number of tweets per minute of those with eclipse emojis and tweets without eclipse emojis correlated to the timing of the eclipse event, Figure 6.4. In addition, Figure 6.4 shows the volume of specific emojis peaked during eclipse phase with the moon having the most the similar appearance to the emoji. During the penumbral phases, the eclipse emojis with no shading were more prevalent. During the partial and total eclipse phases, the shaded moons were prevalent and in order corresponding to the percent of the moon's surface in shadow. In tweets with and without eclipse emojis, the volume during the eclipse event was significantly more than volume pre- and post-event.

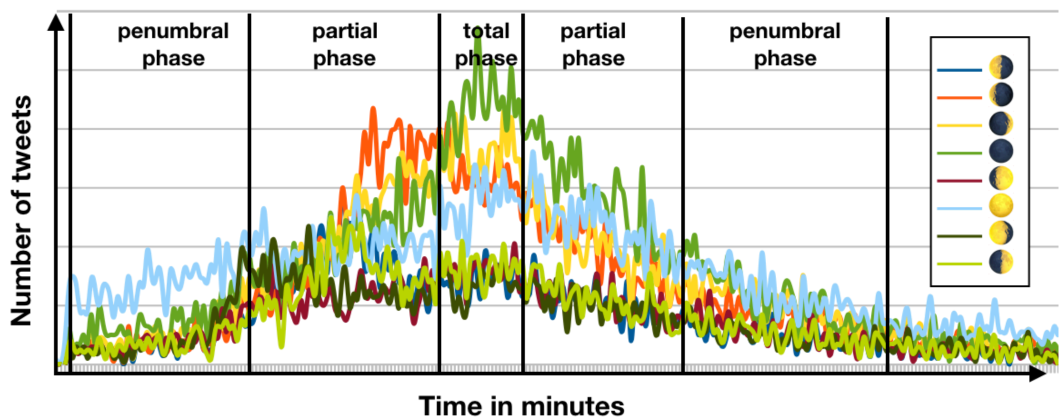


Figure 6.4 Tweet volume per eclipse emoji during 2019 January total lunar eclipse.

6.3.2 Emojis as First-hand Observation of Eclipse event

The collection of tweets consisted of only tweets that were created by the user and not a retweet. There was one emoji that was the most used throughout every phase of the eclipse event, the new-moon-with-face emoji. While the volume of tweets with this emoji also correlated highly to the event timing, the representation of this emoji only matched the appearance of the moon during the totality phase. The popularity of the use of this emoji indicates that it was the emoji used to indicate the eclipse event and the owing to the high percent of these tweets having an eclipse keyword further supports this.

Meanwhile, the second, and third most used emojis during each phase more closely matched the moon's appearance during the associated phase. For example, when the moon was either 25%, 50%, 75%, or 100% obscured by the earth's shadow, the second or third most used emojis during that portion of the eclipse also had similar appearance in terms of shading. There are two emojis for each of the 25%, 50%, and 75% eclipse emojis with shading on either side of the moon. Also accounting to the emojis as first-hand observation, users often chose the emoji with the shading on the side that most reflected the appearance of the moon in their geographic region. During the partial eclipse, the appearance of the location of the shaded part of the moon is flipped between the northern and southern hemispheres. This finding could be used to support location inference of social media posts containing emojis based on use of emojis likely to be used in association with an event of a specific location and time frame.

I further verified the eclipse emoji chosen and the appearance of the moon in some of the posts that included images. In many cases the emoji did match the appearance of the moon in the image of the eclipse included in the same tweet, Figure 6.5. However, emoji and image mis-match could be intentional, an error, or lack of an emoji with similar appearance. Many images that were reviewed appeared to be memes or photos not related to the lunar eclipse event, despite the presence of eclipse emojis or eclipse related keywords in the tweet. Future research could try to systematically classify and compare the images to the emojis appearance included in the same post (Barbieri et al., 2018; Cappallo et al., 2018).

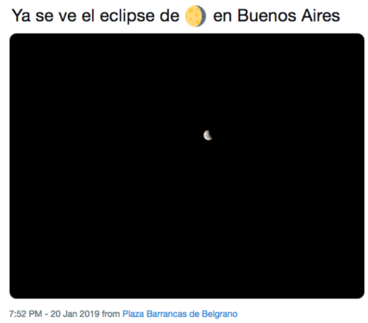


Figure 6.5 A tweet with photo of eclipse and emoji with similar appearance.

The number of tweets containing images and eclipse keywords increased several hours after the eclipse event ended while the number of tweets with eclipse emojis after the event remained very low. This is further support that as users switched roles from real-time to post-event reporting, emojis were used for a distinct communication role during the event as it occurred, in contrast to the use of images and text after the event.

6.4 Conclusion

This chapter demonstrates the value of analyzing emojis in social media to gain insight about the occurrence and observation of events. While the focus was on a single event lasting only a few hours, the event was viewable from many parts of the world simultaneously thus enabling this event as comparison of user observations in social media both geographically and temporally. Emojis were compared by users in different parts of the world as users reported their shared experience of an event happening across a large geographic region at the same time. The same emojis were generally used within a geographic region per each phase of the event. Emojis served a specific role in communication to report user observations during the event distinct from the use of event related keywords or images.

7 EMOJIS DURING EVENTS

7.1 Introduction

National, subnational, religious, cultural, and social events are an important part of establishing and preserving social identity, collective memory, and cohesion (Andrews, 2013). Often these organized annual rituals celebrate and commemorate historic or sacred national, sub-national, religious, and cultural moments, icons, or places and participation in these events is a way of acknowledging the shared identity (Frost & Laing, 2013). For example, participating in the National Day parade for Singapore (Kong & Yeoh, 1997), visiting a temple and lighting candles as part of the Hindu religious celebration for Diwali (Jha, 1976), and partaking in cuisine during food festivals for Cajun cultural heritage in the US (Esman, 1982), or honoring those who have died in conflict during Anzac Day in Australia (Sumartojo, 2016).

In the digitally connected society of today, participation in these events also occurs virtually with the use of social media such as Twitter to both broadcast event participation in person, as well as to provide a mechanism to organize, document, to virtually participate, and to acknowledge support for the participants associated with these events (Drury, 2013; McGarry et al., 2019; Tufekci & Wilson, 2012).The study of online social media communication pertaining to participation and identification with these types of events has focused on the components of social media posts such as the text, hashtags, and images posted on social media during the events (e.g., Croitoru et al., 2013; Highfield & Leaver, 2016). This chapter fills a gap in research by examining social media posts during an event with an emphasis on the role of emoji use as symbolism of collective identity associated with event participation and acknowledgement.

This chapter extends social media research on emojis by exploring how emojis are used during a variety of event types and also examines social events such as protests. It focuses on understanding the similarities and differences in patterns of emoji use as symbolic indicators for collective identity and symbols of support and solidarity during a variety of events. In addressing this research, this chapter also examines the use of emojis to detect events, and compares differences in social media samples, specifically tweets, collected based on event keywords and those tweets collected based on geographic coordinates around an area of an event location. This chapter also discusses variations in meanings of certain emojis as symbols as well as geographic variations in emojis used to symbolize shared identity during the same event spanning a large geographic area, such as International Women's Day.

The following questions addressed in this chapter are:

- What are the similarities and differences in types of emojis used during a variety of event types?
- Is it possible to detect events based on emoji use in geolocated tweets?
- In what way are emojis used as symbolism during events?

This research contributes new insights for the social science research on rituals and events by examining virtual engagement with events through computer mediated communication such as social media and how the use of emojis aids in the symbolic representation of a shared collective identity in the modern era during national, cultural, religious, and social events. In the remainder of this chapter I provide a background on related social science theories and emoji research, Section 7.2. Section 7.3 describes the data and methodology used for this research. The results are presented in Section 7.4 and discussed in Section 7.5. Finally, Section 7.6 concludes with areas for further work.

7.2 Background

Social science research has focused on understanding the role of events on shaping society (Durkheim, 2005). Events take on many shapes and forms ranging from national events, cultural/folk festivals, religious, arts, music, street demonstrations, and more (Meinert & Kapferer, 2015). They are a key aspect of society and have persisted from historic times into the modern era. Participating in these rituals both contributes to fostering a collective identity and is a foundation of culture (Andrews, 2013; Laing & Frost, 2015).

There has been a number of studies that explored the variety of motivations for why people participate in collective gatherings including individual preferences, event design, attachment to the event, availability of others to attend, and many other factors (Crespi-Vallbona & Richards, 2007; Crompton & McKay, 1997; Torres et al., 2018). Participation in local events fosters a sense of place identity and connectedness between the community and their spatial environment associated with the festival (De Bres & Davis, 2001; Ramkissoon, 2015). In addition, events and festivals provide participants with opportunities to share in a collective experience as well as to demonstrate solidarity for the community associated with the event (Jeong & Santos, 2004). Social media offers a unique opportunity to organize and communicate about events and also for event participants to share their experience and also to connect with the community and event virtually (Milan, 2015).

A key part of events and rituals is the use of symbolism as a way to demonstrate solidarity and collective identity. Symbols serve a role in society to unify a group of people around a shared idea or social construct and can signify the collective identity of people associated with that group the symbol is attributed to, whether it be a national, regional, cultural, or religious symbol (Polletta & Jasper, 2001). For example, national symbols typically include a flag, icons, emblems, and colors used as representations of patriotism to unify citizens (Elgenius, 2011; Jackson et al., 1994). The state uses these symbols and rituals such as events for national celebrations to foster identity and community (Anderson, 2006; Jackson et al., 1994).

Similarly, for cultural, religious, and also national symbols such as artistic motifs, icons, foods (e.g., Lupton, 1994), gestures (e.g., Bremmer & Roodenburg, 1992), or people and places of historical, cultural, or religious significance (Draper, 2014; Jeong & Santos, 2004). Symbols can also include flowers and their images as icons or representation as emblems or in motifs (Loy, 2020). Symbols are used to commemorate, show affirmation, or as a representation of identity and are often incorporated into rituals such as events depending on the event and the significance of the symbol (Crespi-Vallbona & Richards, 2007; Laing & Frost, 2015). Participating in events is a way of acknowledging a shared identity and can reinforce social cohesion of society (Edensor, 2002; Wilson, 1954).

As people interact and communicate, individuals present displays of verbal and non-verbal cues that reveal aspects of self-identity (Goffman, 1990). The use of symbols in these communications are a key part of these interactions and how individuals understand and identify with society (Leach, 2003). Further, the collection of symbols used by a group of individuals are also the basis for the establishment and reinforcement of culture (Swidler, 1986; Turner, 2006).

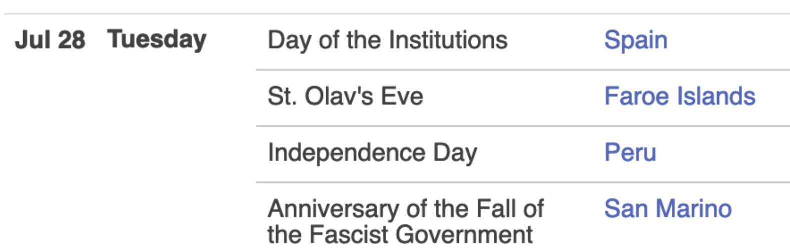
There is limited research that examines emoji use with respect to events such as holidays and protests. Recently, Kariryaa et al., (2020) examined the use of flag emojis in tweets by political leaders in Germany and the U.S. and noted a significant spike of flag emoji use during national holidays such as Day of Unity in Germany and Independence Day, Memorial Day and Veterans Day for the United States, and also for Germany during international sporting events such as FIFA world cup. The use of emojis during natural demonstrations in Charlottesville, VA in 2017 had also been examined by Barach et al., (2020) and as they compare the use of emojis categorized into smileys, gestures, and objects with respect to word use such as pronouns. And differences in emoji use by country have been noted by Ljubešić & Fišer, (2016).

While most research has focused on in person participation, this research explores online event engagement via communication in social media. This research explores the role of symbols in social media, specifically emojis, in how people communicate and self-identity as part of society during national, sub-national cultural, religious, and social events. To date, there has not been an analysis of emoji use with respect to events and holidays of significance to these countries. The research of this chapter is the first to examine differences in emoji use, such as most used emojis and differences in their use symbolically or as icons, with respect to a number of events including cultural, religious, and social events across several countries.

7.3 Methodology

7.3.1 Data

For this research, Twitter was used to explore the variations in emoji use during a variety of events. Types of events included national, sub-national, cultural, and religious holidays worldwide. The event listing on https://www.timeanddate.com/holidays/ was used to identify the date, name and location of the events, Figure 7.1. Of the events listed between January, February, and March of 2019, 33.1 million original tweets were collected across 71 events.



Jul 28	Tuesday	Day of the Institutions	Spain
		St. Olav's Eve	Faroe Islands
		Independence Day	Peru
		Anniversary of the Fall of the Fascist Government	San Marino

Figure 7.1 Event listing showing date, event name, and location.

Events were classified as national, subnational, cultural or religious based on research of an event across from mixed sources included Wikipedia, national, event, and news websites. In addition, a number of social events were also collected relating to protests, concerts, conferences, and parades. For each event, at least two days prior to and after an event, and on event day, tweets were collected. One set of tweets collected using only event keywords and another set based on geographic coordinates around the event location or capital city for national events.

Tweets collected based on keywords were done so using the name of the event in the original language and also translated name using Google Translate. In addition, geo-tagged tweets were collected using geographic coordinates for a bounding box around capital cities and locations affiliated with celebrations or event festivities based on research from event websites and news articles. For cultural events, as many were global, tweets were collected only based on keywords. While for social events, only the geo-tagged tweets were used. So as to not bias the collection of tweets for this analysis, no emojis were included along with the keywords during tweet collection. Table 7.1 provides a summary of the number of events per category, the percent of users of emojis in tweets on event date as reference, and sampling of event names collected.

Table 7.1 Summary of events collected

Event type	*Events collected*	*Average percent of emoji users on event date*	*Event names*
National	26	25%	Estonia Independence Day, Greece Independence Day, Libya Revolution day, Lithuania Day of Restoration, Serbia Day
Sub-national	11	19%	Day of Andalucía (Spain), Ashakalia Day (Kosovo), Johor Sultan Day (Malaysia), Adwa Victory Day (Ethiopia)
Religious	13	28%	Vasant-Panchami (Hindu), Holi (Hindu), World Day of Peace (Christian), Purim (Jewish), Birthday of Ali ibn Abi Talib (Islam)
Cultural	8	36%	Valentine's Day, Nowruz, St. Patrick's Day, Mardi Gras, International Women's Day
Social	13	33%	Concerts (Venezuela Live Aid), conferences (South-by-Southwest), protests (Algiers, Sudan), large public gatherings (Turkey elections)

7.3.2 *Baseline Emoji Use and Tweet Volume for Event Date*

Tweets were collected for national, sub-national, and religious events for two days before and after the event, based on local time of the event. As tweets are provided with a timestamp in UTC time, after tweets were collected for an event, an additional timestamp was generated to indicate the time of the tweet posting in local time with respect to the event. This normalization of time enabled analysis for an event with a single date, based on the local time. Countries having more than one time zone, the time zone of the capital of the country was used for local time. For cultural events, most spanned large geographic areas and multiple time zones so no adjustment to time was made. Most social events collected spanned multiple dates so no correction for local time was made. Using the adjusted time when applicable, or the UTC time, temporal analysis was conducted to establish a baseline of tweet volume and percent of users per emoji during the time period of collection per event. In addition, the volume of tweets, users, and percent of users of specific emoji on event date were used to compare across events and event types.

7.3.3 *Most Used Emojis*

Tweets containing emojis were further processed to describe the emoji attributes to enable comparison of which emojis were used across event tweets and users. For each event, the percent of users of a single emoji and of co-occurring emojis in the same tweet were calculated for each date of the collection. The top 10 emojis and co-occurring emojis on each date collected per event, were identified for comparison. Emojis in the top

10 with less than 0.1% of users were not compared in order to have a standardized threshold and to focus on collective emoji use. Using the top emojis identified in this way enabled temporal comparison of most used emojis, based on percent of daily users, and also event date comparison across events.

An additional comparison of daily emoji use was done against a baseline set of emojis. To demonstrate that emoji use in Twitter is a heavy-tailed distribution with a few emojis used in an extremely large volume of tweets, emojitracker.com was used to identify the number of tweets per emoji and then plotted in Figure 7.2. Using the Pareto curve, the top 10 emojis account for 42% of the volume of emoji tweets. These top 10 most used emojis in tweets as recorded by emojitracker.com are shown in Figure 7.3 with their respective tweet counts for reference. With this set of emojis accounting a large of the emoji use, in addition to comparing emojis by date within an event collection, emoji use is also compared against these top 10 emojis as a baseline.

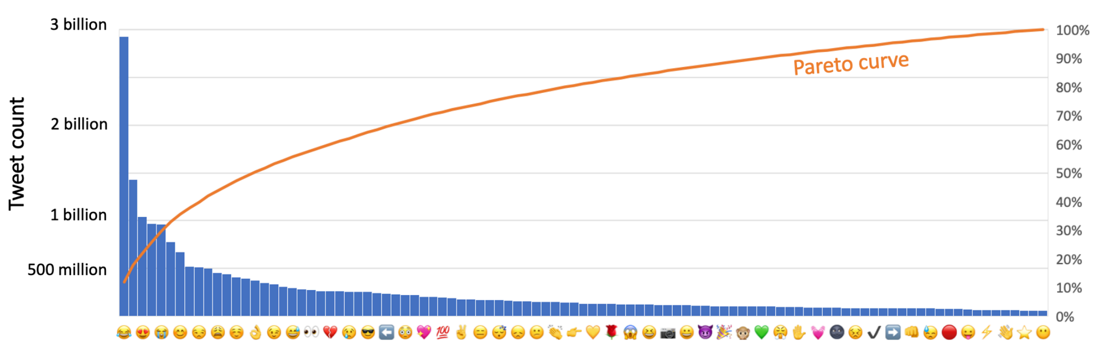


Figure 7.2 Top 100 most used emojis in tweets as of 24 July 2020.

😂	2925416259	❤️	1430764536	😍	1043126847	♻️	966680411	😭	961593535
♥	775846802	😊	672543213	💕	520399648	😒	514879947	😘	501205477

Figure 7.3 Top 10 most used emojis in tweets as of 24 July 2020.

7.3.4 Emoji Attributes

For the top emojis identified on event date, the emoji attributes were summarized in order to compare emoji use across events. The emoji attributes, shown in Table 7.2, are used to describe the top emojis and include features such as Unicode group, subgroup, and from the emoji name the type, color, and shape are derived. Type and anthropomorphic type were mapped based on sub-group. Future research could remove or add attributes such as sentiment or other groupings.

Table 7.2 Example of emoji attributes

Emoji	*Group*	*Sub-group*	*Name*	*Type*	*Anthro-type*	*Shape*	*Color*	*Direction*
☝🏽	People & body	Hand-single-finger	Index pointing up: medium skin tone	Anthro	Hand-gesture	None	Medium	Up
🔵	Symbol	Geometric	Blue circle	Shape	None	Circle	Blue	None
🇪🇸	Flag	Country-flag	Flag: Spain	Other	None	None	Red, Yellow	None
⚜	Symbol	Other	Fleur de lis	Other	None	None	None	None

7.4 Results

This section describes emoji use in tweets for the national, subnational, religious, cultural, and social events collected. It begins with an overview of the temporal analysis of tweet and user volume for an event (Section 7.4.1). This is followed by a summary of the main patterns observed for event type (Section 7.4.2). While this section does not go into detail for all 71 events, the results shown are representative examples. Common patterns of emoji use are summarized and differences of an event compared to other events of the same type are noted.

7.4.1 Temporal Analysis of Emoji Use During Events

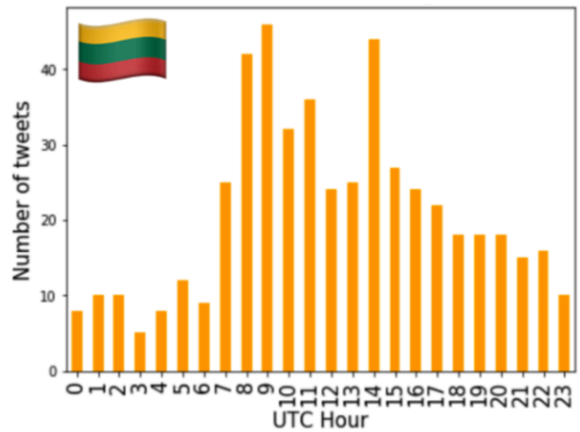
The majority of events analyzed, had a distinct peak of activity on the event date in terms of tweet and user volume, but there was a greater proportion of tweets on event date in the collections based on keywords. In geotagged tweets there was also a peak in activity unless the event did not have a large attendance of users posting public tweets or the bounding box area used to collect tweets was too large and resulted in over sampling of tweets.

Changes in the overall percent of emoji tweets and percent of emoji users on event date were compared to the days before and after the event. The results indicate that there was typically not a big change in the overall percent of emoji users on event date. However, proportions of users of specific emojis on event date, did reveal significant peaks of activity compared to the days leading up to and after the event. In addition, the peaks of activity for some of these most used emojis on event date also corresponded with the local timing of the event. For example, in analysis of the volume of tweets with the Lithuanian flag emoji, two peaks were identified, one corresponded with the timing of a national holiday, The Day of Restoration, and the other an international competition, Eurovision, with a Lithuanian participant. Figure 7.4a shows the relative proportion of users of the Lithuanian flag emoji across several days, and Figure 7.4b shows two tweets from the collection, one for the national holiday, the other for the Eurovision contestant.

(a) (b)

Figure 7.4 Proportional tweet volume with Lithuanian flag (a) and example tweets.

Temporal analysis on the day of the event by hour often correlated with local timing of events. For the Lithuanian flag emoji, these peaks on the day of the national holiday corresponded to the timing of the parade in Vilnius the capital city as well as other festivities on February 16 for the Day of Restoration, (Figure 7.5 top). Analysis of the hourly volumes on February 22 of tweets with the Lithuanian flag emoji shows a concentration of activity during the hour when the Eurovision competition was televised (Figure 7.5 bottom). In general, the temporal pattern associated with specific emojis peaking during the timing of local events was observable in most of the keyword collection datasets across all events, and about half the time in the geo-tagged tweets.



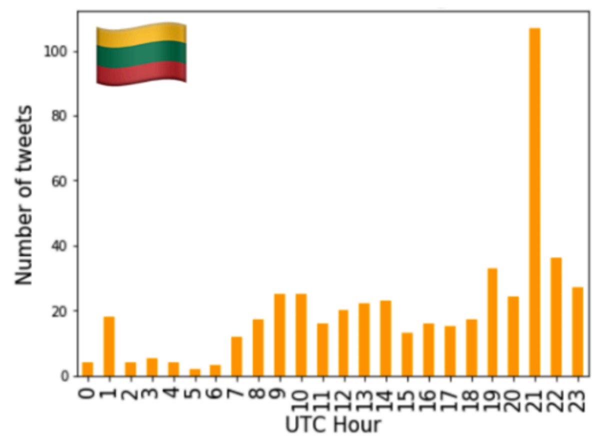


Figure 7.5 Tweet volume with Lithuanian flag emoji on 16 February 2019 national holiday (top) and 22 February Eurovision (bottom).

7.4.2 Summary of Emoji Use by Event Type

National and subnational events. From the analysis of the results of emoji use on event date for national and subnational events, six common patterns of emoji use are identified. National events include Independence Day, Memorial Day and other holidays that typically had the country name as part of the event name. name of the country Use of the emoji of the national flag for the country hosting the national event was the most common pattern of emoji use across users on the date of national events 83% of national events analyzed. For subnational events, the country flag was the most used emoji for

27% of sub-national events. The use of two flag emojis co-occurring in the same tweet occurred in 60% of national events and only 18% of subnational events. Additional common patterns of emoji use during these two types of events included country flag and a single red heart, the country flag and either a symbol or a shape, such as a heart, with a color included in the flag, the use of shapes with colors in order of the flag colors, or a single symbol. Table 7.3 summarizes the patterns and percent of events with that pattern.

Table 7.3 Common patterns of emoji use in national and subnational events

Description	***Emoji Pattern***	***National***	***Subnational***
Single flag	🇧🇦	83% of national events had the country flag as the most used emoji by percent of users, ranging in percent of 16% for Lithuania to 1.2% for Anguilla	27% had the national flag as most used emoji, 53% had the national flag in the top 7 most used emojis
Two different flags	🇧🇦, 🇹🇷	60% of events had the country flag of the event and one other country flag	18% had the country flag and another country flag
Flag with red heart	🇧🇦,❤️	57% had the country flag and a red heart	8% had the country flag and a red heart
Flag with symbol or heart with color of flag	🇧🇦,⚜️, 💙	43% had the flag and a symbol or flag color	27% had the country flag with a symbol or heart with color of the flag
Hearts or symbols with colors of flag	💙, 💛	35% used hearts with colors of the flag or country colors	39% used hearts or symbols with colors of the flag
Other symbol	⚜️	13% also used other symbols	18% used symbols

Religious events. Emoji use during religious events showed distinct emojis per event, with a handful of emojis in common across events of the same religion. Table 7.4 shows the top emojis used on event date for a handful of religious events collected. The differences of emoji use are explained in the discussion.

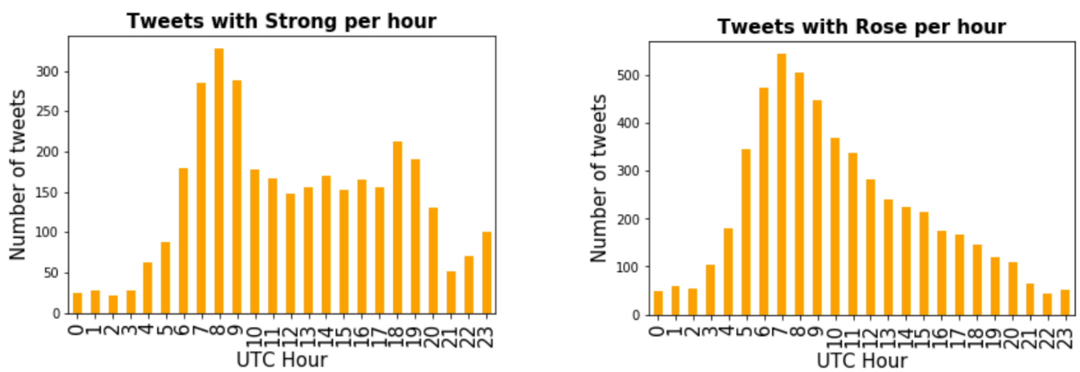
Table 7.4 Summary of emoji use for religious events

Religion	***Event name***	***Most used emojis across users***
Hindu	Vasant Panchami	🙏, 💐, 🌹, 🌻, 🚩, 💛, 📚
Hindu	Maha Shivaratri	🙏, 🌹, 🚩, 💐, 😂, 🙏🏻, 🔱
Christian	World Day of Peace	😇, 🙏🏻, ✌️, ❤️, ✌🏻, 😂, ✨
Christian	Beginning of Lent	❤️, 😂, 💛, 💙, ♥️, 🌿, 😭
Muslim	Birthday of Ali ibn Abi Talib	❤️, 💕, 🌹, 💙, 💜, 💖, 😭
Jewish	Purim	🎉, 🎭, 🤡, 🇮🇱, ❤️, 🥳, 🔯
Hindu	Holi	🙏, 😊, 😍, 💐, 😂, ❤️, 🌈
Christian	Orthodox Feast of Annunciation	🙏, 😂, ❤️, 🌹, ⛪, 🔔, ✝️

Cultural events. For cultural events, because of their global or regional coverage and popularity, I only used keyword tweets for analysis of the top emojis used on event date. Cultural events had the highest average percent of users putting emojis in tweets and also the greatest percent of commonality of emojis use across tweets. The top emojis used for most of these events was also similar across users despite the user location as described in the user profile. For example, hearts "💕,❤️" and roses "🌹" were the predominant emojis used on Valentine's Day, clover and shamrocks on St. Patrick's Day "☘️, 🍀" , and with the same colors and symbols of emojis for Mardi Gras, i.e., "💜, ⚜️, 💚, 🎉, 💛, 🔥, 😏".

One event in particular that did have geographic variation in emoji use by percent of users, was International Women's Day. While the two most used emojis were the rose emoji "🌹" and flexed bicep emoji "💪", the volume of tweets per hour on event date

(Figure 7.6) indicate possible geographic variation in the use of these emojis associated with the event, which is further explained in the discussion.



(a) (b)

Figure 7.6 Volume of tweets by hour during International Women's Day on 8 March 2019 for: (a) 💪flexed bicep emoji and (b) 🌹red rose emoji.

Social events. Geo-tagged tweets were collected for various social events, such as protests, concerts, and conferences based on bounding-box coordinates of approximately a 2-mile radius around the location of events as identified from news or event websites. Many of the social events collected, the top emojis included many of the baseline emojis and emoji use was not further analyzed. For example, the predominant emojis in public tweets during the south-by-southwest conference were "😂, 😭,❤️, 🤣, 😍, 🤔, 💯" which is very similar to the baseline emojis, thus indicating the event did not have enough users of the same emojis specific to the conference and more users than of baseline emojis.

The social events pertaining to national issues, such as protests and elections, did have a large enough volume of users of specific emojis not in the baseline emoji set and thus the event was noticeable based on emoji use and those results are summarized here.

For the Venezuela Aid Live concert in Cucuta, Colombia in 22 February 2019, the top emojis used on the day of the event, based on percent of users were "🇻🇪, 😍,❤️, 😂, 😭, 💛, 👏". Although the face with heart eyes emoji and the red heart are emoji are both in the baseline of emojis, their use peaks significantly higher on event date than for the baseline emojis such as tears of joy of emoji use as shown in Figure 7.7.

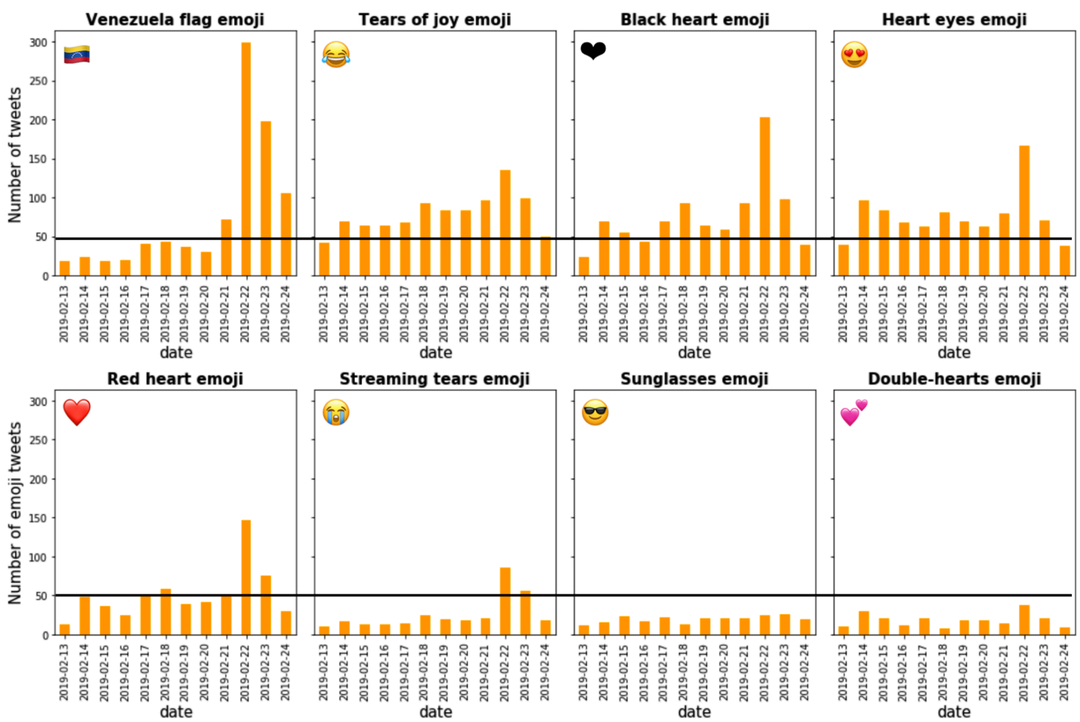


Figure 7.7 Tweet volume at Cucuta, Colombia in February 2019 for emojis.

In Turkey on the date of the elections, although the most used emoji was the baseline tears of joy emoji, there were other emojis in the top 7 most used emojis that were correlated to the timing of the event. The top emoji use was also compared between

Ankara 😂, 😊, 😁, 🙏, 🇹🇷, 🤣, 👍 and Istanbul 😂, 🙏, 👏, 🇹🇷, ❤️, 😊, 🤣. Emojis that entered the most used emojis based on percent of users on that date included 🙏 and the country flag 🇹🇷. The hand gesture for thumbs up 👍 and clapping also were among the most used emojis 👏.

During the time frame of collection were several protests in France, Venezuela, Algeria, Uganda, and Sudan. Many of these multi-day protests were about national reforms or government legitimacy. The use of the country flag emoji and also hand gestures were common within the top most used emojis across these events. For example, in France, the occurrence of the French flag and a thumbs up emoji, 🇫🇷 👍. In Algiers, the country flag, hand with index finger pointing down, and prayer hands, 🇩🇿, 👇, 🙏. In Caracas the country flag 🇻🇪 and use of pressed hands 🙏 and ✋. In Sudan, the use of a broken heart 💔 and fire 🔥 during the protests was common, and on the date that the government was dissolved, the most used emojis changed to 🇸🇩, 🔥, 😂, ❤️, ✌🏽, 😭.

7.5 Discussion

7.5.1 Event Detection with Emojis

For event detection with emojis, it does seem possible for some types of events and for some communities, but requires a baseline of familiarity of emoji use norms for the geographic area or country associated with the event, as well as trends by date in order to compare and identify changes in emoji use that could indicate an event. This could be done using temporal comparison of proportion of users and tweets for specific emojis likely to be associated with an event and compared against a baseline of volume.

However, the use of event keywords in addition to emojis may be more effective and enable filtering out tweets that may include the emojis but not be about the event. This would also yield a larger volume of content for analysis as on average only 28% of users used emojis in original tweets on event date.

In regards to the use of geo-tagged tweets, the predominant source of the coordinates in the geo-tagged are based on the location provided by the user in their account profile, typical a country-name, city, or neighborhood. For most of the events collected, on average about 60% of the users had user locations that were the same as the country of the event being analyzed. Only 2% of the tweets in the data collection contained precise coordinates. Using only these precise coordinates to identify tweets with emojis from users participating in an event was sparse and did not yield enough data to generalize. While this location may not reflect where the user is located at the time of posting a tweet, it does enable the analysis of what people with that user location are saying and which emojis are being used.

7.5.2 Emojis and Symbolism During Events

National and subnational events. Use of emoji flags, emojis as icons of national symbols, and shapes with colors of the country flag were among the most used emojis based on percent of users on event date for national holidays, sub-national events, political events, and during other events such as protests. The national flag and symbols are also used for non-secular events to show support for contestants in international competitions whether it be sporting events or even the televised talent competition Eurovision (Highfield et al., 2013) as shown for Lithuania (Figure 7.8) and are used as symbols of national identity or to represent solidarity, such as sporting events. Kariryaa et al., (2020) also found that use of the emoji with the national flag to peak during national holidays and also sporting events. However, their analysis only measured emoji flag in volume of tweets on event date for two countries United States and Germany.

(a) (b)

Figure 7.8 Flag emojis in tweets for (a) international curling and (b) Eurovision.

The use of the emoji with the national flag, or a symbol associated with the country, were greatest by users during national events who also had the name of the country in their user profile location, which points towards the use of these emojis possibly as symbols of national identity. An example of the use of emojis resembling national symbols was evident in the tweets for Bosnian Independence Day. The fleur-de-lis emblem, used to symbolize the golden lily that was on the coat of arms for the medieval Kingdom of Bosnia, was declared as a national symbol in 1992 when the Republic of Bosnia and Herzegovina became independent (Vincent, 1999). In tweets associated with this event, the emoji of the national flag "🇧🇦" was the most used emoji, followed by the fleur-de-lis emoji ⚜, which was often used on its own and alongside keywords of the event name.

The use of two country flags co-occurring in the same tweet, with one the host of the event and the other flag of another country (e.g., "🇪🇪🇱🇹") were common during national holidays. The use of two country flags in most of the national tweets were as affirmations of support. For example, the two tweets in Figure 7.9 include two country flags as Lithuania acknowledges Estonia (a) and Bhutan to Bangladesh (b) for the respective national Independence Day.

(a) (b)

Figure 7.9 Two tweets showing one country acknowledging the other during national events: (a) Lithuania to Estonia, and (b) Bhutan to Bangladesh.

Emojis with shapes such as hearts or circles with the country colors and displayed in order of the flag colors, such as top-down or left-to-right, was common for some national and sub-national events. For the Lithuanian Day of Restoration, the use of the country flag emoji "🇱🇹" and also a set of shapes with colors in order of the flag, (e.g., "💛💚❤️") were common. Similarly, the use of shapes with colors associated with sub-national regions such as provinces was more common in sub-national events. The use of co-occurring shapes with provincial colors, for example green and white circles "🟢⚪" were common during the Day of Andalucía, in Spain. Also common with this event was the use of the emoji for the Nigerian flag "🇳🇬". The Nigerian flag has similar style as the flag for Andalucía in that both have two green outer stripes and one white stripe in the middle. Although the orientation of the stripes for these two flags is different where the stipes are horizontal for Andalucía Figure 7.10a and vertical for Nigeria Figure 7.10b.

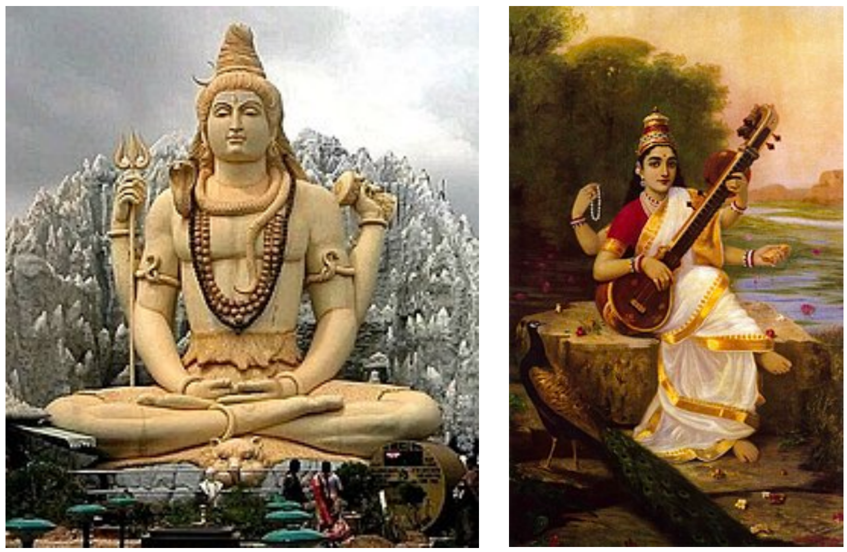
(a) (b)

Figure 7.10 Green and white flags for: (a) Flag of Andalucía and (b) Flag of Nigeria.

Religious events. The use of emojis as symbolic representations and icons during religious events was popular among users on event date. While most of the top emojis per event were unique to the event, there were a few emojis that were common across multiple events. The emoji of pressed hands "🙏" was the most used emoji in both Hindu and Christian religious events, yet the meaning of the emoji differs by religion. For Christian events it symbolizes prayer. For Hindu, the pressed hands emoji is an icon for the gesture of pressing hands together as is done during greetings and as a show of reverence (Singh, 2015). Also common across the Hindu religious events was the bouquet of flowers emoji "💐". The symbolism and icon of this emoji represents the sacred use of flowers as a part of religious rituals and events (Singh, 2015).

For Hindu religious events, besides the common emojis just mentioned, the top emojis used are symbolic specific to each event. During the holiday Maha Shivaratri, the trident emblem emoji "🔱" was popular during this event as it associated with deity Lord Shiva for which the holiday is based (Flood, 1996). Figure 7.11a shows a photograph of a statue representation of Shiva holding a three-pronged spear called a trident (Wikipedia, 2020).

For Vasant Panchami event, kite flying is a common ritual during this celebration (Arora, 1986; Desai, 2010). Yet one of the most used emojis was a red flag "🚩" and not the emoji of a kite "🪁". The reason for this is that the red flag was used to symbolize kite flying as the kite emoji was not yet available mobile devices or Twitter at the time the data was collected. The emoji did become available later in 2019 and a future comparison of emojis for this holiday would be interesting to see if the kite emoji is used. In addition, the use of the emoji resembling a stack of books "📚" was used to symbolize knowledge, as this holiday is a celebration of Saraswati, Figure 7.11b (Varma, 1896), the Hindu goddess of knowledge, music, art, wisdom (Kinsley, 1986).



(a) (b)

Figure 7.11 Depictions of Hindu deities (a) Lord Shiva and (b) Saraswati.

Another example of emojis taking on both iconic meaning to represent a real-world object and at the same time is a symbol of the event, is the use of emojis in social media during the Jewish holiday of Purim. During Purim, the use of the performing arts emoji "🎭" or the clown face emoji "🤡" is both iconic and symbolic. These emojis are iconic since costumes and masks are often worn during the celebration. The use of these emojis is also symbolic for Purim as they represent how Queen Esther hid her Jewish identity in order to save the Jews of Persia (Fishbane, 2016).

Cultural events. Cultural events had the greatest percent of users of emojis associated with the event. While I already mentioned a few in the results, here I will focus on the variation of emojis used across geography, indicating emojis as symbols of cultural norms, during International Women's Day. I include this event as culture because it takes place around the world in many countries and is a celebration of the global culture of women. Culture is a social construct arising from the shared meanings of concepts often expressed via symbols and rituals (Geertz & Darnton, 2017). While the emojis varied, they shared the concept and meanings associated with women.

International Women's Day has been celebrated annually for decades with parades, protests, and gatherings in cities all around the world, and more recently participation is also online such as posting tweets with the hashtag or keywords for the event. Each year the event theme and example resources to organize events are communicated via the website https://www.internationalwomensday.com/ but mobilization for events is done at the grass roots level. On the website, it is explained that the main color associated with the event is purple, the internationally recognized color to symbolize women. While many participants wore purple shirts while they attended the event in person or used purple lettering on banners and signs, the use of shapes with purple color, such as purple heart emojis "💜" were not among the top emojis in the public tweets collected based on keywords for this event.

Also missing from the list of top emojis was the emoji of the female symbol "♀". While it is unclear as to why this emoji was not more widely used, the availability of the emoji does vary based on device. For example, the female symbol does not render with emoji presentation on Apple brand devices, which means it appears as the black glyph form. In addition, it does not appear in the emoji keyboard selection from Apple mobile device, which would limit the ability of users to select this emoji and include in a tweet from these devices. Table 7.5 shows the appearance of the female sign on Twitter and various devices as shown from the web page Emojipedia in 2020 (Emojipedia).

Table 7.5 Renderings of the female sign emoji

Emoji name	***Twitter***	***Apple***	***Google***	***Samsung***
Female sign				

Analysis of these top emojis with precise coordinates and the geographic locations specified in the user profiles revealed there was geographic variation in symbolism used during International Women’s Day. This was also supported by the temporal peaks of user volume of specific emojis by hour. The top most used emojis for North America and Western Europe were similar, however the most used emoji from this region was not used in other regions. Similarly, the most used emoji in Turkey and in India were not among the most used emojis in other areas. The differences in top emoji by geographic region indicates differences in cultural symbolism representing women for this event. A map showing this regional variation in cultural symbolism associated with emoji use on International Women’s Day is shown in Figure 7.12.

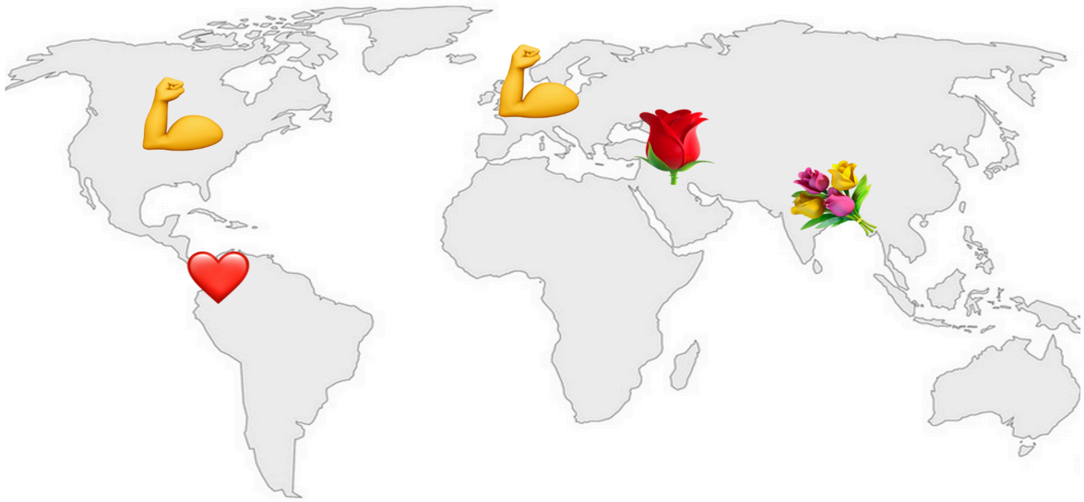


Figure 7.12 Regional emoji use during 2019 International Women's Day.

In the U.S. and Western Europe, one of the most used emoji during the event in 2019 was the flexed bicep emoji "💪" which is considered as a symbol of strength. The emoji has been used with similar symbolism in other collective events and social movements as well, such as protests (Barach et al., 2020). This association is based on the icon of a woman with a flexed bicep working in a U.S. factory, as depicted in the World War 2 poster We Can Do It! by J. Howard Miller (Figure 7.13, (J. Howard Miller, 1943)) and a similar painting by Norman Rockwell depicting a woman with a rivet gun and a lunch pail with the name "Rosie". The association of woman empowerment with Miller's We Can Do It! poster has only been since the 1980s (Kimble & Olson, 2006; Santana, 2016).

Figure 7.13 J. Howard Miller's poster of a woman with flexed bicep.

The most used emoji in India and Turkey during International Women's Day were flowers. It is not unusual for flowers to be symbols of sacred meanings, femininity or womanhood, and even national identity (Loy, 2020). For India, the use of the bouquet of flowers emoji "💐" was the most used emoji during International Women's Day. This emoji was also among the most used emoji in the tweets for Hindu events as flowers are a common part of festivals and celebrations in India.

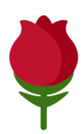
For Turkey, the popularity of the red rose "🌹" emoji may be based on its similarity of appearance to another type of flower, a red tulip. The tulip is a cultural symbol of the Ottoman empire and the is national flower of modern-day Turkey (Karabacak & Sezgin, 2013). It originated in Central Asia and was exported to Europe in the 15th century (Pavord, 2019). The appearance of the of the red rose emoji icon on

Twitter more closely resembles a tulip than does the tulip emoji. Figure 7.14 shows one of the first botanist drawings of a red tulip (Gessner, 1561) and a photograph of a red tulip (Burden, 2016). Table 7.6 shows the differences in the appearance of the red rose emoji and the tulip emoji on Twitter and mobile devices.

(a) (b)

Figure 7.14 Red tulip: (a) drawing and (b) photograph.

Table 7.6 Renderings of the red rose and the tulip emojis

Emoji name	***Twitter***	***Apple***	***Google***	***Samsung***
Red rose				
Tulip				

Social events. Common across the social events collected that had top emojis on event date that were significantly different than the baseline of emojis were events relating to protests or gatherings relating to secularism such as Turkey elections, Venezuela aid live concert, and protests demanding national reforms in France or change of government leaders in Algeria and Sudan. The use of symbols in person and on social media is a way to both communicate and to foster collective identities during social movements such as protests (Gerbaudo & Treré, 2015). During the protest events examined as part of this research, the common emojis used were both iconic and symbolic in their use. The emojis for the respective national flags were among the top emojis used as well as hand gestures which both were used in person events and also online. However, the meanings attributed to these hand gesture emojis is also at times linked with their respective events.

The use of the pressed hands emoji "🙏" was common in the tweets associated with the Hindu and Christian religious events as sign of reverence and prayer, but its appearance among the most used emojis during some of the protests was as symbolism of concern or support (Santhanam et al., 2018). The use of the raised two fingers hand gesture, ✌️ is an emoji that used among the top emojis for the Christian religious event World Day of Peace. It also represents the symbol for "V" as in victory. However, in Sudan the use of this gesture by protestors is also as a show of solidarity during months long protests against deteriorating economic conditions and the Omar al-Bashir regime. The gesture was used during protests, social media, and even painted on billboards, as shown in Figure 7.15a (Leithead, 2019).

Another example of a hand gesture that was a symbol of solidarity was during the protests in Venezuela. The use of the raised open hand "✋" was to signify a show of support for Juan Guaido and is seen in images of a protest in Caracas in January 2019, Figure 7.15b (Parra, 2019). Clenched fist as an emoji represents the raised closed fist hand-gesture that has been associated with social movements and protests as a rallying symbol for solidarity and power (Davidson & Blair, 2018). This symbol is associated with Black Lives Matter protests in the United States and is also was used in Lebanon in late 2019 as a symbol for revolution, Figure 7.15c (Hussein, 2019).

(a) (b) (c)

Figure 7.15 Use of hand gestures during social movements in 2019: (a) Sudan, (b) Venezuela, and (c) Lebanon.

While most emoji analysis of social media uses face-gesture emojis as a valence for sentiment, this research shows that the use of these emojis was not as common during the events collected on event date, thus indicating the need for researchers to consider the broad array of emojis and their role in social media messaging as symbols of identity and other functions.

7.6 Conclusion

Social media use during events can be analyzed to identify the types of symbols used in communications and online interactions associated with national, sub-national, cultural, religious, and social events. The analysis of the emojis associated with events reveals symbols of shared identity and solidarity among users during events. This research also shows that some emojis are used across multiple event types and take on different meanings based on the event. Further, unlike words which are symbols, an emoji can be both iconic and symbolic in its use.

There are some limitations and areas for future research. The sample of users is biased based on the social media platform, demographics, and preferences for use of emojis. In addition, emoji use does not always indicate self-identification or acknowledgement and could be intended for other purposes although this will be difficult to discern. In addition, the authenticity of the users and the location information is not guaranteed. Future research could combine research methods to survey social media users or event participants about which emojis are more likely to be associated with a particular event or why specific emojis were used in social media posts. Additionally, as this study only examined events between January and March, a longitudinal comparison of emoji use over time for specific events would add to the robustness of this study. In addition, additional computational methods such as network analysis and text mining may yield additional insights about the connection of users and differences and overlaps between keywords and emoji use during events.

This is one of the few studies to bridge social science theory and events with a focus on emoji use in social media, however it is the first to examine sub-national, religious, and cultural events. Emoji use does provide additional insight about symbolism, identity, and digital engagement related to an event which keywords alone do not provide. Studying emoji use during events provides information about society norms and values ascribed to various symbols and icons as reflected by emoji use in social media during the event and also their use provides cues about collective identity of users during a variety of national, cultural, and religious events.

8 EMOJIS AND PLACE

Emoji use in social media has been examined for their role as face gestures and emotional reactions in digital communication, and as cues about individual and collective identity. This chapter explores the use of emojis in tweets with respect to place using Washington DC as a test bed. Using geotagged tweets collected 2014 through 2017, tweets with emojis are used to identify local events, map locations based on diversity attributes, and characterize the emojis used associated with a variety of types of places. The results show in addition to identifying individual and collective behavior of users, emojis also provide cues about the function of types of places.

8.1 Introduction

Social media researchers have focused much research on the connections formed of social media users, such as on Twitter, as they interact with other users and content in the form of retweets, mentions, and replies (e.g., Schuchard et al, 2019). Content analysis has also been used to identify common themes across groups of users (e.g., Stefanidis et al, 2017), as well as to identify individuals based on their linguistic styles (e.g., Danescu et al, 2011). While most researchers have focused on user to user interactions, the study of social media with respect to user and place interactions has not been well researched.

Analysis of the content contained within geotagged tweets has been used to map geographic variances in reactions to current events (e.g., Yuan et al, 2020). It has also been used to identify the themes of interest to communities based on census tracts (Cranshaw et al, 2012). With respect to places of interest, the words contained in tweets have been used to describe the geographic areas based on the types of place mentioned in tweets (Panteras et al, 2015) as well as the types of keywords associated with various types of place (Gazaz et al, 2016), and sentiment of places (Padilla 2018). What has not been fully explored is the use of emojis with respect to these types of places.

In this chapter, place is considered a social construct and meaning of place is attributed to a state of mind or geographic space based on an experience or interactions. Places are often designated on a map based on the geographic representation and a place name, such as the city of Washington DC, neighborhood of Dupont Circle, natural or water area called the Tidal Basin, one of the many locations of the coffee shop Starbucks. These geographic areas that serve particular function and typically have defined set of coordinates are typically referred to as places of interest. On maps, the places may appear as a single point with a name, which are referred to as points of interest (POI). This research focuses on these points of interest places for a variety of types for example businesses, parks, schools, restaurants, and natural areas.

Building on previous chapters, this chapter explores the use of emojis with respect to place for cues about individual and collective identity. In Chapter 7 on emojis and events, it was shown that events can be detected from emoji use but cited a challenge for geo-tagged analysis of events if a baseline has not been established. This chapter addresses that gap and examines emoji use for the capital city of Washington DC for a year to identify national, local, cultural, and religious events. Using the diversity language model introduced in Chapter 4, places are mapped based on the use of diversity emojis at a particular place.

This chapter also compares the use of emojis across different types of places. Using the place name, places can be categorized in to types, for example the name George Mason University would be labelled as university. By associating emojis with places of interest, and using the place type labels, a model of emoji to place type mapping can be developed. An emoji to place type mapping is based on collective emoji use associated with places and reveals common use of emojis depicting similar experience which provide cues about the functional role of the place, as well as digital place identity.

The questions this chapter addresses are:

- What does emoji use reveal about the events and rituals associated with the geographic area?
- What does the use of diversity associated emojis in tweets reveal about diversity at places?
- What are the most commonly used emojis with respect to various types of places?

Through this chapter, there are three major contributions. One is the application of previous methodologies to demonstrate event identification and diversity mapping from emojis with respect to Washington DC. Second, the development of a place name to place type categorization which can be used to fill in geographic place datasets. And third, emoji use at places of interest reveals functional roles of places through the iconic use of emojis associated with that type of place. This chapter is organized as follows. Section 8.2 provides a background. Data and methodology are presented in Section 8.3. The results are discussed in Section 8.4. Section 8.5 concludes the chapter and presents areas for further work.

8.2 Background

While there has not yet been research on emojis and place, there has been research on sense of place from social media typically using topic modeling although more can be done (MacEachren et al., 2017). Jenkins et al., (2016) applied this approach to Twitter to identify areas of recreation, entertainment, politics and sport for New York City. Research has also been done to characterize place based on photographs and tags applied to them in Flickr (e.g., Rattenbury & Naaman, 2009). Geotagged tweets can be used to compare mobility space of individuals in a city (e.g., Schwartz & Halegoua, 2015) and to analyze how people describe various places in a city (e.g., Cranshaw et al., 2013).

There has already been some work done to depict the differences in emoji use by geography such as at the country level. For example, Ljubešić et al., (2016) showed the top most used emojis by country and found similarities for countries with similar social and economic conditions. Variations in the use of skin tone with anthropomorphic emojis was observed based on geographic regions and demographics of users (Barbieri et al., 2016). However, analysis of emoji use in social media at the micro level within a city has not yet been studied.

This research builds on this previous work but applies a new approach and lens to analyze digital language patterns in social media content to reveal our collective interactions with place. The next section introduces the methods for analyzing the structure and content of social media posts with respect to place.

8.3 Methodology

8.3.1 Data

To explore the use of emojis and place, this research used publicly available tweets collected from Twitter for Washington DC from 2014-2017. In this collection of tweets, the geographic precision varied from general geographic area to precise location that can be placed on a map. Most tweets included metadata about the precision of the location information and a place name. The level of precision of the coordinates was attributed to geographic areas such as a country, city, or neighborhood. However, some tweets contained metadata that associated the location of the tweet with a particular point of interest (POI) and included, for example, a business name. In addition, some tweets that did not have a place name but did have precise geo-coordinates for the location of the tweet. Tweets with metadata containing POI names or precise geo-coordinates, and that also contained emojis in the tweet text were used for this research. This analysis was done on 1.2 million publicly available geotagged emoji tweets between 2014-2017, that were located or tagged across 7,954 unique places of interest around the Washington DC area. Figure 8.1 shows the density of emoji tweets.

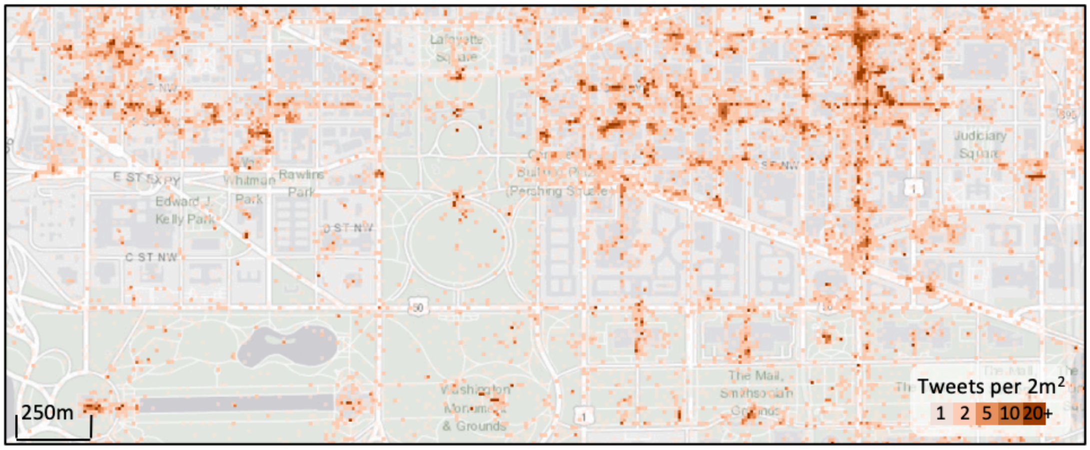


Figure 8.1 Density of public geo-tagged emoji tweets in part of Washington DC, 2014-2017.

8.3.2 Events and Place

Based on research from Chapter 7 on use of emojis at different events, event analysis based on emoji use in tweets for the city of Washington DC was conducted. This was done using temporal analysis to identify the top 3 most used emojis by date. A comparison was made for each date to identify the top three most used emojis based on tweet volume and also unique user volume. The emojis identified within the top 3 across the dates and by user volumes were further compared for their trends over time. Using the most used emoji based on percent of dates in the collection, a baseline of emojis for Washington DC was created. Events were detected when either the most or second most used emoji for a date was not the one of the emojis from the baseline.

8.3.3 Diversity and Place

Using the emojis from the diversity language model from Chapter 4, the skin tone emojis and gender emojis were used to label the skin tones and genders included in the tweets. The use of the five Fitzpatrick modifiers were used to label skin tone. Rather than mapping the location of the tweets and the skin tone emojis and gender emoji used, this analysis looked at the use of these emojis by percent of users that tagged their tweet to the same POI. Mapping the locations of these places based on the emoji use as percent of users enables comparison of broader geographic patterns across the city.

8.3.4 Place name to Type of Place

In order to identify which emojis are associated with a particular type of place, the place names that are sometimes included in tweet metadata needed to be cross-mapped to the type of place. To do this, a customized model was developed to map place name to place types and subtypes based on either keywords in the place name or direct labeling of places based on additional information from DBpedia, Yelp, or Google places. For example, the POI of Starbucks was directly labelled as a type of Eatery and subtype coffee even though the word coffee was not in the name.

For this model, a set of 18 place types and varying number of subtypes for each were identified based on similar categorizations found in DBpedia as they account for a wide variety of places and don't focus on just business and restaurants. For each subtype, keywords were identified that were most likely to be used in association with a POI that would be expected to be labeled as that subtype. In some cases, such as the Starbucks example, the names of national or regional chain business and restaurants were directly associated with a subtype. An example of the mapping of these generic keywords to place type and subtype are shown in Table 8.1. The model can be modified in future research.

Additionally, the 10% of place names that either did not contain keyword cues or were ambiguous, were assigned a place type and subtype categories based on manual review of information across a variety of sources including the Yelp, websites, and Google places. Often these POIs were unique to Washington DC, for example "White House" and "NASA" were manually assigned to subtype of "Community_govt". Using the customized place name to type model, POIs identified in the geo-tagged tweets were labeled with place type and subtype. Some places received multiple types and subtypes. For example, the Georgetown University Medical Center was labeled as both Education and Medical.

Table 8.1 Place name keyword mapping to type and subtype

Type	Subtype	Place name Keywords Generic
Animal	pet	dog park, paws, pet, animal hospital, veterinar, humane
	zoo_other	wildlife, zoo, panda house, aquarium, zoologic
Attraction	amusement_park	amusement, fairground, carousel, carnival
	monument	monument, cemetery, statue, fountain, historic, tour, memorial
	museum_art_sculpture	museum, gallery, paint, sculpture
Bar	bar	nightclub, bar, lounge, pub, brew, distil, beer, vine, wine, liquor
Community_govt	community_govt	courthouse, fire, police, mail, post office, community
Eatery	bakery	bakery, pastry, donut, cake, cookie, pie, patisserie, bagel
	coffee	coffee, café, tea, starbuck
	icecream	cream, coldstone, rita, ice, yogurt, frozen, smoothie, gelato
	fastfood	burger, sub, deli, sandwich, taco, carryout, cafeteria, take
	restaurant	restaurant, grill, cantina, steak, pizza, steak
Education	books_library	library, book store, book shop
	college	university, college, campus, dorm, quad, student
	school	elementary, academy, education, school, HS, ES, prep
Entertainment	cinema	cinema, movie, film, amc, bowtie, imax, regal
	event_facility	concert, conference, arena, armory, convention
	music	symphony, concert, ballroom, music, jazz, blues
	performing_arts	playhouse, performing, stage, comedy, theatre, auditorium
Financial	financial	bank, credit union, wells fargo, capital one
Lodging	lodging	hotel, motel, suites, lodge, inn, hostel, marriott, hilton, hyatt
Medical	medical	hosptial, clinic, dental, emergency, patient, urgent, trauma
Nature	land	trails, mountain, overlook, scenic, park, forest, nature
	water	waterfall, gorge, beach, reservoir, gorge, creek, river, pond, lake
Religious	religious	church, shrine, chapel, synagogue, temple, mosque, cathedral
Service	beauty	cut, beauty, salon, barber, hair, nail, spa, resort
	other	clean, laundr, tattoo
Shopping	shopping	store, shop, grocer, market, retail, shoe, outlet, bargain, discount
Sports_fitness	sports_fitness	recreation, bowling, baseball, sport, aquatic, gym, golf, tennis
Transit	air	afb, airline, airport, airfield, gate, terminal, baggage
	car	car, motorcycle, gas, tire, oil, fuel, automobile
	parking	transit, line, bus, trolley, tram, metro, subway
	train_bus	parking, garage, lot
	water	boat, cruise, ferry, locks
Work	work	work, business, associat, co., industries, inc, corporation, llc
Other	address	street, road, boulevard, route, highway, north, south, east, west
	area	plaza, square, circle, center, centre, ville, row, town
	residence	apartments, condo, homes, lofts, estates

8.3.5 Emojis at POIs and by Place Type

In order to generate a mapping of emojis to place type, first the emojis used at each POI were identified. For each POI, distinguished by place name and coordinates in the tweet metadata, the tweets associated with that POI were aggregated by user. The unique emojis used at that place were identified and then based on the aggregation, the percent of users at that place who used a particular emoji there were calculated. This process identifies the most common used emojis across users at a particular place. From this step is a gazetteer of place names, their geographic coordinates, and a listing of emojis and percent of users.

Using the place name to type model, the place type and subtype were then added to the gazetteer. From this listing, emoji use was then summarized across POIs of the same type and subtype. This yields a model of emojis and place type and subtypes that shows the most common used emojis across users at POIs. For this research, the descriptive analysis of the emojis used per place was the focus. However, this model could be evaluated in future research to compare to other geographic areas as well as to label the type of location possibly represented in a tweet based on emoji use.

8.4 Results and Discussion

8.4.1 Emojis and Events at Place

A baseline of 10 emojis were identified in the collection of emoji geo-tagged tweets for Washington DC, 😂 😩 😍 💯😘 😒 😭☺️🙌❤️. These were the most used emojis based on percent of users per day and greatest percent of dates when one of these emojis was among the top 3 most used emojis. In addition, ten other emojis appeared in the top 3 of most used emojis by users on particular dates. Table 8.2 shows a subset of the dates and the top 3 most used emojis in tweets, and for users on that date, including the national holiday Veteran's Day when the U.S. flag emoji was the second most used across users.

Table 8.2 Most used emojis in tweets and by user on 5 days in Washington DC.

Date	***Top 3 emojis With tweet count***	***Top 3 emojis with user count***
2014-11-09	[(😂, 490), (💯, 262), (😩, 217)]	[(😂, 160), (😍, 78), (😘, 77)]
2014-11-10	[(😂, 430), (😩, 236), (💯, 212)]	[(😂, 143), (😩, 84), (😍, 80)]
2014-11-11	[(😂, 472), (💯, 266), (😩, 265)]	[(😂, 155), (🇺🇸, 127), (😍, 95)]
2014-11-12	[(😂, 403), (😩, 232), (😘, 200)]	[(😂, 158), (😍, 89), (😘, 76)]
2014-11-13	[(😂, 473), (💯, 281), (😩, 249)]	[(😂, 155), (😩, 90), (💯, 87)]

Across the dates of the collection, there were a total of 20 emojis that ever made the top 3 most used emojis based on number of emoji users on that date. Temporal analysis of the percent of users per date for the top 20 emojis over time shows that the baseline emojis remain fairly consistent, shown in shades of gray in Figure 8.2. While the other 10 emojis show peaks in activity that correlate to timing of events acknowledged by several users in their tweets. The events included national holidays such as Independence Day and Veteran's Day with a spike in the use of the US flag emoji. Cultural holidays included as Halloween, New Year's, Valentine's Day, and St. Patrick's Day. Religious holidays included Christmas and Easter. And local events such as weather-related winter storms and the blooming of the cherry blossom along the Tidal Basin were evident based on emoji use in tweets.

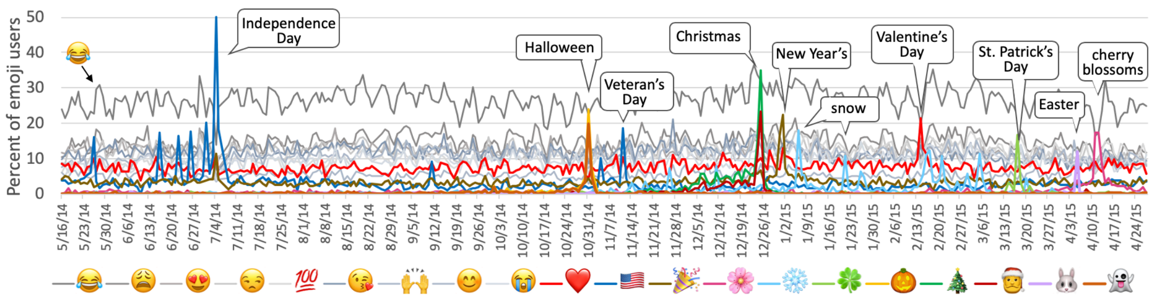


Figure 8.2 Timeline of 20 most used emojis and events in Washington D.C..

The most used emojis during the events are similar to symbols associated with the national, local, and cultural holidays identified from the analysis. However, one event unexpectedly missing from this part of the events timeline Thanksgiving. It was detected in later years because the turkey emoji was introduced mid 2015 and thus was not available during the dates of the time series shown in Figure 8.2.

The emoji tweets on event date typically did not include both the emoji associated with the event and also the event name or keywords for the emoji name. Only 40% of the emoji tweets associated with event symbols contained the event name in the same tweet on event date. For example, on Valentine's day the most used emoji in tweets was "❤️" red heart and only 2% of those tweets also included the keyword "heart" and 47% contained the keyword "Valentine" or "valentine". This indicates these symbolic emojis provide additional information associated with the event not conveyed by text alone.

Overall, the results of emojis and events at place, based on analysis for Washington DC support the findings of Chapter 7. By comparing emojis against a baseline, peaks of activity resulting from collective use of the same emoji on a particular date, may be indicative of timing for national, cultural, and religious events. This research also showed that emoji use different from the baseline can also identify some events local to a place. In addition, the use of specific emojis on the same dates for events show that emojis provide cues about the symbolism associated with a particular event.

8.4.2 Emojis and Diversity at Place

Skin tone with emojis became available in 2015. In this dataset, skin tone emojis were used by 20% of users. Mapping the most used skin tone emoji among users at a particular POI for Washington DC shows places tagged in tweets posted with this emoji, Figure 8.3. However, due to bias in data collection and demographic bias of Twitter, this data may not be representative of the population for this geographic area so broad conclusions about diversity should not be based on this level of aggregation.

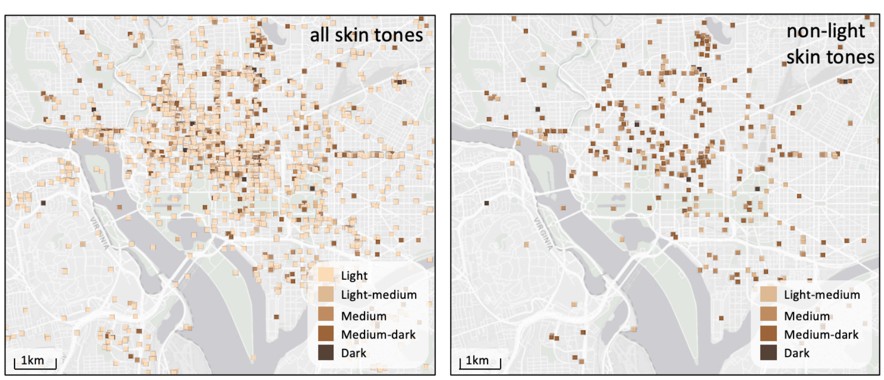


Figure 8.3 Distribution of POIs by most used skin tone emoji for Washington DC.

8.4.3 Places by Type and Subtype

For the geo-tagged emoji tweets of Washington DC that contained POI metadata such as place name and coordinates, the place name to type model was applied. The distribution of POIs with emoji tweets per type is shown in Figure 8.4. This shows that the most common POIs and that had emoji tweets were Eatery and Education. The Other category were POIs that did not fall in one of the main types. For example, emoji tweets tagged to a city name, neighborhood, or residential area such as the name of an apartment complex. Figure 8.5 shows the percent of POIs by type and then subtype for Eatery, Education, Entertainment, and Attractions. Figure 8.6 displays the geographic location of POIs by type.

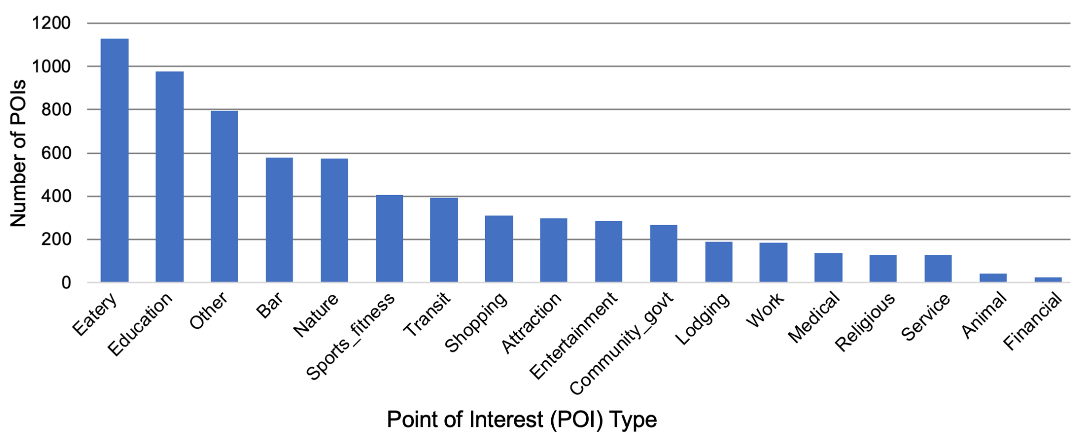


Figure 8.4 POIs by place type in emoji tweets for Washington DC, 2015-2017.

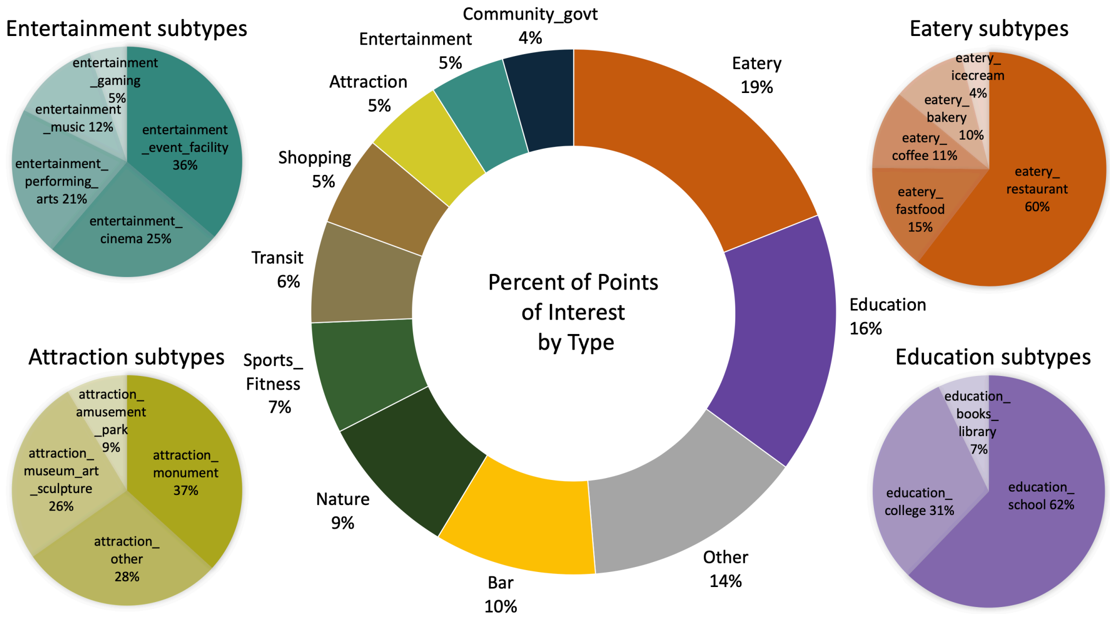


Figure 8.5 Percent of POIs by type and percent per subtype for four POI types.

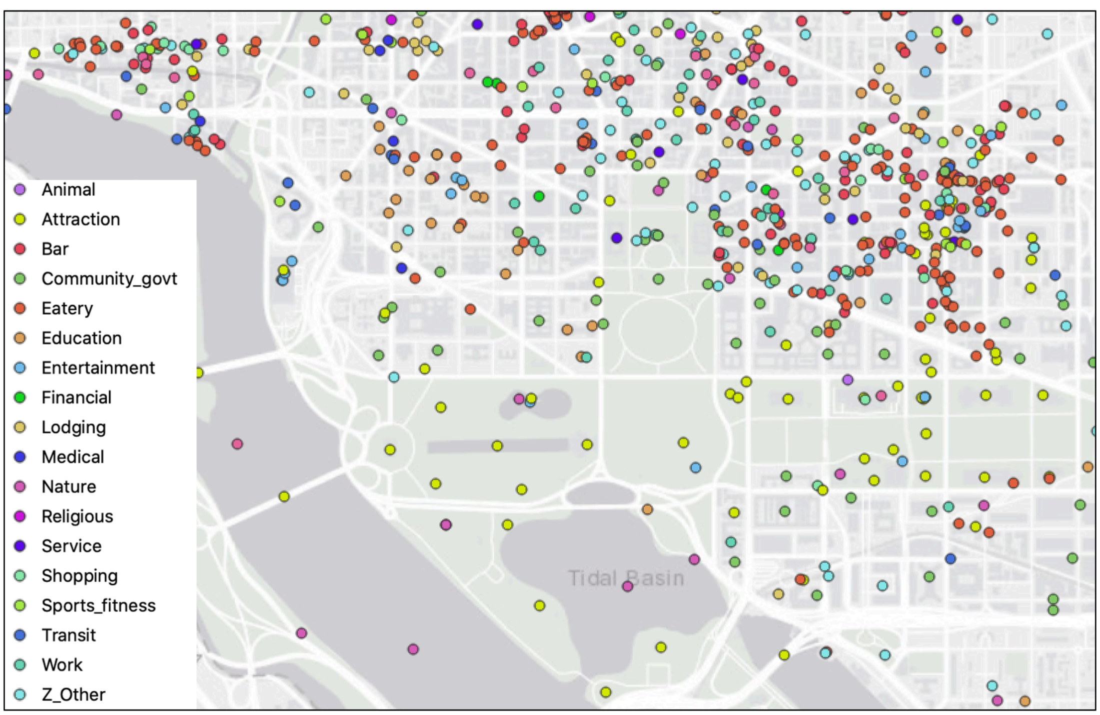


Figure 8.6 POIs by place type in Washington DC.

8.4.4 Emojis at POIs

The most used emojis based on percent of users of emojis at a particular place were analyzed. Figure 8.7 shows a map of the POIs colored based on the POI type, with a select set of POIs displayed with their name, type and subtype label, and the top most used emojis across users at that place. Figure 8.8 shows a comparison of POIs by type and the most used emoji at a particular place.

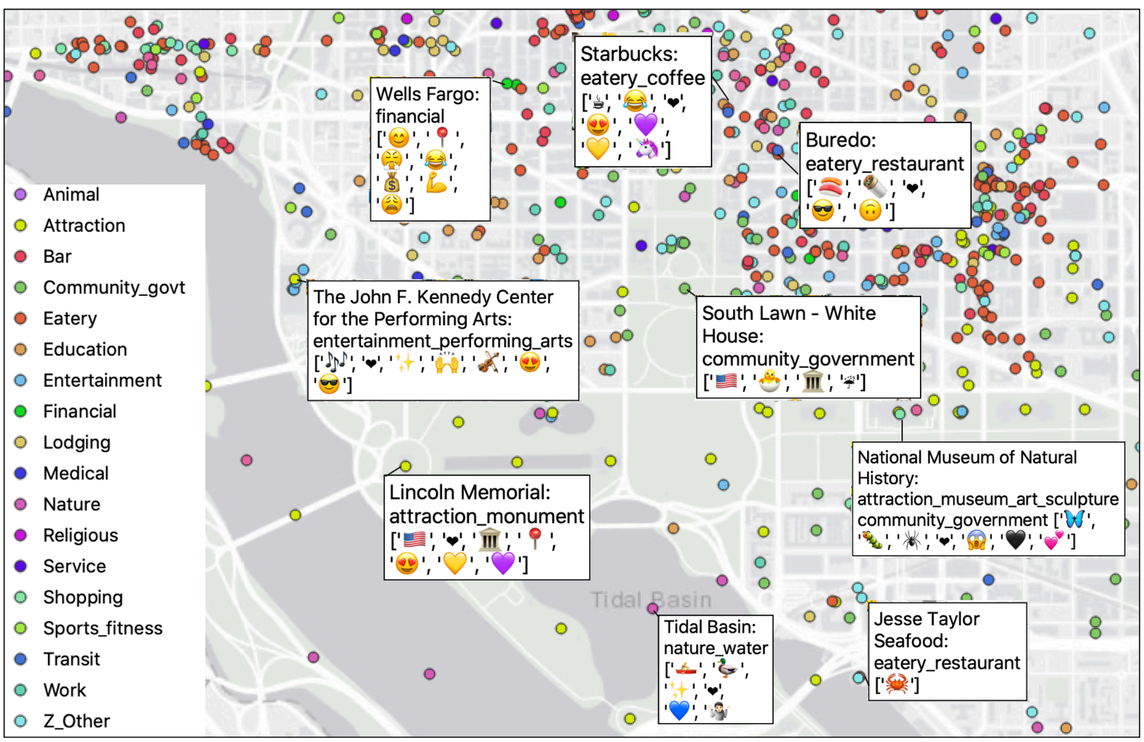


Figure 8.7 POIs shown with name, type, and emojis.

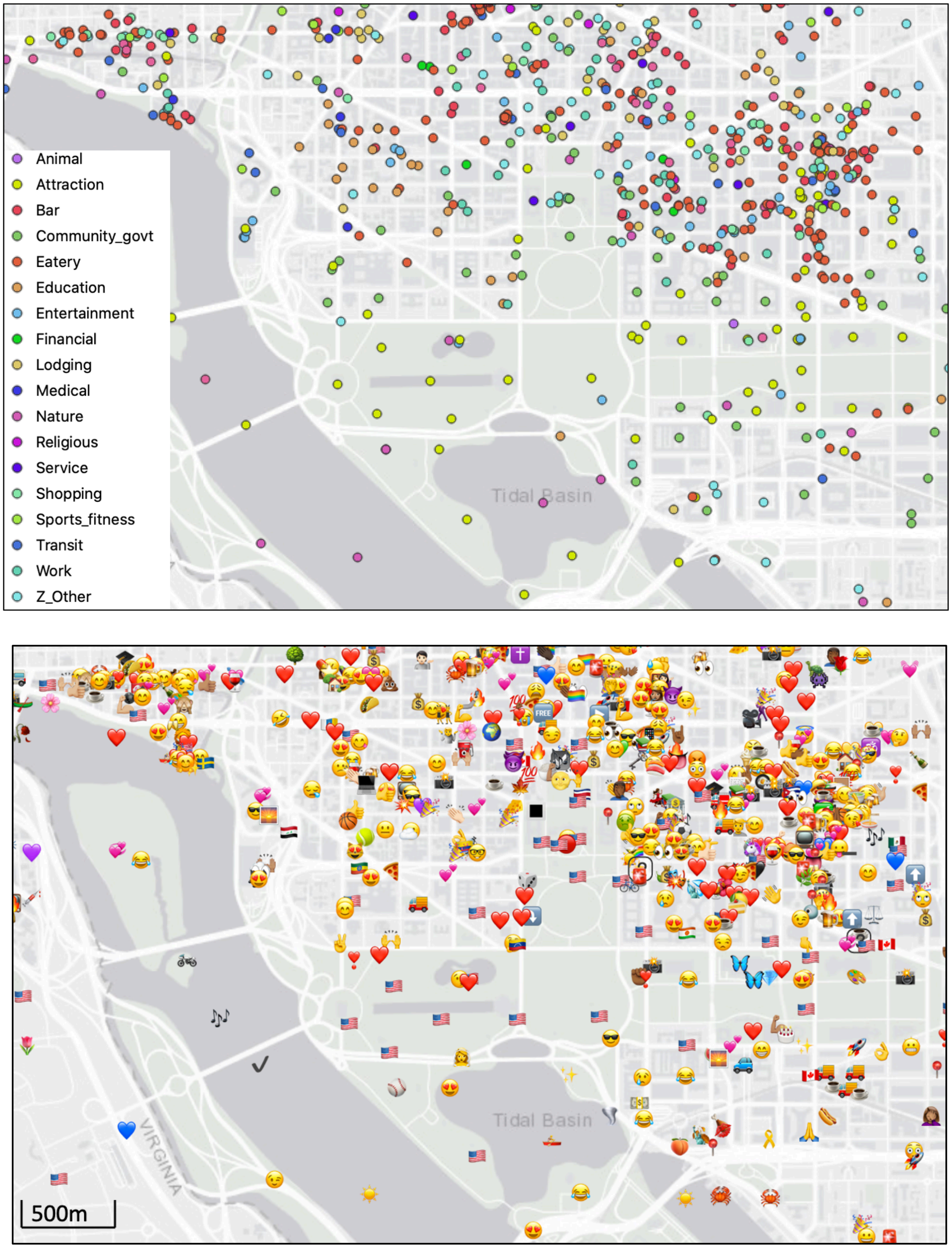


Figure 8.8 Map of POIs by type (top) and most used emoji at POIs (bottom).

8.4.5 *Emojis by Place Type and Subtype*

The most used emojis across users at POIs of the same type and subtype are shown in Table 8.3 and 8.4, respectively. The most used emojis associated with particular place types and subtypes include iconic representations indicative of the function, experience, or description of a particular place. For example, the use of emojis from the Animal Unicode sub-group at zoos, aquariums, and animal hospital. Another example is the use of pre-pared food icons for the type of eatery as well as the experience of eating 😋. And the use of icons representative of the type of service such as nail salon, hair cutting.

Not all place types had distinct or distinguishable emojis associated with that place type. For example, emoji use for places that are categorized as shopping, work, and lodging had greater overlap of emojis used with other categories. Which could indicate the lack of emojis meaningful to that type of place. Or more likely, while emoji tweets at the zoo are about the zoo, the emoji use in tweets from a place labeled as work or other are from a place but not necessarily about the place.

Table 8.3 Top 15 most used emojis across POIs by type

POI Type	Top 15 emojis
Animal	[🐼, 🐶, 🌻, 🐘, 🐒, 🦁, 🐯, 🦍, 🐠, 🐬, 🐱, 🐾, 🐅, 🐢, 💗]
Attraction	[📸, 🎡, ✊🏾, 😉, 🏛, 🎨, 🚀, ✊🏽, 🤔, ✈️, 🌻, 🌈, 💓, 💖, 🌞]
Bar	[🍻, 🍷, 🍺, 🍾, 🍹, 😋, 🍸, 😉, 🍔, 🥂, 🥃, 🏈, 💋, 🌈, 🎤]
Community_govt	[⚖️, 😧, 🚒,🏛,🤔,🚓, 🔴, 😱,👮‍♀️]
Eatery	[😋, ☕, 🤤, 🤔, 🍕, 🍴, 🦀, 🌮, 🍦, 🍩, 🍺, 🍹, 💖, 🍷, ☺️]
Education	[🎓, 📚, 🏈, 🏀, ‼, ⚽, 🤔, 💪🏾, 🏆, ☺️, 😉, 🐢, 👑, 🔴]
Entertainment	[🏀, 🎸, 🙌🏼, 🎤, 🙌, 🤘, 💖, 👏, 🙌🏾, 💥, ⚡, 😛, ‼, 💗]
Financial	[🤘🏻, 🤘, 🐐, 👍, 💰, 😝]
Lodging	[👍, 💋, 💗, 👏, 🌈, 💰, 😌, 📸, 😉, 😄, 😴, 🍾, 🍻, 💍, 🙌🏽]
Medical	[😒, 🙏, 🏥, 👶🏾, 🤕, 😌, 😷, 😑, 🙃, 💉, 🤑, 🙏🏽, 😢, 💪, 😔]
Nature	[🌞, 🌻, 👍, 🐶, ⚽, 😋, 🍃, 😉, 💪, 🍂, ☺️, 🌿]
Religious	[🙌🏾, 🙏🏾, 😇, 🙏🏽, 🙌, ⛪, 🤔, 🎹, 🙏🏼, 🙏, 🙌🏽, 💡, ✊🏾, 🤙🏽]
Service	[💈, 💅🏻, 💁🏻, 💇🏻, 🙌🏾, 💅🏻, ✂️, ☺️, 😅, 💋, 💇, 💉, 😝, 👑, 🍾]
Shopping	[🤔, 😋, 😌, 😏, 💰, 😅, ☺️, 📸, 😒, ☕, 💸, 🤤, 👍, 🍎]
Sports_fitness	[🏈, ⚽, ⚾,💪🏾, 💪, ‼, ⛳, 🏀, 🔴, 🏆, 🤔, 🔵, 💪🏽, 💥,🎾]
Transit	[✈️, 🛬, 🤔, 🙃, ✌🏾, 😏, 😳, ‼, 😉, 🛩, 👍, 🛬, ✈️, 😅, 😢]
Work	[🤓, ☕, 💰, 🏳️‍🌈, 🤔, ☺️, 📺, 👋, 💁🏼, 🚨, 😋, 🎈, 🙌🏾, 👉, 👍🏽]
Other	[💰, 🤔, 😋, 💋, 😌, 🌈, 🗣, 🙃, 🤷🏾‍♂️, 💪🏽, 💀, 💉, 😜, 📸, ‼]

Table 8.4 Top 15 most used emojis across POIs by subtype

POI Type	POI Subtype	Top 15 emojis per POI subtype
Animal	pet	[🐶, 🐾, 🐱, 🐴, 💓, 💗]
	zoo_other	[🐼, 🌻, 🐘, 🐒, 🦁, 🐯, 🦍, 🐠, 🐬, 🐅, 🐢, ☺️, 🐟, 🐻, 🦋]
Attraction	amusement_park	[🎡, 🎢, 🎠, 🐐, 🎃, 🎸, 🙌🏽, 🚜, 🎵, 🔫, 🥁, 💦, 🎤, 🐢, 🍗]
	monument	[🏛, 📸, 🌞, 🏳️‍🌈, 🌈, 🙏🏽, 🌎, 🏈, 🎓, 💓, 🌹, 🌅, ⚓, ✊🏽]
	museum_art_sculpture	[✊🏾, 📸, 🎨, 🚀, 🤔, 🌻, 💫, ✈️, 😉, 🙌🏾, ✊🏽, 👏, 🦋, 👌, 🌈]
	other	[🏀, 😉, 💰, 🎡, ❤️, 💖, 🤑, 👑, 🎨, 💞, 🍻, ☺️, 🤣, 📷, 😏]
Bar	bar	[🍻, 🍷, 🍺, 🍾, 🍹, 😋, 🍸, 😉, 🍔, 🥂, 🥃, 🏈, 💋, 🌈, 🎤]
Community_govt	community_govt	[⚖️, 😟, 🚒, 🏛, 🤔, 🚓, 🔴, 😱, 👮]
Eatery	bakery	[😋, 🍩, ☕, 🍰, 🤤, 🎂, 🍭, 🍴, 😉, 🤔, 🌈, 🌸, 😛, 🤓, 🎸]
	coffee	[☕, 😋, 📚, 🤤, 📸, ☕, 🍺, 😜, 🎤, 🙃, 💞, 🐶, 🌸, 🙏🏽, ☺️]
	fastfood	[😋, 🌮, 🍔, 😡, 🤔, ☺️, 🍟, 🙃, 🙌🏽, 👍, 🌯, 🍳, 💪🏾, 🤤, 💖]
	icecream	[🍦, 😋, 🍧, 😛, ☕, 🐾, 🏝, 🎃, 🍨, 🙌🏻]
	restaurant	[😋, 🤤, 🦀, 🍕, 🍴, 🤔, 🥂, 🍻, 🍹, 🍺, 🙌🏾, 🍣, 🍽, 🌹, 💖]
Education	books_library	[📚, 📖, 🌼, 📷, 🤔, 😔, 💻, 🙌🏽, 🍁, 🌍, 🏈, 😜, 🙃, 😉, 🌎]
	college	[🎓, ‼️, 📚, 🐢, 📸, 💪🏾, 👩🏽‍🎓, 🏈, 🤔, 🏀, 👑, 🗣, ⚽, 🚨, 💰, 🙏🏾]
	school	[🏈, 🏀, 📚, 🎓, ⚽, ‼️, 🔴, 🏆, ⚪, 😉, 🔵, 👏, 💀, ☺️, 🐾]
Entertainment	cinema	[🙌🏾, 🤣, 🎬, 🎥, 🎸, 😋, 🍿, 👏, 👻, ❤️, 💋, 📽, 🙏, 🎤, 😉]
	event_facility	[🏀, 🙌, 🎤, ⚡, 💥, 🙌🏽, 👑, 🙌🏼, 🤘, 💪, 🤓, 💰, 💖]
	gaming	[💰, 👭, 👍, 😜, 💸, 💵, 🍕, 💱, 👱🏼, 💴, 💪🏾, 🍅, 👅, 🍾, ⌚]
	music	[🎸, 💖, 🙌🏼, 🤘, 😛, 🎵, 🎤, 💗, ❤️, 💥, 🤘🏼, 🤘🏻, 🎼, 🐌, 💘]
	performing_arts	[🎻, 😉, 😏, 💪🏾, ❄️, 👏, 🤤, 😋, 🎃, 🎨, 💰, 🙌, ‼️, 🙏🏾, 😝]
Lodging	lodging	[👍, 💋, 💗, 👏, 🌈, 💰, 😌, 📸, 😉, 😄, 😴, 🍾, 🍻, 💍, 🙌🏽]
Medical	medical	[😒, 🙏, 🏥, 👶🏾, 🤕, 😌, 😷, 😑, 🙃, 💉, 🤑, 🙏🏽, 😢, 💪, 😔]
Nature	land	[🌻, 👍, 🐶, ⚽, 😋, 😉, 🍂, 🤔, 📸, ☔, 🍃, 😜, ⛰]
	water	[🌞, 💪, 🌊, ☺️, 🦆, 🍃, ❤️, 🛶, 🍻, 🐶, 🤘, 😛, 🌿, 🤣]
Religious	religious	[🙌🏾, 🙏🏾, 😇, 🙏🏽, 🙌, ⛪, 🤔, 🎹, 🙏🏼, 🙏, 🙌🏽, 💡, ✊🏾, 🤙🏽, 🏀]
Service	beauty	[💈, 💅🏻, 💁🏻, 💇🏻, 💅🏻, ✂️, ☺️, 💋, 💇, 🙌🏾, 🍾, 💆🏽, 🙌🏽, 💆🏻, 😅]
	other	[💉, 🗣, 😛, 😏, ❗, 🕸, 🌐, 🕰, 👑, 😴, 🐄, 🖕, 💳, ⌚]
Shopping	shopping	[🤔, 😋, 😌, 😏, 💰, 😅, ☺️, 📸, 😒, ☕, 💸, 💗, 🤤, 👍, 🍎]
Sports_fitness	sports_fitness	[🏈, ⚽, ⚾, 💪🏾, 💪, ‼️, ⛳, 🏀, 🔴, 🏆, 🤔, 🔵, 💪🏽, 💥, 🎾]
Transit	air	[✈️, 🛬, ✌🏾, 🛩, 🛬, ✈️, 🙃, 😳, 😉, 👍, 😅, 🤔, ‼️, 💔, 😢]
	car	[🚗, 🤔, 😐, 😡, 🎬, 💪🏽, 👌, ‼️, 💞, 🎈, 😌, 🐐, 🌸, ☺️, 😛]
	parking	[😶, ⛽, 💪🏾, 💪🏿, 🤣, 😅, 🏗, ☹️, 🔒, 🤘🏾, 👨🏾‍🎓, 🤙🏿]
	train_bus	[😏, 🤔, 🚂, 💰, 🙃, 🗣, 💨, 😋, 😉, 💀, 😑, 😌, ☺️, 🤷🏽‍♂️, 📸]
	water	[🌻, ⛰, 🌊, ❄️, 🚢, 🍓, 🍻, 🚣🏼, 😌, 🤙🏼, ✔️, ✌🏼, 😋, 🍧, 🛳]
Work	work	[🤓, ☕, 💰, 🏳️‍🌈, 🤔, ☺️, 📺, 👋, 💁🏼‍♂️, 🚨, 😋, 🎈, 🙌🏾, 👉, 👍🏽]
Other	address	[🤣, 😜, 💍, 🙃, 💔, 💰, 💪🏽, 🗣, 😳, 🤦🏾‍♀️, 🤘🏽, 🎾, 💸, 💉, 🤷🏾‍♂️]
	area	[💰, 💋, 😋, 🤷🏾‍♂️, 🤔, 😌, 📸, 🌈, 👍, ❤️, 💉, 💪🏽, 😴, 🗣, 💥]
	residence	[💰, 🏳️‍🌈, 🌈, 🤔, 😌, 😡, 💦, 🌞, 🗣, ❄️, 🚨, 😋, 🤣, 🤷🏽‍♀️, 💋]

8.5 Conclusion

Emoji use in social media provides cues about individual identity, collective behavior, events, and now place. This chapter demonstrated that emojis at a particular place reveal geographic areas and diversity of users, and symbolism of emojis associated with timing of particular events. In addition, emoji use varied based on the type of place with iconic emojis being used collectively to indicate experience with or description of the function of a place. A limitation of this research is that it explored emoji use in only geo-tagged tweets for one city. Future research should compare the results to other cities. In addition, the place name to type model and the emoji to place type model presented in this chapter can be expanded upon or modified in order to tune them to other geographic areas and dataset collections. In addition, a comparison of emojis used and place in other social media platforms may yield additional differences in emoji use with respect to place. Emoji use in social media provides much more information than what is currently focused on in academic research and social media analysis and this is the first study to explore the relationship of emojis and place at a city level.

9 CONCLUSION

9.1 Summary of Dissertation Results

This chapter draws across the methodologies and results presented in the previous chapters to summarize and to show how together they address the four research objectives and three research questions introduced in Chapter 1. First, I summarize how each of the four research objectives have been addressed. For the objective of establishing a baseline of emoji use in a social media data sample, Chapter 3 analyzed differences in individual emoji use and by document type for Twitter. It showed that users typically only used a handful of different emojis but coming from only one or two subcategories, although emoji super users use significantly more. In Chapter 3 and 5, I address the second research objective to develop a framework to enable comparison of emoji use and structure in documents. Results of Chapter 3 showed the differences in which emojis are used based on document type in terms of categories and sub-categories and Chapter 5 presents a methodology to analyze the structure of emojis in documents by describing their attributes to also include color, type, direction and by considering the position, order, and repetition of emojis in a document. Research objective three, identify and describe emoji use relating to individual or collective identities in social media, was addressed in Chapters 4 and 5. The diversity language model was presented in Chapter 4 and shows how emoji use can be used to analyze user social identity associated with diversity characteristics such as race, gender, sexual orientation, religion, skin tone, and political ideology. Chapter 5 showed differences in structural emoji use based on user roles for some user groups such as bots, marketers, journalists, students and more. To address research objective four, identify communication patterns and emojis related to particular events and types of places, emoji use was compared across users during a lunar eclipse in Chapter 6 and for events in Chapter 7 and place in Chapter 8.

Now I turn to the three research questions and summarize the results in the context of each. Research question one asked: What are the differences in emoji use across users and documents in social media? With a focus on Twitter, in Chapter 3 I identified that most users sending non-retweets were more likely to use a handful of anthropomorphic emojis coming from the same sub-categories. Chapter 5 confirmed this showing high level of consistency of which unique emojis are used in non-retweets. The biggest differences in emoji use emerged across users based on document type. In Chapter 3, I showed that difference in the categories of emojis used per document type. Also, that emoji use in a user name and profile description did not necessarily correlate to emoji use in tweets, and vice versa and that there appeared to be different purposes for emoji use based on document type. Chapter 4 confirmed this by showing that emojis in a user profile description are more likely to relate to social identity of the user.

For research question two, it asked: In what way do emojis reveal cues about social identity and individual communication style preferences? Drawing on insights from Chapter 4 emojis use in user profile descriptions can reveal the diversity characteristics associated with the online identity of the user. The communication style of users was analyzed in Chapter 5 and revealed that users can be clustered based on similarity of the behavior of consistency of which emojis are used and the way they are used in a tweet.

Now I summarize the results of research question three: What are the collective patterns and behaviors that arise from individual emoji use and what do they reveal about social norms? Despite the individual differences in emoji use, Chapter 5 showed that some users shared similar styles and behaviors of emoji use which could be identified by clustering users. This revealed that for example, news organizations and retired military veterans had the greatest consistency of structure of emoji use in their tweets. Further, using the specific document structure it is possible to identify users that are possibly related such as identifying additional bot accounts. In addition, similar use of emojis occurred based on type of documents. For example, in Chapter 3 it was shown that tweets which were more likely to be retweeted contained emojis from the group of symbols while users sending mostly non-retweets had greater user of face and gesture emojis. Chapter 4 shows an example of social norms at play for use of emojis in the user profile descriptions as indicators of social identity. There is also collective behavior of emoji use during an events as users include the emoji in their tweet as icon representations of what they see, such as during an eclipse, e.g., Chapter 6, or as symbolism to indicate that solidarity or acknowledge an event such as protests, national holidays, cultural and religion events, Chapter 7. Lastly, in Chapter 8, emoji use collectively across users at particular places were description of the function of place revealing when users are engaging with their environment rather than just tweeting from place.

9.2 Limitations and Future Work

The focus of this research has been on the use of emojis in social media, with an emphasis on Twitter. Future research could expand the scope to study emoji use on other social media platforms such as Facebook, Instagram, Gab, Parler, and Reddit. A comparative analysis of individuals posting on multiple social media accounts on a single or on multiple platforms can be used to investigate individual differences in emoji use associated with each account and platform. As the tweet data used as the basis for much of this analysis being the same dataset which was collected in 2018 and related to American politics, additional datasets can be used to compare the robustness of the findings and results of this dissertation. Similarly, comparison across tweet collections for other topic domains may reveal additional insights about individual and social behavior with respect to sports, tourism, current events, and other topics of interest.

In addition, Twitter is inherently biased in terms of demographics of users and the amount of data available via the API (Morstatter & Liu, 2017) and I made assumptions of user roles based on keywords in the user profile descriptions which could be improved by using other datasets with labeled and validated user roles. Using the methodologies presented in this dissertation, the results can be combined with results of other analysis, such as text mining, user activity metrics, geo-temporal analysis, and networks of interactions for a more complete social media analysis on order to merge the structure of emoji use, text modeling, and the interactions of users. Finally, this analysis represents a snapshot in time even though the timeframe of the dataset collection and the emojis available at that timeframe were kept in mind. Just as the emoji "Face with Tears of Joy" was the Oxford word of the year in 2015 and quickly became the most popular emoji used on Twitter, there could always be another emoji that takes its place or enters the top most used emojis. With the availability of new emojis, temporal analysis of emoji trends and the factors contributing to the prevalence of certain emojis in social media are opportunity for future research. While this research focused on identifying the patterns of emojis used, future research could also focus on polling and interview of emoji users to explain why they use certain emoji during specific events. This research provides examples and also resources such as code posted to GitHub at https://github.com/msemoji to enable other researchers to compare and analyze emojis. This dissertation contributes to both social sciences and communication studies by demonstrating emojis are important to consider in social media research as they provide cues about individual identity and collective behavior. More than just cute faces and indicators of sentiment, emojis in computer mediated communication provide a valuable lens with which to examine social processes of both in-person and online communities as people interact across a variety of groups, events, and places.

REFERENCES

Ai, W., Lu, X., Liu, X., Wang, N., Huang, G., & Mei, Q. (2017, May 15). Untangling Emoji Popularity Through Semantic Embeddings. *Proceedings of the Eleventh International Conference on Web and Social Media*. ICWSM 2017, Montreal QC, Canada. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/viewPaper/15705

Anderson, B. R. O. (2006). *Imagined communities: Reflections on the origin and spread of nationalism* (Rev. ed). Verso.

Andrews, H. (2013). *Events and The Social Sciences* (1st ed.). Routledge. https://doi.org/10.4324/9780203070741

Arora, P. (1986). Hindu Festivals of India. *The Journal of Popular Culture*, *20*(2), 175–182. https://doi.org/10.1111/j.0022-3840.1986.2002_175.x

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, *124*, 150–159. https://doi.org/10.1016/j.paid.2017.12.018

Badawy, A., Lerman, K., & Ferrara, E. (2019). Who Falls for Online Political Manipulation? *Companion Proceedings of The 2019 World Wide Web Conference on - WWW '19*, 162–168. https://doi.org/10.1145/3308560.3316494

Balasuriya, L., Wijeratne, S., Doran, D., & Sheth, A. (2016). Finding street gang members on Twitter. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 685–692. https://doi.org/10.1109/ASONAM.2016.7752311

Barach, E., Srinivasan, V., Fernandes, R., Feldman, L. B., & Shaikh, S. (2020). It's not Just What you Tweet, it's how you Tweet It. *7th European Conference on Social Media*, 52–59.

Barbera, P. (2016). *Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data.* New York University. https://pdfs.semanticscholar.org/7626/4abe68ea6de6c14944990c5e01da1725c860.pdf

Barberá, P., & Zeitzoff, T. (2018). The New Public Address System: Why Do World Leaders Adopt Social Media? *International Studies Quarterly*, *62*(1), 121–130. https://doi.org/10.1093/isq/sqx047

Barbieri, F., Ballesteros, M., Ronzano, F., & Saggion, H. (2018). Multimodal Emoji Prediction. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 679–686. https://doi.org/10.18653/v1/N18-2107

Barbieri, F., & Camacho-Collados, J. (2018). How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter. *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, 101–106. https://doi.org/10.18653/v1/S18-2011

Barbieri, F., Kruszewski, G., Ronzano, F., & Saggion, H. (2016). How Cosmopolitan Are Emojis?: Exploring Emojis Usage and Meaning over Different Languages with Distributional Semantics. *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*, 531–535. https://doi.org/10.1145/2964284.2967278

Baym, N. (1998). The emergence of on-line community. In S. Jones (Ed.), *Cybersociety 2.0: Revisiting Computer-Mediated Communication and Community*. New Media Cultures.

Berea, A. (Ed.). (2019). *A Complex Systems Perspective of Communication from Cells to Societies*. IntechOpen. https://doi.org/10.5772/intechopen.77961

Blagdon, J. (2013, March 4). *How emoji conquered the world*. https://www.theverge.com/2013/3/4/3966140/how-emoji-conquered-the-world

Blank, G., & Lutz, C. (2017). Representativeness of Social Media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, *61*(7), 741–756. https://doi.org/10.1177/0002764217717559

Blumer, H. (1962). Society as symbolic interaction. In A. Rose (Ed.), *Human Behavior and Social Process: An Interactionist Approach*. Houghton-Mifflin.

Bode, L., & Dalrymple, K. E. (2016). Politics in 140 Characters or Less: Campaign Communication, Network Interaction, and Political Participation on Twitter. *Journal of Political Marketing*, *15*(4), 311–332. https://doi.org/10.1080/15377857.2014.959686

Bremmer, J., & Roodenburg, H. (1992). *A cultural history of gesture* (Vol. 11). Cornell University Press.

Burden, A. (2016). *Photograph of single red tulip*. https://unsplash.com/photos/5JHRlfQzm-A

Burger, A., Oz, T., Crooks, A., & Kennedy, W. G. (2017). Generation of Realistic Mega-City Populations and Social Networks for Agent-Based Modeling. *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas on - CSS 2017*, 1–7. https://doi.org/10.1145/3145574.3145593

Burger, A., Oz, T., Kennedy, W. G., & Crooks, A. T. (2019). Computational Social Science of Disasters: Opportunities and Challenges. *Future Internet*, *11*(5), 103.

Cappallo, S., Svetlichnaya, S., Garrigues, P., Mensink, T., & Snoek, C. (2018). The New Modality: Emoji Challenges in Prediction, Anticipation, and Retrieval. *IEEE Transactions on Multimedia*. https://doi.org/10.1109/TMM.2018.2862363

Chakraborty, A., Messias, J., Benevenuto, F., Ghosh, S., Ganguly, N., & Gummadi, K. P. (2017). Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations. *Proceedings of the Eleventh International Conference on Web and Social Media*, 22–31. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15680/14791

Chen, Z., Lu, X., Ai, W., Li, H., Mei, Q., & Liu, X. (2018a). Through a Gender Lens: Learning Usage Patterns of Emojis from Large-Scale Android Users. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 763–772. https://doi.org/10.1145/3178876.3186157

Chen, Z., Lu, X., Ai, W., Li, H., Mei, Q., & Liu, X. (2018b). Through a Gender Lens: Learning Usage Patterns of Emojis from Large-Scale Android Users. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 763–772. https://doi.org/10.1145/3178876.3186157

Chierichetti, F., Kleinberg, J., Kumar, R., Mahdian, M., & Pandey, S. (2014). Event Detection via Communication Pattern Analysis. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 51–60. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8088

Cioffi-Revilla, C. A. (2014). *Introduction to computational social science: Principles and applications*. Springer.

Clarkson, K., Srivastava, G., Meawad, F., & Dwivedi, A. D. (2019). Where's @Waldo?: Finding Users on Twitter. In L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, & J. M. Zurada (Eds.), *Artificial Intelligence and Soft Computing* (pp. 338–349). Springer International Publishing. https://doi.org/10.1007/978-3-030-20915-5_31

Comunello, F., Mulargia, S., Polidoro, P., Casarotti, E., & Lauciani, V. (2015, May 24). No Misunderstandings During Earthquakes: Elaborating and Testing a Standardized Tweet Structure for Automatic Earthquake Detection Information. *12th Proceedings of the International Conference on Information Systems for Crisis Response and Management*. ISCRAM, Krystiansand, Norway. http://idl.iscram.org/files/francescacomunello/2015/1232_FrancescaComunello_etal2015.pdf

Cramer, H., de Juan, P., & Tetreault, J. (2016). Sender-intended functions of emojis in US messaging. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 504–509. https://doi.org/10.1145/2935334.2935370

Crespi-Vallbona, M., & Richards, G. (2007). THE MEANING OF CULTURAL FESTIVALS: Stakeholder perspectives in Catalunya. *International Journal of Cultural Policy*, *13*(1), 103–122. https://doi.org/10.1080/10286630701201830

Croitoru, A., Crooks, A., Radzikowski, J., & Stefanidis, A. (2013). Geosocial gauge: A system prototype for knowledge discovery from social media. *International Journal of Geographical Information Science*, *27*(12), 2483–2508. https://doi.org/10.1080/13658816.2013.825724

Croitoru, A., Crooks, A., & Stefanidis, A. (2017). Geovisualization of Social Media. In D. Richardson, N. Castree, M. Goodchild, A. Kobayashi, W. Liu, & R. Marston (Eds.), *The International Encylopedia of Geography: People, the Earth, Environment, and Technology*. Wiley Blackwell.

Crompton, J. L., & McKay, S. L. (1997). Motives of visitors attending festival events. *Annals of Tourism Research*, *24*(2), 425–439. https://doi.org/10.1016/S0160-7383(97)80010-2

Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). #Earthquake: Twitter as a Distributed Sensor System: #Earthquake: Twitter as a Distributed Sensor System. *Transactions in GIS*, *17*(1), 124–147. https://doi.org/10.1111/j.1467-9671.2012.01359.x

Crooks, A., Pfoser, D., Jenkins, A., Croitoru, A., Stefanidis, A., Smith, D., Karagiorgou, S., Efentakis, A., & Lamprianidis, G. (2015). Crowdsourcing urban form and function. *International Journal of Geographical Information Science*, *29*(5), 720–741. https://doi.org/10.1080/13658816.2014.977905

Crooks, A. T., Croitoru, A., Jenkins, A., Mahabir, R., Agouris, P., & Stefanidis, A. (2016). User-Generated Big Data and Urban Morphology. *Built Environment*, *42*(3), 396–414.

Culliford, E. (2019). *Twitter makes global changes to comply with privacy laws.* https://www.reuters.com/article/us-twitter-privacy/twitter-makes-global-changes-to-comply-with-privacy-laws-idUSKBN1Y622J

Danescu-Niculescu-Mizil, C., Gamon, M., & Dumais, S. (2011). Mark my words!: Linguistic style accommodation in social media. *Proceedings of the 20th International Conference on World Wide Web - WWW '11*, 745. https://doi.org/10.1145/1963405.1963509

Davidson, K., & Blair, J. (2018). Semiotic Analysis of the Raised Fist Emoji As a Sign of Resilience: In *Semiotics* (pp. 31–45). Semiotic Society of America. https://doi.org/10.5840/cpsem20183

Davis, M. (2018). *Unicode Emoji* (Unicode Technical Standard # 51). unicode.org.

Davis, M. (2020, February). *Full Emoji List, v13.0*. https://unicode.org/emoji/charts/full-emoji-list.html

Davis, M., & Edberg, P. (2019, December). *Full Emoji List, v12.1*. Unicode Emoji Charts. https://unicode.org/emoji/charts/full-emoji-list.html

De Bres, K., & Davis, J. (2001). Celebrating group and place identity: A case study of a new regional festival. *Tourism Geographies*, *3*(3), 326–337. https://doi.org/10.1080/14616680110055439

Deb, A., Luceri, L., Badaway, A., & Ferrara, E. (2019). Perils and Challenges of Social Media and Election Manipulation Analysis: The 2018 US Midterms. *Companion Proceedings of The 2019 World Wide Web Conference on - WWW '19*, 237–247. https://doi.org/10.1145/3308560.3316486

Desai, N. (2010). *A different freedom: Kite flying in Western India : culture and traditions*. Cambridge Scholars. http://public.eblib.com/choice/publicfullrecord.aspx?p=1165472

Doherty, C., Kiley, J., & O'Hea, O. (n.d.). *Wide gender gap, growing educational divide in voters' party identification*. Pew Research Center.

Doherty, C., Kiley, J., & O'Hea, O. (2018). *Wide gender gap, growing educational divide in voters' party identification.* Pew Research Center. https://www.people-press.org/2018/03/20/wide-gender-gap-growing-educational-divide-in-voters-party-identification/

Donato, G., & Paggio, P. (2017). Investigating Redundancy in Emoji Use: Study on a Twitter Based Corpus. *Proceedings of the 8th Workshop on Computational*

Approaches to Subjectivity, Sentiment and Social Media Analysis, 118–126. https://doi.org/10.18653/v1/W17-5216
Draper, S. (2014). Effervescence and Solidarity in Religious Organizations: EFFERVESCENCE AND SOLIDARITY. *Journal for the Scientific Study of Religion*, *53*(2), 229–248. https://doi.org/10.1111/jssr.12109
Drury, J. (Ed.). (2013). *Crowds in the 21st century: Perspectives from contemporary social science*. Routledge [u.a.].
Duggan, M., & Brenner, J. (2013). *The Demographics of Social Media Users—2012* (No. 14). Pew Research Center. https://www.pewresearch.org/internet/2013/02/14/the-demographics-of-social-media-users-2012/
Durkheim, É. (2005). The Dualism of Human Nature and its Social Conditions. *Durkheimian Studies*, *11*(1). https://doi.org/10.3167/175223005783472211
Edensor, T. (2002). *National identity, popular culture and everyday life*. Berg Publishers.
Eisner, B., Rocktäschel, T., Augenstein, I., Bosnjak, M., & Riedel, S. (2016). emoji2vec: Learning Emoji Representations from their Description. *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, 48–54. https://doi.org/10.18653/v1/W16-6208
Elgenius, G. (2011). *Symbols of nations and nationalism: Celebrating nationhood*. Palgrave Macmillan.
Ellis, N. C., Larsen-Freeman, D., & Research Club in Language Learning (Ann Arbor, Mich.) (Eds.). (2009). *Language as a complex adaptive system*. Wiley-Blackwell.
Emojipedia. (n.d.). https://emojipedia.org/
Esman, M. R. (1982). Festivals, Change, and Unity: The Celebration of Ethnic Identity among Louisiana Cajuns. *Anthropological Quarterly*, *55*(4), 199. https://doi.org/10.2307/3317149
Espenak, F. (2009, April 29). *Total Lunar Eclipse of 2019 Jan 21*. Eclipse.Gsfc.Nasa.Gov. https://eclipse.gsfc.nasa.gov/LEplot/LEplot2001/LE2019Jan21T.pdf
Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., & Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing - EMNLP*, 1615–1625. https://doi.org/10.18653/v1/D17-1169
Fishbane, S. (2016). Behind the Purim mask: The symbolic representation of the rituals and customs of Purim. In S. Fishbane & E. Levine (Eds.), *Contention, controversy, and change—Evolutions and revolutions in the Jewish experience* (Vol. 2, pp. 135–205). Touro College Press.
Flood, G. D. (1996). *An introduction to Hinduism*. Cambridge University Press.
Frost, W., & Laing, J. (2013). *Commemorative events: Memory, identities, conflict* (1st ed). Routledge.
Gawne, L., & McCulloch, G. (2019). Emoji as digital gestures. *Language@ Internet*, *17*(2). http://nbn-resolving.de/urn:nbn:de:0009-7-48882
Gazaz, H., Croitoru, A., Delamater, P. L., & Pfoser, D. (2016). Geo-fingerprinting social media content. *Proceedings of the Third International ACM SIGMOD Workshop*

on Managing and Mining Enriched Geo-Spatial Data - GeoRich '16, 1–6. https://doi.org/10.1145/2948649.2948654

Ge, J. (2019). Emoji Sequence Use in Enacting Personal Identity. *Companion Proceedings of The 2019 World Wide Web Conference on - WWW '19*, 426–438. https://doi.org/10.1145/3308560.3316545

Geber, S., & Hefner, D. (2019). Social norms as communicative phenomena: A communication perspective on the theory of normative social behavior. *Studies in Communication | Media*, *8*(1), 6–28. https://doi.org/10.5771/2192-4007-2019-1-6

Geertz, C., & Darnton, R. (2017). *The interpretation of cultures: Selected essays* (3rd edition). Basic Books.

Gerbaudo, P., & Treré, E. (2015). In search of the 'we' of social media activism: Introduction to the special issue on social media and protest identities. *Information, Communication & Society*, *18*(8), 865–871. https://doi.org/10.1080/1369118X.2015.1043319

Gessner, C. (1561). *Drawing of a red tulip*. Wikimedia. https://commons.wikimedia.org/wiki/File:1561_Gesner_Tulip.jpg

Giles, H. (1973). Accent Mobility: A Model and Some Data. *Anthropological Linguistics*, *15*(2), 87–105.

Giorgi, S., Lynn, V., Matz, S., Ungar, L., & Schwartz, H. A. (2019). *Correcting sociodemographic selection biases for accurate population prediction from social media.* arXiv preprint arXiv:1911.03855.

Goffman, E. (1990). *The presentation of self in everyday life* (Repr). Penguin.

Guadagno, R. E., Rempala, D. M., Murphy, S., & Okdie, B. M. (2013). What makes a video go viral? An analysis of emotional contagion and Internet memes. *Computers in Human Behavior*, *29*(6), 2312–2319. https://doi.org/10.1016/j.chb.2013.04.016

Guntuku, S. C., Li, M., Tay, L., & Ungar, L. H. (2019). Studying Cultural Differences in Emoji Usage across the East and the West. *Proceedings of the International AAAI Conference on Web and Social Media*, *13*, 226–235. https://www.aaai.org/ojs/index.php/ICWSM/article/view/3224

Halperin, S., & Heath, O. (2020). *Political research: Methods and practical skills*. Oxford University Press.

Hargittai, E. (2020). Potential Biases in Big Data: Omitted Voices on Social Media. *Social Science Computer Review*, *38*(1), 10–24. https://doi.org/10.1177/0894439318788322

Hepp, A., Breiter, A., & Friemel, T. (2018). Digital Traces in Context—An Introduction. *International Journal of Communication*, *12*, 11.

Hernandez, J., Hoque, M. (Ehsan), Drevo, W., & Picard, R. W. (2012). Mood meter: Counting smiles in the wild. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, 301. https://doi.org/10.1145/2370216.2370264

Herring, S., & Dainas, A. (2018, June 25). Receiver Interpretations of Emoji Functions: A Gender Perspective. *Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media (Emoji 2018)*.

https://pdfs.semanticscholar.org/8a3d/6c2da0a4d90b5f60b3dd25fd0aae5d02f3d4.pdf
Hier, S. P. (2005). *Contemporary sociological thought themes and theories*. Canadian Scholars' Press. http://site.ebrary.com/id/10191734
Highfield, T., Harrington, S., & Bruns, A. (2013). TWITTER AS A TECHNOLOGY FOR AUDIENCING AND FANDOM: The #Eurovision phenomenon. *Information, Communication & Society*, *16*(3), 315–339. https://doi.org/10.1080/1369118X.2012.756053
Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, *2*(1), 47–62. https://doi.org/10.1080/22041451.2016.1155332
Hillberg, H., Levonian, Z., Kluver, D., Terveen, L., & Hecht, B. (2018). What I See is What You Don't Get: The Effects of (Not) Seeing Emoji Rendering Differences across Platforms. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1–24. https://doi.org/10.1145/3274393
Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter Topic-Networks Using Social Network Analysis. *Social Media + Society*, *3*(1), 205630511769154. https://doi.org/10.1177/2056305117691545
Hodson, J., & Petersen, B. (2019). Diversity in Canadian election-related Twitter discourses: Influential voices and the media logic of #elxn42 and #cdnpoli hashtags. *Journal of Information Technology & Politics*, *16*(3), 307–323. https://doi.org/10.1080/19331681.2019.1646181
Holland, J. H. (2003). *Hidden order: How adaptation builds complexity* (10. print). Perseus Books.
Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in Twitter. *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, 80–88. https://doi.org/10.1145/1964858.1964870
Hu, T., Guo, H., Sun, H., Nguyen, T. T., & Luo, J. (2017). Spice up Your Chat: The Intentions and Sentiment Effects of Using Emoji. *ArXiv:1703.02860 [Cs]*. http://arxiv.org/abs/1703.02860
Hughes, A., & Asheer, N. (2019). *National politics on Twitter: Small share of U.S. adults produce majority of tweets*. Pew Research Center.
Hussein, B. (2019). *Lebanon protests* [Photograph]. AP Images. http://www.apimages.com/metadata/Index/Lebanon-Protests/156d710a90334b87ad2b35ea545a9cfe/17/0
Jackson, P., Penrose, J., & Association of American Geographers (Eds.). (1994). *Constructions of race, place, and nation*. University of Minnesota Press.
Jenkins, A., Croitoru, A., Crooks, A. T., & Stefanidis, A. (2016). Crowdsourcing a Collective Sense of Place. *PLOS ONE*, *11*(4), e0152932. https://doi.org/10.1371/journal.pone.0152932

Jeong, S., & Santos, C. Almeida. (2004). CULTURAL POLITICS AND CONTESTED PLACE IDENTITY. *Annals of Tourism Research*, *31*(3), 640–656. https://doi.org/10.1016/j.annals.2004.01.004

Jha, J. C. (1976). The Hindu Festival of Divali in the Caribbean. *Caribbean Quarterly*, *22*(1), 53–61. https://doi.org/10.1080/00086495.1976.11829270

Kalimeri, K., Beiró, M. G., Delfino, M., Raleigh, R., & Cattuto, C. (2019). Predicting demographics, moral foundations, and human values from digital behaviours. *Computers in Human Behavior*, *92*, 428–445. https://doi.org/10.1016/j.chb.2018.11.024

Kaneko, T., & Yanai, K. (2016). Event photo mining from Twitter using keyword bursts and image clustering. *Neurocomputing*, *172*, 143–158. https://doi.org/10.1016/j.neucom.2015.02.081

Karabacak, Z., & Sezgin, A. (2013). The Reflection of Cultural Heritage of Ottoman Empire to Everyday Life as a Popular Culture Product. In *Turkish Studies International Periodical for the Languages, Literature and History of Turkish or Turkic* (pp. 299–305).

Kariryaa, A., Rundé, S., Heuer, H., Jungherr, A., & Schöning, J. (2020). The Role of Flag Emoji in Online Political Communication. *Social Science Computer Review*, 089443932090908. https://doi.org/10.1177/0894439320909085

Kavak, H., Kim, J.-S., Crooks, A., Pfoser, D., Wenk, C., & Züfle, A. (2019). Location-Based Social Simulation. *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*, 218–221. https://doi.org/10.1145/3340964.3340995

Kavak, H., Vernon-Bido, D., & Padilla, J. J. (2018). Fine-Scale Prediction of People's Home Location Using Social Media Footprints. In R. Thomson, C. Dancy, A. Hyder, & H. Bisgin (Eds.), *Social, Cultural, and Behavioral Modeling* (Vol. 10899, pp. 183–189). Springer International Publishing. https://doi.org/10.1007/978-3-319-93372-6_20

Kaye, L. K., Wall, H. J., & Malone, S. A. (2016). "Turn that frown upside-down": A contextual account of emoticon usage on different virtual platforms. *Computers in Human Behavior*, *60*, 463–467. https://doi.org/10.1016/j.chb.2016.02.088

Kelly, R., & Watts, L. (2015, September 20). *Characterising the Inventive Appropriation of Emoji as Relationally Meaningful in Mediated Close Personal Relationships*. Experiences of technology appropriation: Unanticipated users, usage, circumstances, and design, Oslo, Norway.

Khalid, O., & Srinivasan, P. (2020). Style Matters! Investigating Linguistic Style in Online Communities. *Proceedings of the International AAAI Conference on Web and Social Media*, *14(1)*, 360–369. https://www.aaai.org/ojs/index.php/ICWSM/article/view/7306

Kim, J.-S., Jin, H., Kavak, H., Rouly, O. C., Crooks, A. T., Pfoser, D., Wenk, C., & Zufle, A. (2020, July 2). Location-Based Social Network Data Generation Based on Patterns of Life. *The 21st IEEE International Conference on Mobile Data Management*.

Kimble, J., & Olson, L. (2006). Visual Rhetoric Representing Rosie the Riveter: Myth and Misconception in J. Howard Miller's "We Can Do It!" Poster. *Rhetoric & Public Affairs*, *9*(4), 533–569.

Kimura, M., & Katsurai, M. (2017). Automatic Construction of an Emoji Sentiment Lexicon. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 1033–1036. https://doi.org/10.1145/3110025.3110139

Kinsley, D. (1986). *Hindu goddesses: Visions of the divine feminine in the Hindu religious tradition* (1st paper. print). Univ. of California Press.

Kong, L., & Yeoh, B. S. A. (1997). The construction of national identity through the production of ritual and spectacle. *Political Geography*, *16*(3), 213–239. https://doi.org/10.1016/0962-6298(95)00135-2

Kosmajac, D., & Keselj, V. (2019). Twitter Bot Detection using Diversity Measures. *Proceedings of the 3rd International Conference on Natural Language and Speech Processing*, 1–8.

Kralj Novak, P., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of Emojis. *PLOS ONE*, *10*(12), e0144296. https://doi.org/10.1371/journal.pone.0144296

Kurita, S. (1999). *Emoji 1998-1999*. MOMA. https://www.moma.org/collection/works/196070

Laing, J., & Frost, W. (Eds.). (2015). *Rituals and traditional events in the modern world*. Routledge.

Leach, E. R. (2003). *Culture & communication: The logic by which symbols are connected ; an introduction to the use of structuralist analysis in social anthropology* (transferred to digital print). Univ. Press.

Leithead, A. (2019). *Sudan stand-off after Omar al-Bashir ousted* [Photograph]. BBC. https://www.bbc.com/news/world-africa-47980568

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*(1), 78–106. https://doi.org/10.1016/0010-0285(82)90005-6

Levin, S. A. (1999). *Fragile dominion: Complexity and the commons*. Perseus Books.

Ljubešić, N., & Fišer, D. (2016). A Global Analysis of Emoji Usage. *Proceedings of the 10th Web as Corpus Workshop*, 82–89. https://doi.org/10.18653/v1/W16-2610

Loy, S. (2020). Spread of Flower Symbolism: From the Victorian Language of Flowers to Modern Flower Emoji. In S. D. Brunn & R. Kehrein (Eds.), *Handbook of the Changing World Language Map* (pp. 4059–4082). Springer International Publishing. https://doi.org/10.1007/978-3-030-02438-3_59

Lupton, D. (1994). Food, Memory and Meaning: The Symbolic and Social Nature of Food Events. *The Sociological Review*, *42*(4), 664–685. https://doi.org/10.1111/j.1467-954X.1994.tb00105.x

Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, *13*(1), 114–133. https://doi.org/10.1177/1461444810365313

Maynard, D. W., & Peräkylä, A. (2006). Language and Social Interaction. In J. Delamater (Ed.), *Handbook of Social Psychology* (pp. 233–257). Springer US. https://doi.org/10.1007/0-387-36921-X_10

McCulloch, G., & Gawne, L. (2018). Emoji grammar as beat gestures. *Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media (Emoji 2018)*, *2130*. http://ceur-ws.org/Vol-2130/short1.pdf

McGarry, A., Jenzen, O., Eslen-Ziya, H., Erhart, I., & Korkut, U. (2019). Beyond the iconic protest images: The performance of 'everyday life' on social media during Gezi Park. *Social Movement Studies*, *18*(3), 284–304. https://doi.org/10.1080/14742837.2018.1561259

McInnes, L., Healy, J., & Astels, S. (2017). *HDBSCAN: Hierarchical density-based clustering*. *2(11)*, 205. https://doi.org/10.21105/joss.00205

Medlock, B., & McCulloch, G. (2016, March 11). *The linguistic secrets found in billions of emojis*. SXSW 2016, Austin TX. https://www.slideshare.net/SwiftKey/the-linguistic-secrets-found-in-billions-of-emoji-sxsw-2016-presentation-59956212

Meinert, L., & Kapferer, B. (Eds.). (2015). *In the event: Toward an anthropology of generic moments*. Berghahn Books.

Milan, S. (2015). From social movements to cloud protesting: The evolution of collective identity. *Information, Communication & Society*, *18*(8), 887–900. https://doi.org/10.1080/1369118X.2015.1043135

Miller, H., Kluver, D., Thebault-Spieker, J., Terveen, L., & Hecht, B. (2017). Understanding Emoji Ambiguity in Context: The Role of Text in Emoji-Related Miscommunication. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*, 152–161. https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15703/14804

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). "Blissfully happy" or "ready to fight": Varying interpretations of emoji. *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*, 259–268. https://experts.umn.edu/en/publications/blissfully-happy-or-ready-to-fight-varying-interpretations-of-emo

Miller, J. H., & Page, S. E. (2007). *Complex Adaptive Systems*. Princeton University Press.

Miller, J. Howard. (1943). *We Can Do It* [Poster]. Smithsonian National Museum of American History, Washington DC. https://americanhistory.si.edu/collections/search/object/nmah_538122

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. N. (2011). Understanding the Demographics of Twitter Users. In L. A. Adamic, R. Baeza-Yates, & S. Counts (Eds.), *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. The AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2816

Morris, C. (1946). *Signs, Language, and Behavior*. Prentice-Hall.

Morstatter, F., & Liu, H. (2017). Discovering, assessing, and mitigating data bias in social media. *Online Social Networks and Media*, *1*, 1–13. https://doi.org/10.1016/j.osnem.2017.01.001

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013, July 8). Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. *Proceedings of the International AAAI Conference on Web and Social Media*. Seventh International AAAI Conference on Weblogs and Social Media, Cambridge, Massachusetts. http://arxiv.org/abs/1306.5204

Morstatter, F., Shu, K., Wang, S., & Liu, H. (2017). Cross-Platform Emoji Interpretation: Analysis, a Solution, and Applications. *ArXiv:1709.04969 [Cs]*. http://arxiv.org/abs/1709.04969

Na'aman, N., Provenza, H., & Montoya, O. (2017a). Varying Linguistic Purposes of Emoji in (Twitter) Context. *Proceedings of ACL 2017, Student Research Workshop*, 136–141. https://www.aclweb.org/anthology/P17-3022

Na'aman, N., Provenza, H., & Montoya, O. (2017b). Varying linguistic purposes of emoji in (Twitter) context. *ACL 2017, Student Research Workshop*, 136–141.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360. https://doi.org/10.1177/026192702237953

OSX Daily. (2011, December 19). *Enable the Emoji Keyboard on an iPhone*. http://osxdaily.com/2011/12/19/enable-emoji-keyboard-iphone/

Oxford Dictionaries Blog. (2015). *Word of the Year 2015*. https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2015

Padilla, J. J., Kavak, H., Lynch, C. J., Gore, R. J., & Diallo, S. Y. (2018). Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. *PloS One*, *13*(6), e0198857.

Pang, A. S.-K. (1993). The Social Event of the Season: Solar Eclipse Expeditions and Victorian Culture. *Isis*, *84*(2), 252–277. JSTOR.

Parra, F. (2019). *Venezuela protests*. Getty Images. https://www.gettyimages.com/detail/news-photo/people-raise-their-hands-during-a-mass-opposition-rally-news-photo/1087690026

Pavalanathan, U., & Eisenstein, J. (2015). Emoticons vs. Emojis on Twitter: A Causal Inference Approach. *ArXiv:1510.08480 [Cs]*. http://arxiv.org/abs/1510.08480

Pavord, A. (2019). *Tulip: The Story of a Flower That Has Made Men Mad.* Bloomsbury. http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9781526602671

Pederson, J. (2016). *It's Not What You Tweet but How You Tweet It: An Experiment of Orientation, Interactivity, and Valence in Twitter* [Dissertation, Texas A&M University]. http://hdl.handle.net/1969.1/157794

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, *54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Pohl, H., Domin, C., & Rohs, M. (2017). Beyond Just Text: Semantic Emoji Similarity Modeling to Support Expressive Communication 👫📲😃. *ACM Transactions on Computer-Human Interaction*, *24*(1), 1–42. https://doi.org/10.1145/3039685

Polletta, F., & Jasper, J. M. (2001). Collective Identity and Social Movements. *Annual Review of Sociology*, *27*(1), 283–305. https://doi.org/10.1146/annurev.soc.27.1.283

Preoţiuc-Pietro, D., & Cohn, T. (2013). Mining user behaviours: A study of check-in patterns in location based social networks. *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13*, 306–315. https://doi.org/10.1145/2464464.2464479

Preoţiuc-Pietro, D., & Ungar, L. (2018). User-level race and ethnicity predictors from Twitter text. *Proceedings of the 27th International Conference on Computational Linguistics*, 1534–1545.

Rajabi, Z., Shehu, A., & Purohit, H. (2019). User Behavior Modelling for Fake Information Mitigation on Social Web. *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 234–244.

Ramkissoon, H. (2015). Authenticity, satisfaction, and place attachment: A conceptual framework for cultural tourism in African island economies. *Development Southern Africa*, *32*(3), 292–302. https://doi.org/10.1080/0376835X.2015.1010711

Revay, P., & Cioffi-Revilla, C. (2017). Modeling the Co-evolution of Culture, Signs and Network Structure. In D. Lee, Y.-R. Lin, N. Osgood, & R. Thomson (Eds.), *Social, Cultural, and Behavioral Modeling* (Vol. 10354, pp. 162–171). Springer International Publishing. https://doi.org/10.1007/978-3-319-60240-0_20

Rice, D. (2019, January 14). Here's your guide to the "super blood wolf moon eclipse" that's coming this weekend. *USA Today*. https://www.usatoday.com/story/news/nation/2019/01/14/super-blood-wolf-moon-eclipse-coming-weekend-what-does-mean

Riordan, M. A. (2017). The communicative role of non-face emojis: Affect and disambiguation. *Computers in Human Behavior*, *76*, 75–86. https://doi.org/10.1016/j.chb.2017.07.009

Robertson, A., Magdy, W., & Goldwater, S. (2018). Self-Representation on Twitter Using Emoji Skin Color Modifiers. *Proceedings of the Twelth International AAAI Conference on Web and Social Media (ICWSM 2018)*, 680–683. http://arxiv.org/abs/1803.10738

Rodrigues, D., Lopes, D., Prada, M., Thompson, D., & Garrido, M. V. (2017). A frown emoji can be worth a thousand words: Perceptions of emoji use in text messages exchanged between romantic partners. *Telematics and Informatics*, *34*(8), 1532–1543. https://doi.org/10.1016/j.tele.2017.07.001

Saetre, A., & Browning, L. (2008). Complexity Theories and ICTs. In *Information and Communication Technology in Action: Linking Theory and Narratives of Practice*. Taylor & Francis.

Santana, M. C. (2016). From Empowerment to Domesticity: The Case of Rosie the Riveter and the WWII Campaign. *Frontiers in Sociology*, *1*. https://doi.org/10.3389/fsoc.2016.00016

Santhanam, S., Srinivasan, V., Glass, S., & Shaikh, S. (2018). I Stand With You: Using Emojis to Study Solidarity in Crisis Events. *1st International Workshop on Emoji Understanding and Applications in Social Media*. Emoji '18, Stanford, CA. http://ceur-ws.org/Vol-2130/paper1.pdf

Schuchard, R., Crooks, A. T., Stefanidis, A., & Croitoru, A. (2018). Bots in Nets: Empirical Comparative Analysis of Bot Evidence in Social Networks. In L. M. Aiello, C. Cherifi, H. Cherifi, R. Lambiotte, P. Lió, & L. M. Rocha (Eds.), *Proceedings of the 7th International Conference on Complex Networks and Their Applications* (Vol. 2, pp. 424–436). Springer.

Schuchard, Ross, Crooks, A. T., Stefanidis, A., & Croitoru, A. (2019). Bot stamina: Examining the influence and staying power of bots in online social networks. *Applied Network Science*, *4*(1), 55. https://doi.org/10.1007/s41109-019-0164-x

Schuldt, J. P., & Pearson, A. R. (2016). The role of race and ethnicity in climate change polarization: Evidence from a U.S. national survey experiment. *Climatic Change*, *136*(3–4), 495–505. https://doi.org/10.1007/s10584-016-1631-3

Seden Dogan, & Ayse Collins. (2019). Does communication with emoticons/emojis work efficiently? Evidence from Turkey. In *Advances in Global Business and Marketing* (Vol. 2, pp. 381–407).

Shearer, E., & Gottfried, J. (2017). *News use across social media platforms 2017*. Pew Research Center.

Simon, H. (1952). A formal theory of interaction in social groups. *American Sociological Review*, *17*(2), 202–212.

Singh, K. (2015). *Hindu Rites and Rituals: Origins and Meanings*. Penguin.

Sloan, L., Morgan, J., Housley, W., Williams, M., Edwards, A., Burnap, P., & Rana, O. (2013). Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter. *Sociological Research Online*, *18*(3), 74–84. https://doi.org/10.5153/sro.3001

Soares, F. B., Recuero, R., & Zago, G. (2019). Asymmetric Polarization on Twitter and the 2018 Brazilian Presidential Elections. *Proceedings of the 10th International Conference on Social Media and Society - SMSociety '19*, 67–76. https://doi.org/10.1145/3328529.3328546

Srivastava, G. (2018). Gauging Ecliptic Sentiment. *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, 1–5. https://doi.org/10.1109/TSP.2018.8441433

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal*, *78*(2), 319–338. https://doi.org/10.1007/s10708-011-9438-2

Stefanidis, A., Jenkins, A., Croitoru, A., & Crooks, A. T. (2016). Megacities Through the Lens of Social Media. *Journal of the Homeland Defense & Security Information Analysis Center*, *3*(1), 24–29.

Stefanidis, A., Vraga, E., Lamprianidis, G., Radzikowski, J., Delamater, P. L., Jacobsen, K. H., Pfoser, D., Croitoru, A., & Crooks, A. (2017). Zika in Twitter: Temporal Variations of Locations, Actors, and Concepts. *JMIR Public Health and Surveillance*, *3*(2), e22. https://doi.org/10.2196/publichealth.6925

Sterling, J., Jost, J. T., & Bonneau, R. (2020). Political psycholinguistics: A comprehensive analysis of the language habits of liberal and conservative social media users. *Journal of Personality and Social Psychology*, *118*(4), 805–834. https://doi.org/10.1037/pspp0000275

Stets, J. E., & Burke, P. J. (2000). Identity Theory and Social Identity Theory. *Social Psychology Quarterly*, *63*(3), 224–237.

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, 156–168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter. *Political Communication*, *35*(1), 50–74. https://doi.org/10.1080/10584609.2017.1334728

Sumartojo, S. (2016). Commemorative atmospheres: Memorial sites, collective events and the experience of national identity. *Transactions of the Institute of British Geographers*, *41*(4), 541–553. https://doi.org/10.1111/tran.12144

Swartz, M., & Crooks, A. (2020, February 2). Comparison of Emoji Use in Names, Profiles, and Tweets. *The Eighth IEEE International Workshop on Semantic Computing for Social Networks and Organization Sciences*.

Swidler, A. (1986). Culture in Action: Symbols and Strategies. *American Sociological Review*, *51*(2), 273. https://doi.org/10.2307/2095521

Swiftkey. (2015). Most-used emoji revealed: Americans love skulls, Brazilians love cats, the French love hearts. *Swiftkey*. https://blog.swiftkey.com/americans-love-skulls-brazilians-love-cats-swiftkey-emoji-meanings-report/

Tigwell, G. W., & Flatla, D. R. (2016). Oh that's what you meant!: Reducing emoji misunderstanding. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '16*, 859–866. https://doi.org/10.1145/2957265.2961844

Torres, E. C., Moreira, S., & Lopes, R. C. (2018). Understanding how and why people participate in crowd events. *Social Science Information*, *57*(2), 304–321. https://doi.org/10.1177/0539018418757714

Tufekci, Z., & Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. *Journal of Communication*, *62*(2), 363–379. https://doi.org/10.1111/j.1460-2466.2012.01629.x

Tulk, S., Bagheri-Jebelli, N., & Kennedy, W. G. (2018). Modeling the Impact of Fake News on Citizens. *Proceedings of the 16th Annual Meeting of the International Conference on Cognitive Modelling*, 187–192.

Turner, V. (2006). *Dramas, fields, and metaphors: Symbolic action in human society* (Nachdr.). Cornell Univ. Press.

Tyson, A. (2018). *The 2018 midterm vote: Divisions by race, gender, and education*. Pew Research Center.
Unicode. (2010, October 12). Unicode 6.0: Popular Symbols for Asia. *Unicode News and Announcements from the Unicode Consortium*. http://unicode.org/announcements/pr-6.0.html
Unicode. (2019, February 5). The Unicode Blog: Unicode Emoji 12.0—Final for 2019. *Unicode News and Announcements from the Unicode Consortium*. http://blog.unicode.org/2019/02/unicode-emoji-12-final-for-2019.html
Varma, R. (1896). *Painting of the Goddess Saraswati*. https://commons.wikimedia.org/wiki/File:Saraswati.jpg
Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions:Detection, Estimation, and Characterization. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, 280–289. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15587/14817
Vergeer, M., Hermans, L., & Sams, S. (2013). Online social networks and micro-blogging in political campaigning: The exploration of a new campaign tool and a new campaign style. *Party Politics*, *19*(3), 477–501. https://doi.org/10.1177/1354068811407580
Vincent, J. (1999). Symbols of nationalism in Bosnia and Hercegovina. In K. Cameron (Ed.), *National Identity* (pp. 46–63).
Vraga, E. K., Stefanidis, A., Lamprianidis, G., Croitoru, A., Crooks, A. T., Delamater, P. L., Pfoser, D., Radzikowski, J. R., & Jacobsen, K. H. (2018). Cancer and Social Media: A Comparison of Traffic about Breast Cancer, Prostate Cancer, and Other Reproductive Cancers on Twitter and Instagram. *Journal of Health Communication*, *23*(2), 181–189. https://doi.org/10.1080/10810730.2017.1421730
Walther, J. B. (2012). Interaction Through Technological Lenses: Computer-Mediated Communication and Language. *Journal of Language and Social Psychology*, *31*(4), 397–414. https://doi.org/10.1177/0261927X12446610
Wijeratne, S., Balasuriya, L., Doran, D., & Sheth, A. (2016). Word Embeddings to Enhance Twitter Gang Member Profile Identification. *ArXiv:1610.08597 [Cs]*. http://arxiv.org/abs/1610.08597
Wikipedia. (2020). *Shiva* [Photograph]. Wikipedia. https://en.wikipedia.org/w/index.php?title=Shiva&oldid=969073359
Wilson, M. (1954). Nyakyusa Ritual and Symbolism. *American Anthropologist*, *56*(2), 228–241. https://doi.org/10.1525/aa.1954.56.2.02a00060
Wirth, K., Menchen-Trevino, E., & Moore, R. T. (2019). Bots By Topic: Exploring Differences in Bot Activity by Conversation Topic. *Proceedings of the 10th International Conference on Social Media and Society - SMSociety '19*, 77–82. https://doi.org/10.1145/3328529.3328547
Wiseman, S., & Gould, S. J. J. (2018). Repurposing Emoji for Personalised Communication: Why 🍕 means "I love you." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–10. https://doi.org/10.1145/3173574.3173726

Wood, I., & Ruder, S. (2016). Emoji as emotion tags for tweets. *Proceedings of the Emotion and Sentiment Analysis Workshop (LREC2016)*, 76–79.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using Social Media to Quantify Nature-based Tourism and Recreation. *Scientific Reports*, *3*, 2976.

Yakın, V., & Eru, O. (2017). An application to determine the efficacy of emoji use on social marketing ads. *International Journal of Social Sciences and Education Research*, *3*(1), 230–230. https://doi.org/10.24289/ijsser.270652

Yuan, X., & Crooks, A. T. (2018). Examining Online Vaccination Discussion and Communities in Twitter. *Proceedings of the 9th International Conference on Social Media and Society*, 197–206. https://doi.org/10.1145/3217804.3217912

Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019a). Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter. *Social Media + Society*, *5*(3), 205630511986546. https://doi.org/10.1177/2056305119865465

Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019b). Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter. *Social Media + Society*, *5*(3), 205630511986546. https://doi.org/10.1177/2056305119865465

Zannettou, S., Caulfield, T., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Blackburn, J. (2019). Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. *Companion Proceedings of The 2019 World Wide Web Conference on - WWW '19*, 218–226. https://doi.org/10.1145/3308560.3316495

Zhao, S., Zhong, L., Wickramasuriya, J., & Vasudevan, V. (2011). *Human as Real-Time Sensors of Social and Physical Events: A Case Study of Twitter and Sports Games*. Technical Report TR0620-2011, Rice University and Motorola Labs, arXiv:1106.4300v1.

Zhou, R., Hentschel, J., & Kumar, N. (2017). Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 748–759. https://doi.org/10.1145/3025453.3025800

BIOGRAPHY

Melanie Swartz received her Bachelor of Science degree from The Pennsylvania State University in 1999 in the field of Geography. In 2010 she began taking graduate courses in the social sciences, however her interest in the quantitative aspects of her social studies led her to the Computational Social Science (CSS) program at George Mason University. In the CSS program she honed her research skills focusing on social-environmental interactions through modeling and simulation and received a master's degree in 2015. Continuing on with the program in the Computational and Data Sciences Department, she conducted research at the intersection of social behavior, linguistics, and data science, documented in this dissertation, which earned her a PhD in CSS in the summer of 2020.