Comparison of Emoji Use in Names, Profiles, and Tweets

Melanie Swartz
Dept. of Computational and Data Sciences
George Mason University
Fairfax, Virginia, USA
mswartz2@gmu.edu

Andrew Crooks
Dept. of Computational and Data Sciences
George Mason University
Fairfax, Virginia, USA
acrooks2@gmu.edu

***Abstract*—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 paper fills this gap by comparing emoji use across users and collectively in user names, profiles, and in original and re-shared content. We present a methodology that enables comparison of semantically similar emojis based on Unicode emoji categories and subcategories. We apply this methodology 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, our analysis reveals emoji super users who have a significantly higher proportion and diversity of emoji use. Our methodology offers a novel approach for summarizing emoji use and enables systematic comparison of emojis across individual user profiles and communication patterns, thus expanding methods for semantic analysis of social media data beyond just text.**

Keywords—emoji; social media analytics; content analysis; online social networks

I. 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 [1, 2].

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 [3, 4]. In addition, some social networking sites, such as Twitter, limit the size of the content or number of characters allowed. In that regard, emojis can be more efficient than their textual equivalent [5].

Although emojis originated in the late 1990s, their use only recently become popular on social networking sites [6]. 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 [7], 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 [8]. 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 we focus on in this paper.

The overall contributions of this paper are: (1) we present a methodology to extract, aggregate, and compare emoji use across a collection of documents based on Unicode emoji category and subcategories, (2) we present a baseline of statistics of emoji use in user names, profile descriptions, and tweets, and (3) we compare 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 paper is organized as follows. In Section II we briefly highlight related work which is followed by our methodology to aggregate and analyze emojis (Section III). Next we present the results and comparison of emoji use applied to a corpus of tweets and user profiles we collected related to the 2018 US midterm elections (Section IV). Finally, in Section V we conclude and identify areas for further research.

II. 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.

A. 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 [9], gender [10], and at the country level [11]. At the individual level, research has correlated emoji use to social identity [12]. Other researchers have identified personal preferences on emoji use related to marketing and how people respond to them [13]. 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 we explore and present in this paper.

978-1-7281-6332-1/20/$31.00 ©2020 IEEE
DOI 10.1109/ICSC.2020.00075

B. 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 [14, 15]. 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 [16] or to omit emojis all together [17]. 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 [14, 18]. 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 paper addresses this gap by considering emojis grouped by Unicode categories and subcategories as semantically similar. Further, we use these groupings to enable comparison of emoji use.

C. 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 [19, 20]. At the document or user group level such as gender or country, emoji use is often compared based on the most frequently used emojis [9-11]. 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. We 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.

III. Methodology

To summarize and compare emoji use in documents and per user, our 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. Fig 1. shows the four parts of our workflow.

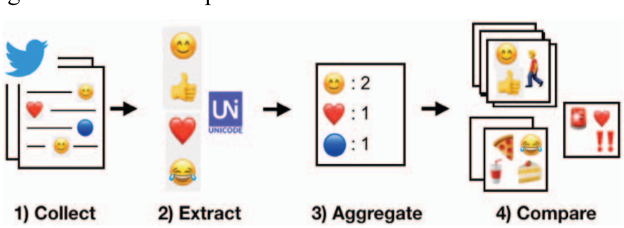

Fig. 1. Workflow for emoji comparison

A. Data Collection

For our analysis, we analyzed emoji use in tweets collected from the free public Twitter streaming application programming interface (API). We then used the Twitter standard search API to collect user profiles of all the users who authored tweets or were retweeted in our 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 [21]. After this step, we 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). The keywords we used to collect tweets and our case study will be introduced in Section IV.

B. 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; for example, 👩🏾‍💻 is composed of 👩, skin-tone modifier, and 💻. In addition, country flag emojis are made up of two emojis, such as 🇺 🇸 for the flag of the United States of America 🇺🇸. Finally, 5) to generate the unique list of emojis, we removed duplicate occurrences and sorted the unique emojis. Our rationale for doing this was to enable comparison and counting of emojis used per user and tweet during aggregation.

For each unique emoji extracted, we 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, Fig. 2 shows the Unicode category Smileys & Emotion and corresponding subcategory of face-smiling for two emojis. We used the Unicode data files version 12.1 [22] to generate the full mapping of each of the 3000 plus emojis to their respective category and subcategory.

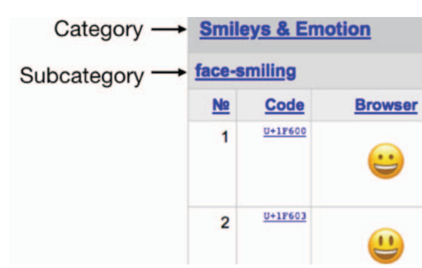

Fig. 2. Unicode emoji category and subcategory example

C. Aggregate by Unit of Analysis

For this paper we summarized and compared emoji use by user and document type, thus we consider user and document type as our units of analysis for aggregation. The document types we analyzed are the user name, profile description, retweet content, and non-retweet content. For each user, we generated a list of their unique emojis, categories, and subcategories they used per document type. Next we created a sorted list in descending order based on the count of user retweets and non-retweets per emoji, categories, and subcategories. We also summarized emoji use across all users per each document type. For emoji, category, and subcategory, we 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.

D. Comparison of Emoji Use

After the emojis for each user and document type have been aggregated, the final step is to compare their use. We 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, we measured similarity of emojis and subcategories that were used by at least 50% of the user population. To do this, we used the Jaccard similarity coefficient (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). We report Jaccard similarity measures between document types in the results.

$$J(\mathrm{A},\mathrm{B}) = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A} \cup \mathrm{B}|} \quad (1)$$

IV. RESULTS AND ANALYSIS

Utilizing the methodology from Section III, we 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. We 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, we 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, we present our results measuring behavior of emoji use and comparison of emojis used in names, profiles, and tweets.

A. Dataset and Percent of Emoji Use

We collected over 44 million tweets 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 (e.g., BetoORourke). Table I summarizes the dataset as tweet or author counts and percent utilizing emojis. The percent of tweets that contained an emoji was 8.28%. We divided them into retweets, which accounted for 83% of our collected tweets, and non-retweets. Non-retweets had a slightly higher percent of tweets containing emoji (9.54%) compared to retweets (8.03%).

Our 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 our 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 I. DATASET OVERVIEW OF 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
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

B. Number of Unique Emojis used 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, Fig. 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 we show emoji use by category and subcategory.

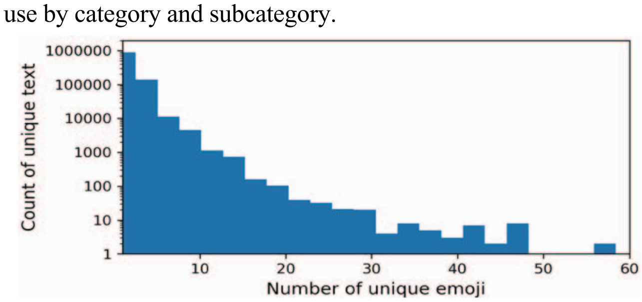

Fig. 3. Histogram of the number of unique emojis per unique text

C. Behavior of Emoji Use and Emoji Super Users

During the one-month timeframe of our data collection, we found that it was more common for emoji users to send both non-retweets and retweets containing emojis. With this in mind, we 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%. Fig. 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%). We 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.

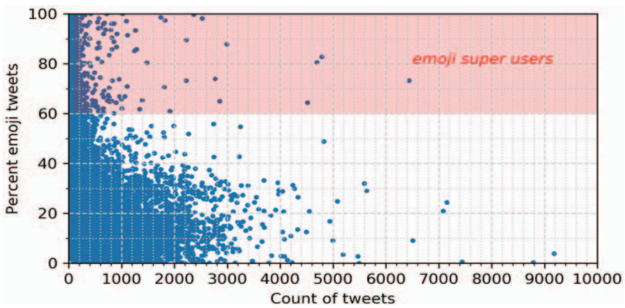

Fig. 4. Plot of count of tweets and percent emoji tweets per user

We 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 we compared emojis used in names and profile descriptions and diversity of emojis used. Earlier we 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 we compare emoji categories and subcategories used collectively across all users for the document types of user names, profile descriptions, retweets, and non-retweets.

D. *Emoji Categories*

As noted in Section III, there are nine overarching categories that encompass the full set of emojis. In this section we present the results of our 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 Figs. 5 and 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). Fig. 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. Fig. 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 our 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, we compare groupings of emojis by subcategories within the Unicode emoji categories.

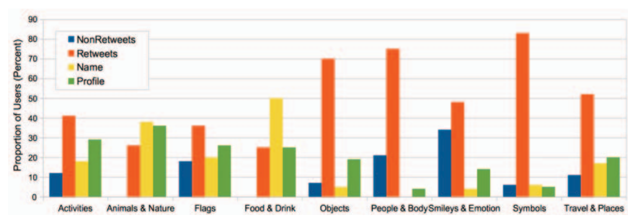

Fig. 5. Comparison of use by emoji category per document type

E. *Emoji Subcategories*

In this section we compare the proportion of emoji use between the four document types of user profile descriptions, user names, retweets, and non-retweets across the 95 Unicode emoji subcategories. We consider emojis within the same subcategory to be semantically similar. Fig. 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

Fig. 6. Proportion of emoji use in profiles, names, retweets, and non-retweets, ordered by category

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 our dataset was emotion and the most-used emoji in that subcategory was ❤️. 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 🇺🇸 emoji used the most, which is expected as our data was collected on keywords related to the 2018 U.S. midterm elections. The fourth most popular subcategory, was sky & weather with the 🌊 emoji having greatest use which was associated with the Democratic party campaign slogan, blue wave. The mail subcategory with the 🗳️ 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.

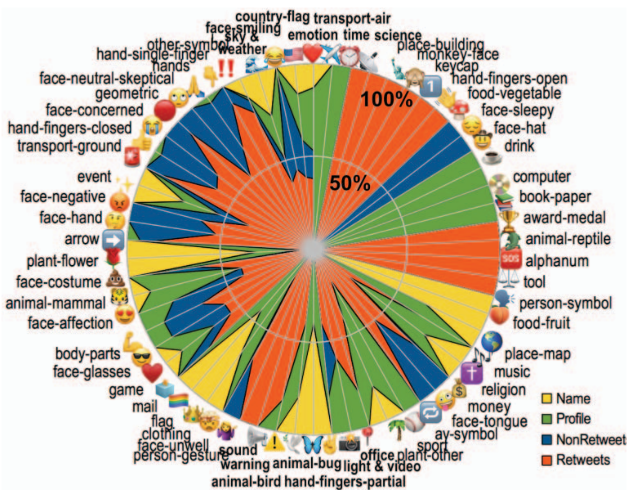

Fig. 7. Proportion of emoji use by subcategory

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, alphanum, 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 we measured similarity of the top emojis and subcategories per document type.

F. Emoji and Subcategory Similarity

We 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 III. 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, we 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 our analysis, based on the above rationale, we 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 II. The Jaccard similarity coefficients between the top 250 emojis for each document type (Table III) 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, we 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 IV shows the top 15 emoji subcategories per document type. The Jaccard similarity of top 35 subcategories between each document type is shown in Table V. 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 II. TOP EMOJI FOR EACH COMMUNICATION TYPE

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

TABLE III. JACCARD SIMILARITY FOR 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 IV. 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 V. 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

V. CONCLUSION

Analysis of the role of emoji use in online communication is still a growing area of research. This paper contributes to this research by presenting a 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, we move the research on emojis beyond just comparing individual emojis and broad aggregations. In applying our methodology to a set of 44 million tweets and over 3 million user profiles relating to the 2018 U.S. midterm elections, we 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, our 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: our work only looks at one use case and thus a question is how representative are our 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 paper.

While these limitations exist, our work shows that emojis are more than just text and the methodology in this paper supports semantic content analysis of documents containing emojis. Our 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, our 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 our understanding of the methods and styles of digital communication in online social networks.

REFERENCES

[1] R. Kelly and L. Watts, "Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships," presented at *Experiences of technology appropriation: Unanticipated users, usage, circumstances, and design*, Oslo, Norway, September 20, 2015.

[2] V. Yakın and O. Eru, "An application to determine the efficacy of emoji use on social marketing ads," *International Journal of Social Sciences and Education Research*, vol. 3, no. 1, 2017, pp. 230-240.

[3] P. K. Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PloS one*, vol. 10, no. 12, 2015, e0144296.

[4] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2017, pp. 1615–1625.

[5] S. Wiseman, and S. Gould, "Repurposing emoji for personalised communication: Why🍕 means "I love you"," in *Proc. Conf. CHI Human Factors in Computing Systems*, 2018, pp. 152.

[6] T. Highfield, and T. Leaver, "Instagrammatics and digital methods: Studying visual social media, from selfies and GIFs to memes and emoji," *Communication Research and Practice,* vol. 2, no. 1, 2016, pp. 47-62.

[7] "Unicode version 6.0: Support for popular symbols in Asia", *Unicode News and Announcements from the Unicode Consortium,* 12 October 2010. http://blog.unicode.org/2010/10/unicode-version-60-support-for-popular.html

[8] "Unicode Emoji 12.0 – final for 2019", *Unicode News and Announcements from the Unicode Consortium*, 5 February 2019. http://blog.unicode.org/2019/02/unicode-emoji-12-final-for-2019.html

[9] S. Guntuku, M. Li, L. Tay, and L. Ungar, "Studying cultural differences in emoji usage across the east and the west," in *Proc. of the International AAAI Conference on Web and Social Media*, vol. 13, no. 01, 2019, pp. 226-235.

[10] Z. Chen, X. Lu, S. Shen, W. Ai, X. Liu, and Q. Mei, "Through a gender lens: Learning usage patterns of emojis from large-scale android users," in *Proc. Conf. World Wide Web*, 2018, pp. 763-772.

[11] N. Ljubešić, and D. Fišer, "A global analysis of emoji usage," in *Proc. 10th Web as Corpus Workshop*, 2016, pp. 82-89.

[12] J. Ge, "Emoji sequence use in enacting personal identity," in *Proc. Companion of the World Wide Web Conference,* 2019, pp. 426-438.

[13] S. Dogan, and A. Collins, "Does communication with emoticons/emojis work efficiently? Evidence from Turkey," *Advances in Global Business and Economics*, vol. 2, 2019, pp. 374-381.

[14] F. Morstatter, K. Shu, S. Wang, and H. Liu, "Cross-platform emoji interpretation: analysis, a solution, and applications," arXiv:1709.04969 [cs.CL], 2017.

[15] L. Kaye, H. Wall, and S. Malone, "Turn that frown upside-down: a contextual account of emoticon usage on different virtual platforms," *Comput. Hum. Behav.,* vol. 60, July, 2016, pp. 463-467.

[16] H. Miller, D. Kluver, J. Thebault-Spieker, L. Terveen, and B. Hecht, "Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication, " in *Proc. Conf. Eleventh International AAAI Conference on Web and Social Media*, 2017, pp. 152-161.

[17] X. Yuan, R. Schuchard, and A. Crooks, "Examining emergent communities and social bots within the polarized online vaccination debate in twitter," *Social Media+ Society,* vol. 5, no. 3, 2019.

[18] H. Miller, Z. Levonian, D. Kluver, L. Terveen, and B. Hecht, "What I See is What You Don't Get: The Effects of (Not) Seeing Emoji Rendering Differences across Platforms," in *Proc. Conf. on Human-Computer Interaction,* vol. 2, no. CSCW, article 124, November, 2018.

[19] H. Pohl, C. Domin, and M. Rohs, "Beyond just text: Semantic emoji similarity modeling to support expressive communication👫➡📱😃," *ACM Trans. on Computer-Human Interaction*, vol. 24, no. 1, 2017, article 6.

[20] G. McCulloch, and L. Gawne, "Emoji grammar as beat gestures," in *Proc. 1st Int. Workshop on Emoji Understanding and Applications in Social Media*, 2018.

[21] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, ''Is the sample good enough? Comparing data from twitter's streaming APIwith Twitter's firehose,'' in *Seventh international AAAI conference on weblogs and social media*. 2013.

[22] M. Davis and P. Edberg, "Full Emoji List, v12.1", Unicode Emoji Charts, December, 2019. https://unicode.org/emoji/charts/full-emoji-list.html