Through a Gender Lens: An Empirical Study of Emoji Usage over Large-Scale Android Users

Article · May 2017

6 authors, including:

Zhenpeng Chen
Peking University
22 PUBLICATIONS   **209** CITATIONS

SEE PROFILE

Wei Ai
University of Maryland, College Park
13 PUBLICATIONS   **313** CITATIONS

SEE PROFILE

Xuanzhe Liu
Peking University
143 PUBLICATIONS   **2,329** CITATIONS

SEE PROFILE

Qiaozhu Mei
University of Michigan
123 PUBLICATIONS   **10,525** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Through a Gender Lens: An Empirical Study of Emoji Usage over Large-Scale Android Users View project

Project   Mobile Web View project

Through a Gender Lens: An Empirical Study of Emoji Usage over Large-Scale Android Users*

Zhenpeng Chen, Xuan Lu, Sheng Shen, Wei Ai, Xuanzhe Liu, Qiaozhu Mei

Abstract

Emojis have gained incredible popularity in recent years and become a new ubiquitous language for Computer-Mediated Communication (CMC) by worldwide users. Various research efforts have been made to understand the behaviors of using emojis. Gender-specific study is always meaningful for HCI community, however, so far we know very little about whether and how much males and females vary in emoji usage. To bridge such a knowledge gap, this paper makes the first effort to explore the emoji usage through a gender lens. Our analysis is based on the largest data set to date, which covers 134,419 users from 183 countries, along with their over 401 million messages collected in three months. We conduct a multi-dimensional statistical analysis from various aspects of emoji usage, including the frequency, preferences, input patterns, public/private CMC-scenario patterns, temporal patterns, and sentiment patterns. The results demonstrate that emoji usage can significantly vary between males and females. Accordingly, we propose some implications that can raise useful insights to HCI community.

1 Introduction

On April 11, 2015, Andy Murray, a world-wide known tennis player, announced his wedding on Twitter[1]. Unlike any other formal announcement, such an inspiring tweet consists of no words, but 51 **emojis** instead.

Undoubtedly, emojis have gained incredible popularity in recent years. Compared to traditional information representations such as text messages, pictures, or even emoticons, emojis are considered to be more lively, more expressive, and more semantically rich, and thus appreciated by Internet users, particularly on smartphones. The prevalence of emojis has been an amazing phenomenon of social innovation and appreciation. Interwoven into our daily communications, they have been a new ubiquitous language [1].

Prior research in the fields of Computer-Mediated Communication (CMC) and Human-Computer Interaction (HCI) has taken an active interest in studying the similar instant messaging elements to emoji, such as emoticons (e.g., ;-)). HCI research on emoji use, by contrast, is still in an early stage. Given that users generate a large volume of emojis, some recent research interests have been made to understand the behaviors of using emojis across apps [2], across platforms [3], and even across cultures [1]. Among these efforts, one missing issue is the emoji-usage behaviors across genders.

Is the gender issue significant and worthy? Probably yes. It is worth mentioning that Google recently announced the support of some gender-oriented emojis[2], implying the gender is a non-trivial issue in emoji. Furthermore, identifying the gender differences is always an important topic in the HCI research community. Existing studies have demonstrated that there could exist some differences in how males and females use non-verbal cues in face-to-face offline speech [4, 5, 6]. As a result, one may be curious if there are differences in how males and females use in online CMC. In particular, with a conversational

---

user interface, an estimate of gender can be quite useful to infer the possible user profiles. For example, web content/ad providers can make use of such gender differences in user's behavior and decide to deliver proper contents or recommend advertisements accordingly.

Indeed, due to the lack of labelled gender information in most state-of-the-art research, we currently know quite little about the difference of emojis usage between male and female users. In this paper, we made the first descriptive analysis to bridge such a knowledge gap. Our work takes uniqueness at two folds. First, we introduce the largest data set to date, from a leading input method app, namely Kika. Such a data set covers 134,419 users from 183 countries, and their 401 million messages collected in three months. Second, we conduct a multi-dimensional statistical analysis from various aspects of emoji usage, including the frequency, preferences, input patterns, CMC-scenario sensitive patterns, temporal patterns, and sentiment patterns. The analysis over the large-scale data set can at best make our results comprehensive, and thus evidences the statistically significant differences of emoji usage between female and male users.

More specifically, we make the following findings:

- The frequency of emoji usage is quite diverse. Females are more likely to use emojis than males do.

- The emoji usage patterns seem to be different . Female users are more likely to use only one emoji or discretely use multiple emojis in a message, while male users tend to consecutively use multiple emojis.

- The preferred emojis are different. For example, females are more likely to use *face-related* emojis, while male users are more likely to use *heart-related* emojis.

- The emoji usage can be significantly affected by different CMC scenarios for both males and females, i.e., females are more likely to use emojis than males in public communication such as Twitter; instead, in private communication, the situation is reversed.

- The sentiment implied by emojis can vary between males and females for a specific time, such as weekdays, weekends, and festivals.

The rest of this paper is organized as follows. Section 2 summarizes status quo research efforts. Section 3 describes our data set and how the ethical issues are preserved. Section 4 presents our measurement approach and the studied research questions. Section 5 makes the descriptive analysis at a macro level, covering the frequency, choice, and input preferences of emoji, and evidences the gender difference exactly exists in emoji usage. Section 6 and 7 extends the basic analysis from two contexts, i.e., CMC scenarios and temporal-sentiment. Besides the preceding analysis, Section 8 discusses about the implications that could help further potential applications, e.g., how our findings along with some knowledge their derived can help improve the UI layout design and play as possible indicators of inferring the gender information by using only emojis without any texts, and analyzes the limitations that could narrow our study and results. Section 9 ends up the paper with concluding remarks.

2 Related Work

We start with summarizing the relevant background and literature. As the marked popularity of emoji, many researchers have taken an active interest in investigating it. However, there is very limited research on emoji usage from gender perspective. Inspired by the prior literature about emoticons, sentiment analysis, and gender differences in non-verbal expressivity, we consider emoji as a new kind of non-verbal cue and investigate gender impacts on emoji usage in mobile communication.

2.1 Emoticons and Emojis

In everyday verbal communications, we often use body language and facial expressions to better express our complex emotions like humor, doubt, and sarcasm. These cues take up about 93% of everyday

communication [7]. However, they cannot be used in text communication. With the growth of computer-mediated communication, to supply the absence of these cues, people gradually turn to a kind of symbolic representations where emotion or affect is referenced pictorially using alphanumerics, punctuations, or other characters, commonly called "emoticons" [8]. Researchers have been trying to understand the sentiment and non-verbal cues provided by emoticons. Emoticons can be used to strengthen the expression [9] or express emotions [10], humor [11], intimacy [12], and irony [13]. There has also been some research on the emoticon usage across countries [14], across cultures [8], and across statuses [15]. What's more, some prior work focused on the gender gap in emoticon usage. Tossell et al. found that females send more messages with emoticons while males use a more diverse range of emoticons [16]. Hwang reported female students are more likely to use emoticons to express emotion or intimacy and manage meaning than male ones based on a Korean sample [17]. Wolf found that females are expressive in emoticon usage but the difference in frequency of emoticon usage is not statistically significant on the mixed gender newsgroups [18].

However, the limited morphological variation of ASCII symbols limits the expressive power of emoticons. The nonstandard creation and use of them introduce some challenge to data analysis. On the contrast, emojis are preloaded and defined standardly. They can not only be used to express emotions, but also be able to represent various objects such as foods and sports. Since the debut on Twitter and Instagram, emojis quickly won a lot of fans. Some findings even showed that the Twitter users who adopt emojis tend to reduce their usage of emoticons [19]. Besides linguistic functions such as replacing words and describing contents, emojis also have non-verbal effects such as decorating text, adjusting tones, providing additional emotional or situational information, and engaging the recipient [20, 21, 22, 23, 24]. Similar to emoticons, there exists prior research on emoji usage across countries and cultures [25, 1, 26]. However, there are very limited studies on emoji usage from gender perspective. Some related work that did not focus on gender perspective reported several findings about gender impacts on emojis. For example, Nishimura examined data from a blog site in Japan and found that women tend to use more emojis than men [27]. Pohl et al. investigated the gender distribution of users tweeting with emoji and found females are more than males [23]. However, there is no existing work to study gender and emoji through user input data at scale. To bridge this research gap, we investigate the gender effects on emoji usage from large-scale Android users comprehensively.

2.2   Sentiment analysis

Traditionally, sentiments and emotions are collected through survey-based [28], audio/video based [29, 30], biometric-based [31, 32], and behavior-based approaches [33, 34, 35]. As the growth of text communication, researchers turned to identify users' sentiment from the text they typed. Although the researchers have proposed various advanced sentiment analysis techniques, it is still challenging to identify the sentiment and emotions from free text. Some researchers started to integrate non-verbal cues into sentiment analysis of text. For example, Filho et al. proposed an approach which detected users' facial expression reactions to conversations when text chatting [36]. In addition, some researchers attempt to use emoticons and emojis to model the text sentiment [37, 38].

In this paper, we need to know the sentiment behind emojis and take emojis as indictors to sense users' sentiment. Therefore, we investigate the existing approaches to measure the sentiment or semantic representations of emojis. In the previous literature, there are three main approaches: (1) based on the sentiment annotated by participants [39, 3, 40], (2) based on the their textual descriptions and annotation tags defined by the Unicode Standard [1, 23, 41], and (3) based on the context in which the emojis occur [23, 42, 43, 44, 45]. We adopt the third approach in this paper. We apply a state-of-art embedding model to get the semantic representations of emojis and then infer their sentiment.

2.3   Gender and Non-verbal Expressivity

Similar to emoticons, emojis are also proved to offer non-verbal cues that poorly conveyed by text [24]. Therefore, to some extent, this work is about gender and non-verbal expressivity. Conventional wisdom

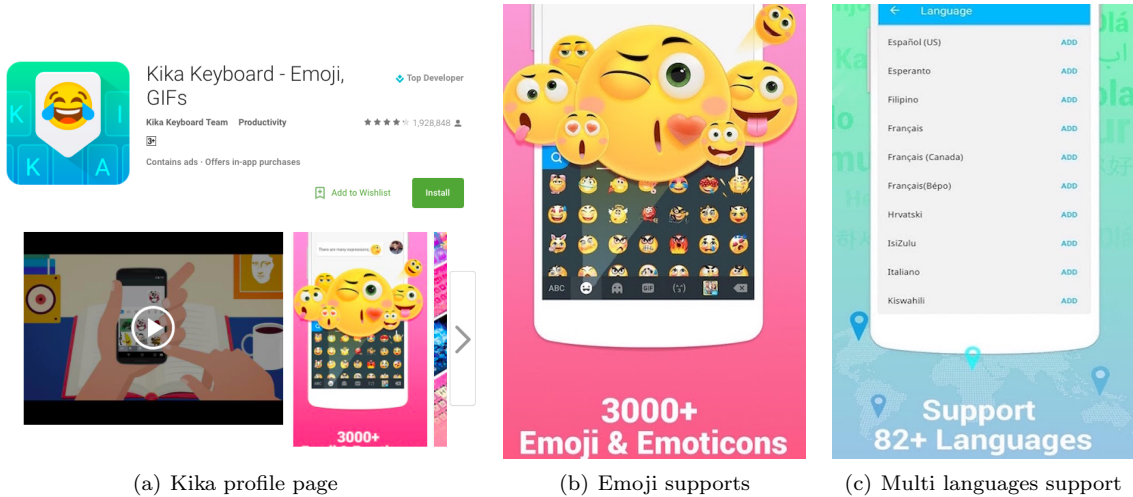| (a) Kika profile page | (b) Emoji supports | (c) Multi languages support |

Figure 1: The Snapshots of Kika Keyboard. Fig. 1(a) describes the profile page on Google Play; Fig. 1(b) illustrates the number of emojis supported in Kika; Fig. 1(c) illustrates the multiple languages supported in Kika

leads us to believe that females are more emotionally expressive than males [46]. The intense expressivity of emotion is also reflected in non-verbal communication. There are considerable prior studies reported that females are more non-verbally expressivity than males [4, 5, 6]. Females are evidenced to show a greater number of facial activity than males [47, 48] and observers can identify emotional states from female more accurately than from male faces [49]. Further, as the growth of the "facial expression" (emoticons) in text, researchers started to investigate the relationship between gender and emoticon usage and found female superiority in emoticon usage [18, 17, 16]. Now, in the times of emojis, there are very limited studies on gender and the new kind of non-verbal cues as is mentioned above. In this paper, we devote to enriching the findings of gender impacts on the rising non-verbal cues – emoji.

3 The Data Set

In this section, we briefly describe our data set and how we process the data with strict ethical considerations.

3.1 Data Collection and Description

Following our previous study in emoji usage [1], this paper still uses the data set that is collected via the Kika Keyboard[3] (abbreviated as Kika in the rest of this paper), a leading Android input method app in Google play (Fig. 1). As a free third-party keyboard supporting 82 languages and more than 3,000 emojis and emoticons, it has gained millions of downloads and installs across the world, and was ranked as the top 25 most downloaded apps of Google play in 2015. During the period of data collection, Emoji v4.0[4] (containing 2,389 emojis) was the latest released version by Unicode Standard. We use it as the list of candidate emojis to search our data set. Finally, we captured a total of 1,356 different emojis from the corpus.

To improve the user experience, Kika explicitly notifies its users that some information will be collected, such as the user profile information (optional at user registration) and the user-input messages

---

[3]https://play.google.com/store/apps/details?id=com.qisiemoji.inputmethod
[4]http://unicode.org/emoji/charts/full-emoji-list.html. Note that the emoji list has recently been updated. The latest release version is Emoji v5.0 now.

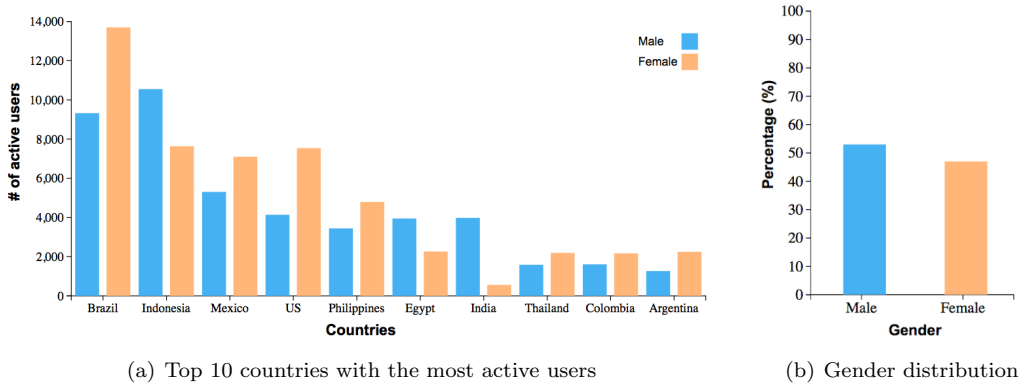(a) Top 10 countries with the most active users   (b) Gender distribution

Figure 2: Demographic distribution of users. Fig. 2(a) illustrates the top 10 countries with the most active users and their gender distributions in the data set; Fig. 2(b) shows the gender distribution in the aggregate

(defined as the content user typed before "Send" action) for analysis. However, such information can be collected from only the users who agree with the user-term statements. As is declared in Kika's Privacy Policy[5], no passwords or sensitive data are recorded and all data have been anonymized before the analysis.

As an input method which can run at a system-wide level, Kika can collect data that are not limited to particular applications such as Twitter (compared to studies using Twitter data [25, 20, 43]). Moreover, Kika can capture the contexts in other apps where it is enabled and launched. Hence, it enables us to make more comprehensive analysis of emoji usage.

The data covers 134,419 active users who offer their profile information. These users come from 183 countries and regions and their 401 million messages from December 4, 2016 to February 28, 2017. All active users posted at least one message through Kika keyboard during this period. We use an anonymized user identifier (device ID replaced with random string) to represent every single active user. For each user, we use the demographic information (including gender and the country where he/she comes from) and the messages typed by him/her in this period (with timestamps and the apps where the messaged were typed).

Then, we report the demographic distribution of users covered in our data set. As is illustrated in Fig. 2(a), the top 10 countries ranked by their active users are Brazil, Indonesia, Mexico, US, Philippines, Egypt, India, Thailand, Colombia, and Argentina. The country distribution is heavily long-tailed, so that a small fraction of countries have a disproportionately large share of users. Users from these 10 countries constitute about 70.9% of all users. However, in the aggregate, we have comparable male and female users. As shown in Fig. 2(b), 47% of the users are male and the other 53% are female.

3.2 User Privacy and Ethical Consideration

Indeed, our study is based on the sensitive gender information of users, hence we take careful steps to preserve the ethics in our research. First, our work is approved by the Research Ethical Committee of the institutes (a.k.a, IRB) of the authors. In the entire life cycle of this study, we consider user privacy as a critical concern. Kika replaced the device ID of each user with a randomized string before storage, so one can not identify any individual user with information in the data set. All the data is stored on a private, HIPPA-compliant cloud server, with strict access authorized by Kika. Although we calculated the message length in Section 6 and labelled emojis with sentiment by means of analyzing the contexts in Section 7.1, our analysis pipeline was entirely governed by Kika employees to ensure the compliance

---

[5]http://www.kika.tech/privacy/

with the public privacy policy stated by Kika. For other analysis, we removed all textual contents and extracted only the metadata for research.

3.3 Limitations

Indeed, every empirical study could have some limitations in the used data set, which can potentially affect the analysis along with the derived results. Note that our data set only contains the users who voluntarily share the gender information, and the user distribution is not uniform among all countries. Hence, there could exist some selection bias. However, to the best of knowledge, our data set is the largest to date and contains a representative number of users whose gender profiles are known. In particular, compared to previous HCI research that studied gender difference based on limited or controlled user groups [50, 51], e.g., by questionnaire from volunteers, the scale of our data set can make our analysis more comprehensive and the results statistically meaningful.

4 Methodology

Our empirical study aims to explore whether there exist some differences in emoji usage between male and female users. As an emerging popular CMC fashion, emoji is considered to be quite semantically rich, and can imply some sentiment even without any texts. In our study, we mainly make descriptive analysis by employing various statistic measures such as OLS regression, two-tailed z-test, and Pairwise Mutual Information.

We focus on the following analysis.

- **Usage Pattern Analysis**. We first make a macro-level descriptive analysis of emoji-usage patterns between male and female users from various aspects, including frequency pattern, consecutive/discrete pattern, co-used pattern, and so on.

- **CMC Scenario Analysis**. Beyond the macro-level analysis of emoji usage, we further focus on the CMC scenario analysis, i.e., whether the emoji usage can vary in different communication situations. Such an analysis is motivated by the fact that emoji can be used in both public apps (e.g., Twitter) and private communication apps (e.g., WhatsApp), so it is interesting whether the emoji usage can reflect gender difference under different contexts.

- **Temporal-Sentiment Analysis**. Emoji is widely used to express sentiment beyond plain text only. We finally explore the emoji usage along with the potential sentiment of male and female users. More specifically, we focus on the difference from a temporal perspective, as time is an important incentive factor that can lead to the changes of sentiment.

5 Usage Pattern Analysis

From the data set, we observe 30.5 million messages containing at least one emoji, accounting for 7.6% of all messages. Beyond the overall popularity of emojis, however, we are curious if males and females use emojis differently. How often do male and female users tend to use emojis in text messages? Which emojis do male and female users tend to use, respectively? Are there any input patterns of using emojis? In this section, we begin with some macro-level descriptive analysis of gender difference in emoji usage.

5.1 Frequency of Emoji in Messages

Previous studies have suggested that females are more non-verbally expressive than males [4, 5, 6]. More specifically, gender difference has been observed in the frequency of emoticon usage [11, 16]. Since emojis are another type of non-verbal cue, could we observe a similar difference in their usage frequency? We begin with examining whether males or females use emojis more frequently in their messages. That is,

Table 1: Male and female frequency of emoji usage. ("**": p-value <0.01)

Gender	%emoji-msg(in total)	%emoji-msg(2016.12)	%emoji-msg(2017.01)	%emoji-msg(2017.02)
Male	7.02	6.88	6.84	7.45
Female	7.96	7.78	7.77	8.41
z-score	-347.58**	-197.79**	-215.37**	-182.97**

Table 2: Emoji usage patterns. ("**": p-value <0.01 after Bonferroni correction)

	Male (%)	Female (%)	z-score
Pat1	39.50	40.61	-59.87**
Pat2	57.55	56.09	77.96**
Pat2.1	37.85	36.82	55.87**
Pat3	10.16	10.56	-35.08**

we compare the percentage of messages containing emojis (%emoji-msg) between male and female users, both for the entire time range and in each month.

As shown in Table 1, in general, the 7.02% of messages sent by male users contain at least one emoji, while the percentage for female user is 7.96%. The difference is significant under two-tailed *z-test*[52] and holds in each month during the period. Therefore, we can conclude that females have higher proportion of messages containing emojis in their input.

> **Finding (F1)**: In general, female users are more likely to use emojis in text messages than males. This superiority is reflected from female higher proportion of messages containing emoji.

5.2 Consecutive/Discrete Usage Patterns

Above, we looked at the message-level difference, yet those emoji-messages aren't quite the same. Some may have only one emoji, while some other may contain multiple emojis. We would like to explore if males and females have different input patterns of using emojis. Below are the three typical patterns how users input emojis:

- *Pat1:* Only one emoji in one message, e.g., I ❤️ you.

- *Pat2:* Consecutive use of multiple emojis in one message, e.g., I love you❤️❤️.

- *Pat3:* Discrete use of multiple emojis in one message, e.g., ❤️I love you❤️.

Note that, the *Pat2* and *Pat3* are not mutually exclusive. A user may use emojis in one message both consecutively and discretely (e.g., ❤️ I love you❤️❤️). Furthermore, a consecutive use of the same emoji is considered to be an emphasis and thus can imply much stronger sentiment [21]. Similarly, people sometimes lengthen words to emphasize their sentiment, such as "cooooooooooooollllll" [53] and females are reported to have a higher frequency of use this skill than males in Twitter [54]. Additionally, users often lengthen the "mouth" of emoticon to indicate a strong affect, such as :)) [8]. Inspired by these previous lessons, we also focus on male and female consecutive use of the same emoji. We define the pattern as:

- *Pat2.1:* Consecutive use of the same emoji in one message.

For both male and female users, we compute the proportion of the four patterns in the emoji-messages. Then we conduct the *z-test* to measure the gender differences in usage of these patterns. We apply Bonferroni correction [55] to adjust p-values for multi-hypothesis testing. The results are summarized in Table 2.
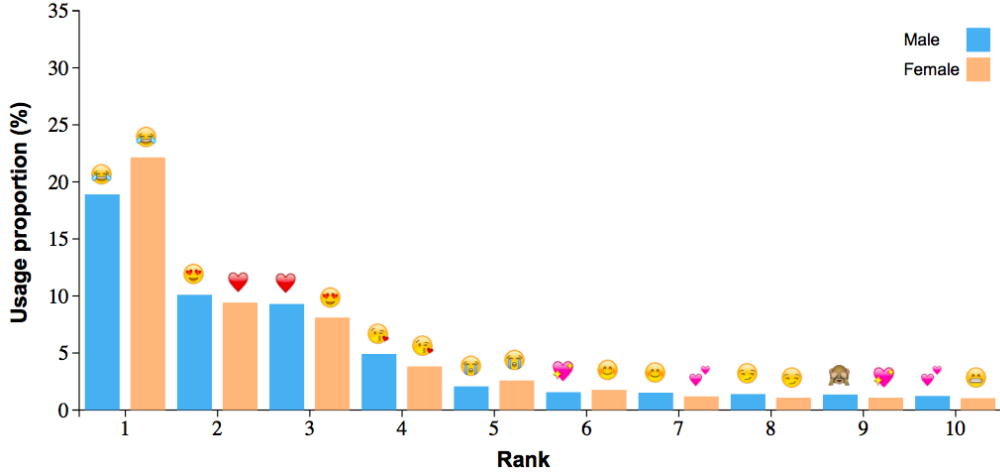
Figure 3: The top 10 most used emojis in male and female users

From Table 2, we see interesting differences between female and male users. Female users are more likely to use only one emoji or discretely use multiple emojis in a message. In the contrast, male users are more likely to consecutively use multiple emojis. Such a preference for consecutive emojis holds true even if we focus solely on the consecutive use of the same emoji.

Finding (F2): The emoji usage patterns are different between genders. Female users are more likely to use only one emoji or discretely use multiple emojis in a message, while male users tend to consecutively use multiple emojis. Furthermore, male users are more likely to consecutively use the same emoji to reinforce their sentiment.

5.3 The Choice of Emojis

We have found gender differences in *how* emojis are used in messages. Yet little do we know about *which* emojis are used. Do men and women have the same "goto" emojis? Do they co-use emojis in the same way? In this part, we will examine if male and female users differ in their choice of emojis. We not only investigate frequently used emojis or emoji categories, but also explore which emojis are likely to be co-used by male and female users.

5.3.1 Frequently Used Emojis

We start by examining the favorite emojis of male and female users. For each emoji, we calculate the percentage of its occurrence in all emoji occurrence used by male or female users, and summarize the ten most frequently used emojis in Fig. 3. Both emoji usage distribution of males and females are long-tailed. A small set of emojis have a disproportionately large share of all occurrence. There are not significant differences in the top 10 frequently used emojis. The top 5 emojis are even the same for male and female users, i.e., 😂 (face with tears of joy), ❤️ (red heart), 😍 (smiling face with heart-eyes), 😘 (face blowing a kiss), and 😭 (loudly crying face). However, we do notice differences for particular emojis. For example, although 😂 is the most popular emoji in both males and females, it accounts for 18.9% in male emoji usage, but 22.1% in females, with a difference of 3.2%. This difference is non-negligible given the heavily skewed distribution.

Finding (F3): Male and female users share the same top 5 most used emojis. However, there is obvious difference in usage proportion of some emojis.

Table 3: Face-related and heart-related emojis. ("**": p-value <0.01)

	Examples	Male usage (%)	Female usage (%)	z-score
Face-related emojis	😊 😔 😡	56.11	58.17	-201.60**
Heart-related emojis	❤️ 💗 💙	19.41	17.62	222.01**

5.3.2 Frequently Used Categories

Above, we can find that the most popular emojis are about expression, being either *faces* or *hearts*. In our data set, we find 69 face-related emojis and 15 heart-related emojis in total. The face-related emojis emphasize the facial expressions through the eye, eyebrow, or mouth shape. Different shapes are used to express different affects (e.g., positive and negative) and meanings such as happy (😊), blue (😔), and angry (😡). Instead of reflecting emotions in the facial expression, the heart-related emojis emphasize the color and shape of heart (such as ❤️, 💗, and 💙) to convey the love and affects of users directly. These 84 emojis comprise 75.5% of total emoji usage for males and 75.8% for females. Indeed, these are the two most popular categories of emojis.

In the traditional verbal communication studies for the real life [5, 47, 48, 56], females are reported to show more facial-related activities than males. Previous studies also suggest that females are more likely to express love in real life [57, 58, 59]. Since emojis act as non-verbal cues in text communication, could we observe the similar behavior characteristics in their usage? In other words, are females more likely to use face- and heart-related emojis in text communication, just like their habits in real life? To answer these questions, we aggregate the usage of face and heart emojis, and see if males and females use them differently.

We calculate the proportion of face- and heart-related emojis in total emoji usage for both male and female users, and conduct a *z-test* to compare gender preferences for these emojis. Table 3 summarizes the results. Female users are significantly more likely to use face-related emojis. Such an observation could be interpreted by previous studies on verbal communication. However, we are surprised to find that male users are more likely to use heart-related emojis than females in online communication. Such an observation is contrary to psychological literature where males are reported to be less willing to express love in real life. This finding implies that males reserve to express their love in the forms of verbal and text in real life, but they turn to the ubiquitous language, emoji, to supplement their love expression for online communication.

> **Finding (F4)**: Females are more likely to use *face-related* emojis than males, while male users are more likely to use *heart-related* emojis than females.

5.3.3 Co-used Emojis

We have already demonstrated gender difference in using the popular emojis. Yet we are also interested in the difference at the long tail. Since the distributions of the long-tail emojis are small and trivial, it is not straightforward to directly characterize such difference. Therefore, we group similar emojis together by their co-occurrences and quantitatively characterize the difference between genders.

We leverage male and female data to cluster emojis that are frequently used together, respectively. We use Point Mutual Information (PMI) [60] to measure the co-occurrence of every two emojis. The PMI of two emojis e_1 and e_2 can be computed as

$$\text{PMI}(e_1, e_2) = log \frac{p(e_1, e_2)}{p(e_1) * p(e_2)}$$

where $p(e_1)$ represents the usage frequency of e_1, $p(e_2)$ represents the usage frequency of e_2, and $p(e_1, e_2)$ represents the frequency of the co-occurrences. A larger PMI indicates the two emojis are more likely to occur together. We use the PMI of every two emojis to build a network for males and females,

respectively. We connect each emoji to five emojis that have the largest positive PMI with it. In such a network, each node represents an emoji and the weight of each edge is PMI of the two nodes.

We then perform community detection using the classic Fast Unfolding algorithm [61] and split the network into several communities. The nodes within one community have more connections (larger PMI) with each other, while the nodes from different communities have fewer connections (lower PMI). We detect 56 communities for males and 55 communities for females. Then we analyze these communities and find some interesting phenomena. For example, males like to use sport-related emojis (such as ⚽ and 🏀) with 🏅 and 🏆, while female prefer to use the sport-related emojis with 🎿, 🥤 and 🥤. What's more, females like to co-use the clothes-, shoes-, and bag-related emojis with 🛍 (shopping bags), while male users do not like this usage.

> **Finding (F5)**: Male and female users are different in emoji co-occurrences. For example, males like to use sport-related emojis with 🏅 and 🏆, while female prefer to use the sport-related emojis with 🎿, 🥤 and 🥤. Females like to co-use the clothes-, shoes-, and bag-related emojis with 🛍 (shopping bags), while male users do not like this usage.

5.4 Summary of Findings

From the findings **F1**-**F5** derived from our descriptive analysis, we can have a basic understanding of how emojis are used by females and males. Generally, female and male users can statistically have quite significantly different behaviors in using emojis. Such findings evidence our initial hypotheses. We further explore whether there could exist some complex context-aware usage that could be reflected by gender differences. In the following sections, we choose typical contexts from two major aspects, i.e., communication scenario and temporal-sentiment.

6 CMC Scenario analysis

Based on the macro-level descriptive analysis, we have demonstrated how males and females use emoji differently. However, we argue that knowing the general gender difference is definitely not enough, but can be explored more deeply. Since users widely interweave emojis in daily life and work, it is more interesting to investigate how users use emojis under different contexts. In this section, we raise a hypothesis that the usage of emojis can vary between genders under different CMC scenarios. Such a hypothesis is motivated from two folds. First, female and male users could have different representation preferences in CMC. Second, even though there have been some non-emoji-oriented studies exploring gender differences in CMC, most of them are conducted at the public channel such as Twitter. It is unclear whether the results can be still consistent for private channels such as WhatsApp. Since Kika keyboard is a system-wide app, it makes us capable of conducting such a study.

For simplicity, we choose two popular representative apps from these two categories, i.e., Twitter and Whatsapp, respectively. *Accordingly, we only take US users into account in this section.* For public communication scenarios, we focus on the messages collected from Twitter, since Twitter is a popular CMC channel for propagating information [62]. For private communication scenarios, we select messages collected from WhatsApp. Although Twitter can also be used to make private conversations now, WhatsApp is relatively more frequently used for private communication purpose. We investigate the emoji usage of male and female users in these two scenarios, respectively.

We measure the frequency of emoji usage using the metric %emoji-msg defined in Section 5. Table 4 displays the %emoji-msg and the *z-test* result of male and female users in public and private communication. In Twitter, it is observed that females are more likely to use emojis compared to males. Such a finding is consistent with the previous work that has discovered that females have a higher frequency of emoticon use than men in Twitter [54].

In contrast, it is interesting to find that males tend to use emojis more frequently than females in private communication scenario. In order to better understand this phenomenon, we further analyze the

Table 4: Frequency (%emoji-msg) in different scenarios ("**": p-value <0.01)

	Male	Female	z-score
Twitter	11.18	12.21	-4.67**
WhatsApp	7.92	7.05	29.44**

Table 5: OLS Regression: Gender and number of emoji ("**": p-value<0.01)

	Dependent variable	
	# of emoji in a message	
	(1) WhatsApp	(2) Twitter
Gender	-0.1287**	0.0573
Message length (excluding emojis)	0.1438**	0.0263**
Constant	2.3298**	1.8226**
Observations	255,490	11,455
R^2	0.114	0.003

number of emojis in each message typed in the two scenarios. We perform regressions to analyze the gender difference and set *Gender* as a dummy variable (Male=0, Female=1) . Since one may expect that more emojis are used in longer messages, we control the message length excluding emojis in the regressions. Then we conduct the multiple-variable Ordinary Least Square regression [63] for Twitter and WhatsApp, respectively. Regression results are summarized in Table 5. As expected, the longer the message is, the more emojis it contains. The coefficients of the dummy variable *Gender* are negative and significant in the regression on the WhatsApp data. Such a result indicates that males prefer using more emojis than females, given the same length of a message in WhatsApp. Combining the finding shown above, we can conclude that males not only use emojis more frequently, but also tend to use more emojis in one message than females in private communication. However, in Twitter, gender does not have such an impact on the number of emojis used in one message.

In addition, at the first sight of Table 4, we may get the hasty conclusion that both males and females are more likely to use emojis in public communication. In other words, emojis are more popular in information propagation (i.e., Twitter) than in conversations (i.e., WhatsApp). However, previous work has reported that emoticons are more popularly used in conversations than in information propagation [8]. Do emojis and emoticons differ in this dimension? With this doubt, we consider the potential interpretation behind this phenomenon.

As is shown above, message length could be an influential factor to emoji usage, so it is unconscionable to compare the emoji usage in public and private communication directly. We can not rule out the possibility that users tend to send longer messages in Twitter than in WhatsApp (median length of messages in Twitter / WhatsApp is 5 words / 3 words, respectively), resulting in the higher likelihood to use emojis in Twitter. Therefore, we measure the frequency by the number of emoji in one message. We define a dummy variable *Type* (Private communication=0, Public communication=1) to distinguish the two scenarios. For male and females users, we conduct OLS regressions, respectively. We report the results of these regressions in Table 6. When typing messages of the same length, both male and female users are more likely to use more emojis in private communication than in public one. From this perspective, indeed, emojis are more popularly used in private communication than in public communication.

Finding (F6): Females and males can perform quite differently in emoji usage under different CMC scenarios. Females are more likely to use emojis than males in public communication. However, in private communication, males are more active in emoji usage, both in the frequency of using emoji and the number of emojis contained in one message. Emojis are more popularly used in private communication than in public communication.

Table 6: OLS Regression: Communication scenarios and number of emoji in one message ("**": p-value<0.01)

	Dependent variable	
	# of emoji in a message	
	(1) Male	(3) Female
Type	-1.1749**	-1.1964**
Message length (excluding emojis)	0.1832**	0.1187**
Constant	2.0370**	2.4369**
Observations	104,606	165,786
R^2	0.171	0.084

7 Temporal-Sentiment analysis

Finally, we aim to examine the emoji usage with temporal-sentiment context. Since emojis can imply richer semantics and sentiment [21], analyzing the sentiment of emojis provides a more detailed summarization of emoji usage. In addition, usually, it is argued that time and social events can be incentives to affect sentiment changes. As a result, we explore whether emojis can play as an new indicator of sentiment to sense male and female response to some specific time or events, e.g., weekends and festivals.

7.1 Inferring Sentiment behind Emojis

To sense users' response from the emoji usage, at first, we need to infer the sentiment behind emojis. As is summarized in Section 2, there are three main approaches to measure the sentiment or semantic representations of emojis: (1) based on the sentiment annotated by participants, (2) based on the their textual descriptions and annotation tags defined by the Unicode Standard, and (3) based on the context in which the emojis occur. As previous work has evidenced that everyone can interpret emojis in his/her own way [3], we do not adopt the first approach. What's more, the official descriptions are so simple that sometimes they can not provide the sentiment of emojis. For example, the Unicode Consortium defines the emoji ❤ as "red heart". We employ a text analysis tool, named LIWC (Linguistic Inquiry and Word Count)[6], to validate the sentiment of the text "red heart" and find that there is no sentiment conveyed in it. The short description only describes the ideogram but covered up its emotional contents. However, we often use the emoji ❤ to convey certain sentiment such as "I ❤ you". Therefore, it seems to be unreasonable to take the description as users' interpretation. We finally adopt the third approach and analyze the emoji sentiment from their contexts. Novak et al. [43] engaged human annotators to label the tweets that contained emojis and further inferred the sentiment of emojis from the sentiment of the tweets that they are used in. However, it is so challenging to annotate in the same way due to the large scale of our data set. Instead, we choose to apply a embedding model to project words and emojis onto the same high-dimensional vector space as in [23, 44, 45] and then use the vector space to generate the semantic representations of emojis.

We choose LINE [64], a network embedding model, to compute the similar words of each emoji from the context. It should be noted that we do not have the expertise in processing multiple languages other than English. Prior work has reported that some emojis are interpreted in a different way from language to language [25]. In order to rule out the impacts of culture and language on emoji interpretation, we only focus on English messages collected from American users to analyze the emoji sentiment in this section. We follow the approach in [45], setting LINE with the *second-order proximity* (LINE-2nd) to get semantic representation of each emoji. First, we construct a co-occurrence graph from the corpus to represent the semantic structure. In the co-occurrence graph, each node represents a token (it could be a

---

[6]http://liwc.wpengine.com

Table 7: Categorization of emojis with sentiment

Affect	Condition	# of emojis
POS	$S_{posemo}>S_{negemo}$	747
NEU	$S_{posemo}=S_{negemo}>0$	150
NEG	$S_{posemo}<S_{negemo}$	352

word, an emoticon, or an emoji), and the edge between the nodes represents the co-occurrence of the pair of tokens. The *second-order proximity* between two nodes is the similarity between their neighborhood network, describing how likely the two tokens occur in contexts. For example, we often use "I ♥ you" to replace "I love you". In this case, "love" is a similar word of ♥ because they have similar "neighborhood" (contexts). We then compute the euclidean distance of two tokens in the embedding space to represent their semantic similarity. For each emoji, we use a *k*-near-neighbors algorithm (*k*NN) to get the most semantically similar words in embedding space, where k is set to 150, as its semantic representation. We further employ LIWC to calculate the sentiment for the semantic representation of each emoji. LIWC has language limitations and can not analyze text from all languages. It is another reason why we only focus on English messages in this subsection. We only consider two measures of affect derived from LIWC: *posemo* (positive affect) and *negemo* (negative affect). After applying LIWC, semantic representations of 1,249 emojis are classified with sentiment. In other words, we obtain sentiment scores of the 1,249 emojis. We further label each of them with one of the three sentiment polarity: positive (POS), neutral (NEU), and negative (NEG). The result of categorization is summarized in Table 7. Most of these emojis are positive, which is consistent with previous work [1, 43].

7.2 Emoji Response to Some Specific Time

Based on the sentiment of emojis, we then move to the temporal analysis of emoji usage along with the potential sentiment of male and female users, respectively. As is mentioned above, we focus on US in this section. We use the timestamps of messages as the time indicator (when the message containing the emojis are posted). We have introduced in Section 3 that the timestamps of messages are collected as the server-side time (GMT-8). We can not determine the real system time when the messages are posted based on the country information because US spans four different time zones (including GMT-8) and we don't have more detailed location information of each user. However, if we conduct daily-grained analysis, this time bias can be negligible. For each day in the period that our data set spans (December 4, 2016 to February 28, 2017), we calculate the daily sentiment of male and female users reflected from emoji usage, respectively. We derive day-to-day sentiment by counting positive and negative emojis. We define the positive score $s_{d,g}$ on day d of gender g as the ratio of positive versus negative emojis, counting from that day's messages typed by g:

$$s_{d,g} = \frac{\text{count}_d \text{ (pos. emoji} \wedge g)}{\text{count}_d \text{ (neg. emoji} \wedge g)} = \frac{p \text{ (pos. emoji} \mid d, g)}{p \text{ (neg. emoji} \mid d, g)}$$

We then leverage the results of the daily emoji usage to investigate the male and female users' temporal "emoji response" to some specific time or events, e.g., weekends and festivals.

7.2.1 At Weekdays and Weekends

We first aggregate the daily results and look at the differences of emoji usage in weekdays and weekends. As shown in Table 8, both male and females are more likely to use relatively more positive emojis in weekends than in weekdays. It implies that people generally tend to use more positive emojis at non-work time. What's more, male users seem to be more sensitive to weekends in emoji usage as the growth rates of positive scores are higher than females. This may be correlated with the different kinds of stress that males and females suffer from. In previous work, males are reported to mainly feel the stress from work while females have more daily stress such as family and health-related events [65]. At weekends, males

Table 8: Positive scores in weekdays and weekends

	Weekdays	Weekends	Growth rate (%)
Male	15.85	18.38	15.96
Female	10.89	12.61	15.79

Table 9: Rank of Christmas-related emojis in male and female usage

	🎄		🎅	
	Male	Female	Male	Female
December 24, 2016	12	7	>20	11
December 25, 2016	9	5	>20	10

may be temporarily away from work pressure (the main pressure in their mind), but females still suffer from their annoying family and health stress although they don't need to go to work at weekends, either.

Finding (F7): Both males and females use emoji more positively in weekends compared to weekdays, especially males.

7.2.2 Near Festival

To investigate the "emoji response" of males and females to festivals, we select the Christmas. We present the variation tendency of positive scores in December, 2016 in Fig. 4 to help us understand the Christmas' influence that can be reflected by emojis. In Fig. 4, we can find the obvious sentiment fluctuation near Christmas for both males and females. From December 23 to 25, the sentiment of males and females became more and more positive compared to other periods. Both the male and female users were immersed in the festive joy. However, females seem to have more serious so-called "*pre-holiday blue*". "Pre-holiday blue" means that people get stressed and depressed about the holidays before they arrive. We can see that males underwent a decrease of positive scores from December 21 to 23, while the scores of females started to decrease much earlier, i.e., from December 18, a week before Christmas. In addition, both males and females underwent quite obvious "*post-holiday blue*" reflected from their decrease in positive scores after Christmas.

Another interesting finding is that people's enthusiasm towards the Christmas is likely to be reflected by those festival-related emojis. We calculate the most used 20 emojis of male and female users per day in the three months. On December 24 and 25, we find the obvious increase in the usage of 🎄 and 🎅, especially for female users. We list the rank of the two emojis (ranked by the usage frequency) in December 24 and 25, for male and female users, respectively. As is illustrated in Table 9, both males and females are *sensitive* to Christmas as they use the two emojis more frequently on December 24 and 25 (the two emojis don't occur in the top 20 emoji list except the two days). Additionally, female users seem to be more sensitive, reflected from the rank.

Finding (F8): Both males and females are relatively more positive and use Christmas-related emojis (🎄 and 🎅) more frequently on December 24 and 25, especially female users. Additionally, male and females users have obvious pre- and post-holiday blue, especially females.

---

8 Discussion

So far we have demonstrated the emoji-usage behaviors through a gender lens and indeed found some gender-specific differences. We then discuss some implications following our previous findings, and try to explore some possible opportunities for app developers. Meanwhile, we analyze the potential threats that could affect the results of this study.
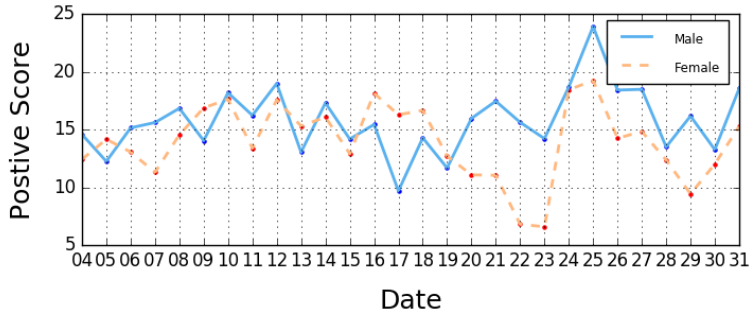
Figure 4: Positive scores in December, 2016. *Note:* For simplicity, we use *day* to represent *2016-12-day* in the figures

8.1 Implications

8.1.1 Improving Emoji Keyboard Layout

As we have found some gender differences in emoji usage, the most intuitive and straightforward implication is to improve user experiences on current smartphone keyboard. In current OS-native and third-party input methods, emojis are always shown in paging control and each page contains some emojis that are displayed in a rather fixed layout. When we want to type in an emoji, we need to swipe left or right to search for it. This approach could be problematic as the number of emoji grows, which attracts many researchers to optimize emoji entry speed such as [66]. Indeed, our analysis can provide some hints for smartphone input-method developers, not limited to Kika but also including other keyboard developers or even OS vendors, to optimize their keyboard layout. For example, besides those top emojis such as 😂, we find that females prefer face-related emojis while males prefer heart-related emojis. Therefore, the ranking list of emojis shown on the keyboards' layout should be more gender-aware to users. Additionally, current keyboards can recommend the possible words or emojis that users may type in the next input. Based on our observations, keyboard developers can improve their algorithms from gender perspective instead of only listing the Most-Recently-Used emojis. In addition, we also find some scenario-related and temporal-related emoji usage patterns, which can further help keyboard developers improve their emoji recommendations under some specific contexts.

8.1.2 Non Privacy-Invasive User Profiling

User profiling is crucial for Internet service or content providers. Knowing the possible user profiles, such as gender, age, and other preferences, has been proved to be an efficient way that not only improves the user experiences, but also increases the accuracy of online ads recommendation to increase potential ad clicks and revenues. In the times of "app economy", app developers rely on in-app ads that require accurate user profiling [67]. Today's in-app ads are mostly irrelevant, justly derided as taking a "spray and pray" [68], which greatly affects the revenue of app developers. In the social applications such as facebook, a user is asked to fill in his/her profiles, e.g., typically at registration phase, so that others can better know his/her possible background. However, in most mobile applications, developers can not collect such information. Indeed, collecting users' typed texts in an input-method and using NLP techniques to analyze the user profile is a feasible approach. However, it may result in collecting some sensitive information and thus hurt user privacy. Note that the analysis in this paper does not touch any texts, but relies on the usage of emoji only. At least, it has been indicated that the inferred emoji usage can help distinguish a users' gender, and implies that the emoji usage could be a possible signal for inferring user profiles in a non privacy-invasive fashion. In our future work, we plan to synthesize the derived emoji-usage patterns and more information such as input traces and smartphone brands, and devise machine learning models for better user profiling. We believe that such efforts are exactly

15

actionable, as we can use the derived patterns as features to train the model and use the existing "gender-labelled" users as ground truth for validation. In addition, since smartphone keyboard is a system-wide application, it could expose some APIs to other applications that can predict their users profiles, which may result in a new business model.

8.2   Threats to Validity

Every empirical study can have its own limitations and thus affect the generalization of results. One major potential threat of this study could come from the coverage of our data set. We focus on the active users in the data set collected by the Kika keyboard. Indeed, most popular smartphone manufacturers support emojis in their built-in input methods, and there are also some other popular third-party input methods supporting emojis in the market. In our opinion, such a threat could be not so significant to the current macro-level analysis, as our data set covers a large number of real-world users from various countries and thus can promise the gender-specific study statistically comprehensive. However, the temporal and communication scenario analysis of emoji usage is conducted over only US users, hence the derived patterns can not be fully generalized to users from other countries.

Indeed, besides gender, there are still some confounding user-profile factors that may influence the emoji usage such as country and language. It would be interesting to synthesize such information with gender in our future work, and thus make the gender difference at a finer granularity.

9   Conclusion and future Work

In this paper, we have presented an empirical study of emoji usage through a gender lens. Our study was based on a unique and large data set collected by Kika – a popular input method app. The data set covers 401 million messages typed by 134,419 active users from 183 countries and regions over three months. 47% of these users are male and the other 53% are female. We conduct a multi-dimensional statistical analysis from various aspects of emoji usage, including the frequency, preferences, input patterns, CMC-scenario sensitive patterns, temporal patterns, and sentiment patterns. We applied rigorous statistical tests to ensure the credibility of our findings. Indeed, we demonstrate that there are exactly some gender differences in emoji usage. We drew on our observations and findings to present some implications such as distinguishing users' gender through their emoji usage and then providing more effective recommendations. To the best of our knowledge, our study is the first to qualitatively analyze emoji usage through a gender lens based on large-scale users and their realistic usage data.

In the future, we would like to further bridge the gap between gender and emoji usage. We plan to synthesize the derived emoji-usage patterns and more information such as input traces and smartphone brands, and devise machine learning models for better user profiling.

References

[1] X. Lu, W. Ai, X. Liu, Q. Li, N. Wang, G. Huang, and Q. Mei, "Learning from the ubiquitous language: An empirical analysis of emoji usage of smartphone users," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2016*, 2016, pp. 770–780.

[2] C. Tauch and E. Kanjo, "The roles of emojis in mobile phone notifications," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp Adjunct 2016*, 2016, pp. 1560–1565.

[3] H. Miller, J. Thebault-Spieker, S. Chang, I. L. Johnson, L. G. Terveen, and B. Hecht, ""blissfully happy" or "ready to fight": Varying interpretations of emoji," in *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*, 2016, pp. 259–268.

[4] S. L. Ablon, D. P. Brown, E. J. Khantzian, and J. E. Mack, *Explorations in affect development and meaning.* Routledge, 2013.

[5] L. Marianne and B. Mahzarin, "Toward a reconsideration of the gender-emotion relationship," *Emotion and Social Behavior*, vol. 14, pp. 178–201, 1992.

[6] B. N. J. and J. A. Hall., "Beliefs about female and male nonverbal communication," *Sex Roles*, vol. 32, no. 1, pp. 79–90, 1995.

[7] M. A, *Silent Messages (1st ed.).* Belmont, CA:Wadsworth, 1971.

[8] J. Park, V. Barash, C. Fink, and M. Cha, "Emoticon style: Interpreting differences in emoticons across cultures," in *Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM 2013*, 2013.

[9] W. J. B and D. K. P, "The impacts of emoticons on message interpretation in computer-mediated communication," *Social Science Computer Review*, vol. 19, no. 3, pp. 324–347, 2001.

[10] L. Shao-Kang, "The nonverbal communication functions of emoticons in computer-mediated communication," *CyberPsychology & Behavior*, vol. 11, no. 5, pp. 595–597, 2008.

[11] E. Dresner and S. C. Herring, "Functions of the nonverbal in CMC: Emoticons and illocutionary force," *Communication theory*, vol. 20, no. 3, pp. 249–268, 2010.

[12] D. Daantje, B. A. ER, and V. G. Jasper, "Emoticons and online message interpretation," *Social Science Computer Review*, 2007.

[13] F. Ruth, urcan Alexandra, T. Dominic, H. Nicole, D. Harriet, and T. Amelia, "Sarcasm and emoticons: Comprehension and emotional impact," *The Quarterly Journal of Experimental Psychology*, vol. 69, no. 11, pp. 2130–2146, 2016.

[14] M. K. M and O. Sae, "Pragmatic play? some possible functions of english emoticons and japanese kaomoji in computer-mediated discourse," in *Association of Internet Researchers Annual Conference*, vol. 8, 2007.

[15] S. E. Tchokni, D. Ó. Séaghdha, and D. Quercia, "Emoticons and phrases: Status symbols in social media," in *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014*, 2014.

[16] C. Tossell, P. T. Kortum, C. Shepard, L. H. Barg-Walkow, A. Rahmati, and L. Zhong, "A longitudinal study of emoticon use in text messaging from smartphones," *Computers in Human Behavior*, vol. 28, no. 2, pp. 659–663, 2012.

[17] H. H. Sung, "Gender differences in emoticon use on mobile text messaging: evidence from a korean sample," *International Journal of Journalism & Mass Communication*, vol. 2014, 2014.

[18] A. Wolf, "Emotional expression online: Gender differences in emoticon use," *Cyberpsy., Behavior, and Soc. Networking*, vol. 3, no. 5, pp. 827–833, 2000.

[19] P. Umashanthi and E. Jacob, "Emoticons vs. emojis on twitter: A causal inference approach," *arXiv preprint arXiv:1510.08480*, 2015.

[20] K. Ryan and L. Watts, "Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships," *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design*, 2015.

[21] H. Cramer, P. de Juan, and J. R. Tetreault, "Sender-intended functions of emojis in US messaging," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2016*, 2016, pp. 504–509.

[22] H. Tianran, G. Han, S. Hao, N. T. vy Thi, and L. Jiebo, "Spice up your chat: The intentions and sentiment effects of using emoji," in *Proceedings of the 11th International Conference on Weblogs and Social Media, ICWSM 2017*, 2017, p. to appear.

[23] H. Pohl, C. Domin, and M. Rohs, "Beyond just text: semantic emoji similarity modeling to support expressive communication 👫➡🏢😄," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 24, no. 1, pp. 6:1–6:42, 2017.

[24] R. Zhou, J. Hentschel, and N. Kumar, "Goodbye text, hello emoji: Mobile communication on wechat in china," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI 2017*, 2017.

[25] F. Barbieri, G. Kruszewski, F. Ronzano, and H. Saggion, "How cosmopolitan are emojis?: Exploring emojis usage and meaning over different languages with distributional semantics," in *Proceedings of the 2016 ACM Conference on Multimedia Conference, MM 2016*, 2016, pp. 531–535.

[26] L. Nikola and F. Darja, "A global analysis of emoji usage," *ACL 2016*, pp. 82–89, 2016.

[27] Y. Nishimura, "A sociolinguistic analysis of emoticon usage in japanese blogs: Variation by age, gender, and topic," *Selected Papers of Internet Research*, vol. 16, 2015.

[28] R. Wang, G. M. Harari, P. Hao, X. Zhou, and A. T. Campbell, "Smartgpa: How smartphones can assess and predict academic performance of college students," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2015*, pp. 295–306.

[29] C. Orellana-Rodriguez, E. Diaz-Aviles, and W. Nejdl, "Mining emotions in short films: User comments or crowdsourcing?" in *Proceedings of the 22nd International World Wide Web Conference, WWW 2013*, 2013, pp. 69–70.

[30] D. Sanchez-Cortes, J. Biel, S. Kumano, J. Yamato, K. Otsuka, and D. Gatica-Perez, "Inferring mood in ubiquitous conversational video," in *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia, MUM 2013*, 2013, pp. 22:1–22:9.

[31] E. A. Carroll, M. Czerwinski, A. Roseway, A. Kapoor, P. Johns, K. Rowan, and M. M. C. Schrae-fel, "Food and mood: Just-in-time support for emotional eating," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*, 2013, pp. 252–257.

[32] J. M. dos Reis Costa, A. T. Adams, M. F. Jung, F. Guimbetière, and T. Choudhury, "Emotioncheck: Leveraging bodily signals and false feedback to regulate our emotions," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2016*, 2016, pp. 758–769.

[33] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "Moodscope: building a mood sensor from smart-phone usage patterns," in *The 11th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys 2013*, 2013, pp. 389–402.

[34] J. Hernandez, M. E. Hoque, W. Drevo, and R. W. Picard, "Mood meter: Counting smiles in the wild," in *The 2012 ACM Conference on Ubiquitous Computing, Ubicomp 2012*, 2012, pp. 301–310.

[35] C. Coutrix and N. Mandran, "Identifying emotions expressed by mobile users through 2d surface and 3d motion gestures," in *The 2012 ACM Conference on Ubiquitous Computing, Ubicomp 2012*, 2012, pp. 311–320.

[36] J. F. Filho, T. Valle, and W. Prata, "Non-verbal communications in mobile text chat: emotion-enhanced mobile chat," in *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services, MobileHCI 2014*, 2014, pp. 443–446.

[37] W. Wolny, "Emotion analysis of twitter data that use emoticons and emoji ideograms," in *Information Systems Development: Complexity in Information Systems Development - Proceedings of the 25th International Conference on Information Systems Development, ISD 2016*, 2016.

[38] J. Zhao, L. Dong, J. Wu, and K. Xu, "Moodlens: An emoticon-based sentiment analysis system for chinese tweets," in *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2012*, 2012, pp. 1528–1531.

[39] T. G. W and F. D. R, "Oh that's what you meant!: Reducing emoji misunderstanding," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, MobileHCI Adjunct 2016*. ACM, 2016, pp. 859–866.

[40] 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 *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, 2017, p. to appear.

[41] E. Ben, R. Tim, A. Isabelle, B. Matko, and R. Sebastian, "emoji2vec: Learning emoji representations from their description," *arXiv preprint arXiv:1609.08359*, 2016.

[42] D. Thomas, "Emojineering part 1: Machine learning for emoji trends," *Instagram Engineering Blog*, 2015.

[43] P. K. Novak, J. Smailovic, B. Sluban, and I. Mozetic, "Sentiment of emojis," *PloS One*, vol. 10, no. 12, 2015.

[44] F. Barbieri, F. Ronzano, and H. Saggion, "What does this emoji mean? A vector space skip-gram model for twitter emojis," in *Proceedings of the 10th International Conference on Language Resources and Evaluation LREC 2016*, 2016.

[45] W. Ai, X. Lu, X. Liu, N. Wang, G. Huang, and Q. Mei, "Untangling emoji popularity through semantic embeddings," in *Proceedings of the 11th International Conference on Weblogs and Social Media, ICWSM 2017*, 2017, p. to appear.

[46] K. A. M and G. A. H, "Sex differences in emotion: Expression, experience, and physiology," *Journal of Personality and Social Psychology*, vol. 74, no. 3, p. 686, 1998.

[47] B. Ross, B. R. M, G. Nancy, and S. Beth, "Unitization of spontaneous nonverbal behavior in the study of emotion communication," *Journal of Personality and Social Psychology*, vol. 39, no. 3, pp. 522–529, 1980.

[48] B. Ross, B. Reuben, and B. Dana, "Temporal organization of spontaneous emotional expression: A segmentation analysis," *Journal of Personality and Social Psychology*, vol. 42, no. 3, pp. 506–517, 1982.

[49] B. Ross, M. R. E, and C. W. F, "Sex, personality, and physiological variables in the communication of affect via facial expression." *Journal of Personality and Social Psychology*, vol. 30, no. 4, p. 587, 1974.

[50] X. Sun, Q. Zhang, S. Wiedenbeck, and T. Chintakovid, "Gender differences in trust perception when using IM and video," in *Extended Abstracts Proceedings of the 2006 Conference on Human Factors in Computing Systems, CHI 2006*, 2006, pp. 1373–1378.

[51] P. Mudliar and N. Rangaswamy, "Offline strangers, online friends: Bridging classroom gender segregation with whatsapp," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI 2015*, 2015, pp. 3799–3808.

[52] F. R. Aylmer, *Statistical methods for research workers*. Genesis Publishing Pvt Ltd, 1925.

[53] S. Brody and N. Diakopoulos, "Cooooooooooooooollllllllllllllll!!!!!!!!!!!!!! using word lengthening to detect sentiment in microblogs," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011*, 2011, pp. 562–570.

[54] P. Róisín, "Gender and emotional expressiveness: An analysis of prosodic features in emotional expression," *Pragmatics and Intercultural Communication*, vol. 5, no. 1, pp. 46–54, 2012.

[55] D. C. W, "A multiple comparison procedure for comparing several treatments with a control," *Journal of the American Statistical Association*, vol. 50, no. 272, pp. 1096–1121, 1955.

[56] B. R. W, S. V. J, M. R. E, and C. W. F, "Communication of affect through facial expression in humans," *Journal of Personality and Social Psychology*, vol. 23, no. 3, pp. 362–371, 1972.

[57] F. R. A and M. C. Lynn, "Gender and age stereotypes of emotionality," *Personality and Social Psychology Bulletin*, vol. 17, no. 5, pp. 532–540, 1991.

[58] B. J. O and P. C. W, "The inexpressive male: A tragedy of american society," *Family Coordinator*, pp. 363–368, 1971.

[59] W. Richard and G. Elisabeth, "Emotion expression and the locution i love you: A cross-cultural study," *International Journal of Intercultural Relations*, vol. 30, no. 1, pp. 51–75, 2006.

[60] C. K. Ward and H. Patrick, "Word association norms, mutual information, and lexicography," *Computational Linguistics*, vol. 16, no. 1, pp. 22–29, 1990.

[61] B. V. D, G. Jean-Loup, L. Renaud, and L. Etienne, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: Theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[62] S. Ye and S. F. Wu, "Measuring message propagation and social influence on twitter.com," in *Social Informatics - Second International Conference, SocInfo 2010*, 2010, pp. 216–231.

[63] W. J. M, *Introductory econometrics: A modern approach*. Nelson Education, 2015.

[64] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "LINE: large-scale information network embedding," in *Proceedings of the 24th International Conference on World Wide Web, WWW 2015*, 2015, pp. 1067–1077.

[65] M. M. Pilar, "Gender differences in stress and coping styles," *Personality and Individual Differences*, vol. 37, no. 7, pp. 1401–1415, 2004.

[66] H. Pohl, D. Stanke, and M. Rohs, "Emojizoom: emoji entry via large overview maps," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2016*, 2016, pp. 510–517.

[67] J. Gui, S. McIlroy, M. Nagappan, and W. G. J. Halfond, "Truth in advertising: The hidden cost of mobile ads for software developers," in *Proceedings of the 37th IEEE/ACM International Conference on Software Engineering, ICSE 2015*, 2015, pp. 100–110.

[68] S. Nath, F. X. Lin, L. Ravindranath, and J. Padhye, "Smartads: Bringing contextual ads to mobile apps," in *Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys 2013*, 2013, pp. 111–124.