From Adoption to Adaption: Tracing the Diffusion of New Emojis on Twitter

Anonymous ACL submission

Abstract

In the rapidly evolving landscape of social media, the introduction of new emojis in Unicode release versions presents a structured opportunity to explore digital language evolution. Analyzing a large dataset of sampled English tweets, we examine how newly released emojis gain traction and evolve in meaning. We find that community size of early adopters and emoji semantics are crucial in determining their popularity. Certain emojis experienced notable shifts in the meanings and sentiment associations during the diffusion process. Additionally, we propose a novel framework utilizing language models to extract words and pre-existing emojis with semantically similar contexts, which enhances interpretation of new emojis. The framework demonstrates its effectiveness in improving sentiment classification performance by substituting unknown new emojis with familiar ones. This study offers a new perspective in understanding how new language units are adopted, adapted, and integrated into the fabric of online communication.

1 Introduction

The language landscape of the Web era is ever-evolving, characterized by the emergence and evolution of new language units. Individuals have creatively crafted out-of-vocabulary language units, such as Internet memes and viral hashtags, to encapsulate and convey complex ideas, sentiments, and cultural phenomena, fostering shared lexicons that resonate across digital communities. Understanding the adoption and adaptation of these language units is crucial for gaining insights into the dynamic nature of online communication, the information diffusion process in social networks, and the underlying social trends and movements. However, analyzing the dynamics of these emerging language units in online communication presents unique challenges. These units lack universal conventions and standards and their characteristics may vary during diffusion, making it complex to track their initial appearances, early adoption, and frequency of use.

As a recent addition to this landscape, emojis offer a distinctive opportunity to explore the diffusion and evolution of new language units. Emojis are visual symbols that are embedded into text. These non-verbal symbols go beyond a single word or phrase, encapsulating rich semantics spanning a wide spectrum of emotions, actions, objects, and concepts. Originating as emoticons in the early Internet culture, emojis have evolved into a standardized and universally recognized visual language. Unlike other language units like hashtags or internet memes, emojis undergo a standardized process prior to their inclusion in the language. They are proposed to the Unicode consortium, formally defined by the Unicode standard, uniquely coded as Unicode strings, such as U+1F603 for emoji 😃, and then rendered by various platforms.

Since the Unicode started to adopt emojis in 2010, we have witnessed emojis' remarkable rise on the Web, with their adoption consistently increasing across multiple platforms (Rong et al., 2022; Lu et al., 2018; Kejriwal et al., 2021; Halverson et al., 2023). New emojis continue to be introduced in response to user requests. From 2018 to 2022, five new versions of emojis (Unicode 11.0 to 15.0) have been released, some of which, like the pleading face emoji (🥺) and the partying face emoji (🥳), have gained widespread adoption among Twitter users.

This standardized approach ensures that emojis maintain a stable form throughout their journey within social networks, enabling precise tracking of emoji adoption and diffusion. These attributes - precise definition, standardized implementation, stable form, and accurate release information — underscore the unique suitability of emojis for studying the evolution of language in online networks.

We take the initiative to study the diffusion of

emerging language units on social media through the unique perspective of Unicode-versioned emojis. We investigate the diffusion of newly created emojis introduced in Unicode versions 11.0 to 13.0, across the Twitter platform, particularly the usage frequency and semantic shift as the new emojis cascade through social media.

Because of the rich semantics, the Unicode definition (such as "pleading face") is far from enough to interpret the meaning of an emoji. To this end, we propose an interpretation framework that leverage language models (LMs) to identify words and existing emojis that share similar semantic contexts with the new emojis. Finally, we evaluate the practical implications of our framework by substituting new emojis with semantically similar ones in sentiment classification tasks, which demonstrates the effectiveness of our approach in helping NLP models interpret emerging language units for downstream tasks. We summarize our major contributions as follows:

- We explore the pattern of emoji diffusion initial adoption to widespread usage, with a focus on frequency and sentiment aspects.
- We introduce an interpretation framework to interpret the semantics of new emojis by exploring the words or old emojis with similar semantics.
- To validate the effectiveness of our interpretation framework, we replace emojis in texts with surrogates and improve the model performance in the sentiment classification task.

2 Related Work

Our work is based on two lines of existing work: the emoji understanding as well as its applications and the information cascade in the social media.

2.1 Emoji Understanding and Applications

The prevalence of emojis on many platforms, especially the social media platform, has gained increasing interest from researchers in various areas. They have studied the emoji functions in multiple aspects, such as conveying sentiments (Ai et al., 2017), highlighting topics (Lu et al., 2016), indicating identities (Ge, 2019), promoting communication (Zhou et al., 2023). In addition to Lu et al. (2018) which describes the emoji development process on the GitHub platform and Feng et al. (2020) which examines patterns of new emoji requests, the diffusion and interpretation of emojis in the recent version have not been fully studied.

Note that many researchers have utilized emojis to help with downstream tasks. Sentiment classification is the primary application to include emoji information (Chen et al., 2019; Felbo et al., 2017; Lou et al., 2020; Chen et al., 2018a; Eisner et al., 2016), and later research uses emoji usage to predict the developer dropout (Lu et al., 2022), the hashtag used (Zhou and Ai, 2022), or user gender (Chen et al., 2018b). However, these research works do not enable the generalizability of the framework on the newly appeared emojis. In this paper, we propose an emoji substitution method to improve the models' effectiveness on the new emojis without further parameter updates.

2.2 Innovation Diffusion on the Social Media

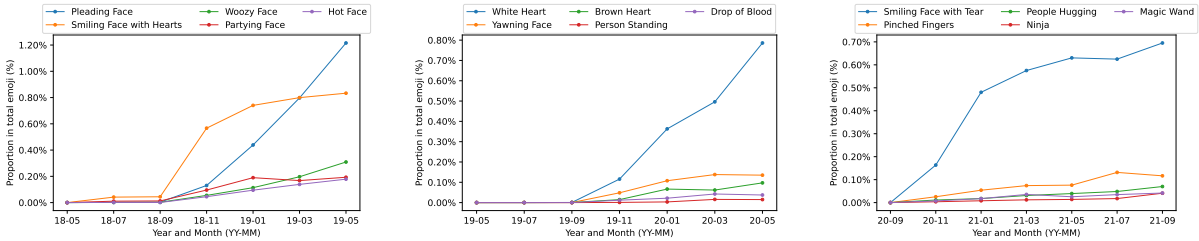
Previous work has noticed and explored the innovation sharing pattern on the social media, such as hashtags, meme, blog articles, and news diffusion (Ma et al., 2014; Johann and Bülow, 2019; Spitzberg, 2014; Kümpel et al., 2015; Yang et al., 2012; Ahmed et al., 2013; Bakshy et al., 2011; Cheng, 2017; Cunha et al., 2011) and proposed multiple models to understand and simulate the information cascade process (Zhou et al., 2021). In addition to content innovation, users create new language units and distribute them through social networks (Grieve et al., 2018; Kershaw, 2018) but few works systematically studied the reason, pattern, and interpretation of new language units.

In our work, we explore the diffusion of new emojis to reveal the diffusion mechanism of social media and frame the new emoji cascade as the innovation diffusion process. The previous study explores the factors influencing innovation popularity by adopting the theory of innovation diffusion (Rogers et al., 2014; Ma et al., 2014; Steffes and Burgee, 2009; Chu and Kim, 2011).

3 Understanding Emojis' Diffusion

We start exploring the diffusion of new emojis by analyzing the their usage frequency and semantic shift. We collected English Tweets from May 2018 to May 2022 using the Twitter API,[1] during which new emojis are released with Unicode 11.0 (February 2018), 12.0 (March 2019), 13.0 (January 2020). Note that it usually takes a few months for

[1] https://developer.twitter.com/en/docs/twitter-api



(a) Emoji 11.0 (b) Emoji 12.0 (c) Emoji 13.0

Figure 1: Frequency trends of the top 5 popular emojis in each emoji version over the two years following their first appearance. We show the frequency of emojis every two months, and the y-axis represents the proportion of each emoji in the total of emojis in that month.

platforms to support inputting and rendering new emojis. For example, the emojis in Emoji 11.0 first appear on Twitter around July 2018.

3.1 Emoji Popularity Increases

We visualize the frequency change of popular emojis of each version in Figure 1. Most of their popularity continuously increase over the two-year period following release. However, the adoption rates differs dramatically both between and within emojis. Even for the most popular emojis of the same version, their frequencies may vary by orders of magnitude within months, showing a power-law distribution similar to old emojis (Lu et al., 2016). The adoption speed is also uneven over time. For example, the frequency of 🥺 (pleading face) is still smaller than 🥳 (partying face) and 🥰 (smiling face with hearts) in January 2019. But by May 2019, the popularity of 🥺 has exceeded all emojis in Emoji 11.0. It suggests that some emojis are diffused to restricted user communities, while others emojis keep spreading and reach larger communities.

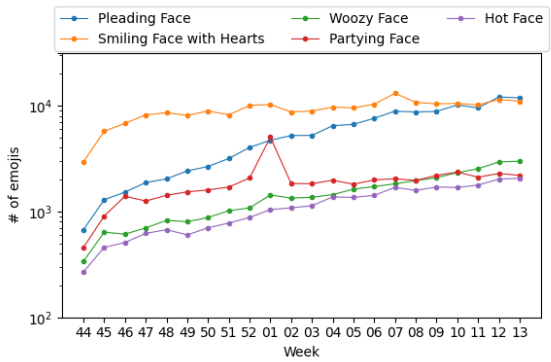


Figure 2: Frequency trends by week of the top 5 emojis in Emoji 11.0 from November 2018 to March 2019.

Seeing the continuous growth of new emojis' popularity after release, we then examine the change in emoji frequency in a more fine-grained time period. We visualize the count of the 5 most popular emojis of Emoji 11.0 at each week from November 2018 to March 2019 in Figure 2. We observe two significant bumps on the lines of 🥳 (partying face) and 🥰 (smiling face with hearts): a bump for 🥳 in Week 1 of 2019 and the a bump for 🥰 in Week 7 of 2019.

The bumps coincide bursting external events (New Year in Week 1 and Valentine's day in Week 7), which have been discussed in literature as possible triggers for information cascade (Zhou et al., 2021; Crane and Sornette, 2008). We hypothesize that external events also influence the adoption of new emojis. To verify, we examine whether the words associated with the new emoji in that period are related to external events. We collect the tweets with emojis in the first and seventh week of 2019 and calculate pointwise mutual information (PMI) between each emoji e and each word w. The PMI equation can be formulated as $\text{PMI}(e, w) = \log \frac{p(e,w)}{p(e)p(w)}$, where $p(w)$, $p(e)$ and $p(e, w)$ refer to the probability of a tweet containing the word w, the emoji e, and both of them, respectively. We present the top 10 associated emojis (based on PMI) for emoji 🥳 and 🥰 in different weeks in Table 1 and highlight the words related to external events, as recognized by the authors.

Emoji	Time period	Top 10 PMI words
🥳	W51, 2018	birthday, happy, hope, great, days, we, day, **year**, amazing, christmas
🥳	W01, 2019	happy, **new**, **year**, birthday, **2019**, **happynewyear**, may, everyone, **years**, **2018**
🥰	W51, 2018	amazing, you, wait, love, thank, thanks, see, beautiful, heart, so
🥰	W07, 2019	looking, **valentines**, tomorrow, **valentine**, ever, amazing, pretty, sweet, beautiful, you

Table 1: Highly-associated words (by PMI) of emoji 🥳 and 🥰 in different weeks. The semantics of the associated words coincide with the external events happened at that week.

From Table 1, we observe that for emoji 🥳 from the 51st week of 2018 to the 1st week of 2019, the words about the New Year event such as "*happynewyear*" and "*2019*" appear in the associated words, and for emoji 🥰, the associated words about Valentine's days show in the 7th week of 2019. The observation verifies our hypothesis and suggests that external events can influence the adoption of new emojis.

3.2 Influencing Factors of Emoji Diffusion

The popularity discrepancy of emojis in the same version raises the question of what influence the diffusion process. Since emojis can be considered a digital innovation, we apply Rogers' diffusion of innovation theory (Rogers et al., 2014), which describe the diffusion network and innovation itself as the influencing factors. We model each factor and describe their correlation with the new emojis' popularity.

Community size: Tracing the diffusion network is extremely challenging, because our tweets is collected through the Twitter's 1% sampling API, and the low access rate make it impossible to construct the diffusion network. However, previous work suggests that hashtags can indicate community identity in social networks (Yang et al., 2012; Zhou and Ai, 2022). We thus use the hashtags cooccuring with emojis as the proxy of the early adapter community of the new emojis, and more popular hashtags mean a larger size of communities.

The Spearman's correlation (Hauke and Kossowski, 2011) of hashtag popularity in the early period and emoji popularity in the late period is 0.580, 0.530, and 0.370 for emojis of Emoji 11.0, 12.0 and 13.0. It indicates that a larger early adopter community may promote the emoji diffusion. (Detailed analysis in Appendix A.1.)

Emoji semantics Emojis' semantics can be considered the innovation itself, so we hypothesize that emojis with more popular semantics are more adopted(Ai et al., 2017)). To proxy the popularity of emoji semantics and avoid the circular reasoning, we prompt GPT-4 (OpenAI, 2023) to generate words with semantics similar to given emojis and count these words' frequency in the early stage.

The Pearson correlation (Cohen et al., 2009) of word occurrence and emoji popularity is 0.413, 0.812 and 0.11 for emojis from Emoji 11.0, 12.0 and 13.0. The significant correlation for Emoji 11.0 and 12.0 supports our hypothesis. For Emoji 13.0, the correlation is weak, likely because users do not directly state the semantics inside the emojis. For example, for 🤌 (pinched fingers), GPT-4 shows us similar words as: *gesture, expressive, Italian, emphasis, and talkative*, which may not be presented in users' tweets. We present the detailed setup of the experiment in Appendix A.2.

3.3 Emoji Meaning Evolves during Diffusion

With the increasing popularity of the new emojis, one may wonder if their meanings are adapted during the diffusion process. However, understanding the emojis' meaning is hard, as will be discussed in Section 4.1. To get a first glimpse of the adaption, we focus on the context where emojis are used, and using the lens of sentiment, because one main functionality of using emojis is to convey sentiments (Ai et al., 2017).

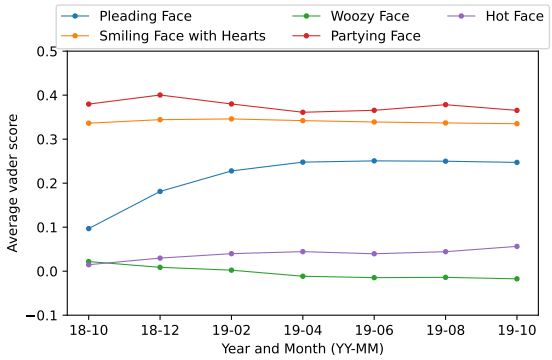


Figure 3: Average Vader scores of the tweets containing the top 5 popular emojis in Emoji 11.0 from October 2018 to October 2019.

To quantify the sentiment context of emojis, we calculate the *Vader* score for each tweet (Hutto and Gilbert, 2014), which outputs the sentiment score from -1 (negative) to 1 (positive). The higher the absolute value of the score, the more sentimental the tweet is. We average the Vader score of all tweets with an emoji in two months as the emoji's sentiment score in that time period and visualize its trend of the top 5 popular emojis from Emoji 11.0 every two months in Figure 3.

For most emojis, their sentiment scores remain constant, hinting that the user's understanding of most emojis is unchanged during the adaption. However, for the emoji 🥺 (pleading face), its sentiment score grew continuously in the first 8 months after its release, suggesting that users use 🥺 in increasingly positive sentiment context. We further visualize the Vader score distribution of tweets containing 🥺 in two time periods a year apart (October 2018 and 2019) in Figure 4.

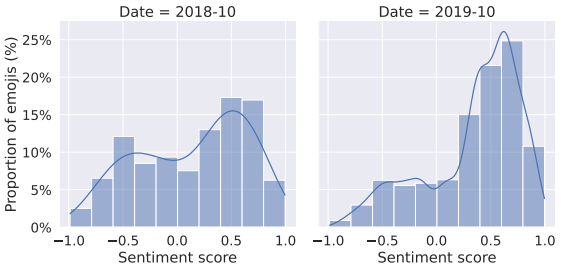


Figure 4: Distribution of Vader scores of tweets with emoji 🥺 in October 2018 and October 2019. During one year period, 🥺 concentrate more on positive tweets.

Initially (October 2018), many users use the emoji 🥺 in negative tweets; but during diffusion, more users tend to use it in positive tweets. We hypothesize that in the early adoption of 🥺, users used this emoji to convey sad feelings, but later, the emoji evolved to convey moved or other positive feelings. Indeed, we see that the associated words (measured by PMI) with the emoji 🥺 changes over time, which provide initial evidence that supports the hypothesis. In October 2018, *sad, sorry, idk* are among the sentimental words with highest PMI, while a year later, we see more positive words such as *prettiest, precious, cutest*. (The details are shown in Table 8 in Appendix A.3.) Through the lens of sentiment, we observe that the emojis' meaning may evolve during the diffusion process.

4 Interpret Emojis with Language Models

The analysis so far shows that the semantic of emojis not only affect their diffusion process, but also evolves during the diffusion process, both of which highlights the importance of an effective framework to interpret new emojis dynamically. Although ChatGPT is capable of showing us similar words to emojis (OpenAI, 2023; Zhou et al., 2024), it lack the understanding based on the application scenario, and cannot capture the semantic evolution during the diffusion. Moreover, the pre-training data of ChatGPT may not cover the recently-released emojis.

To address these challenges, we utilize the corpus containing new emojis and open-source language models (LMs) to investigate the application scenario and semantic meaning of emojis. Previous researchers have relied on the attention mechanism in LMs to explore the inner association of the fine-tuning dataset (Wang et al., 2021). In this section, we use the attention score method and the cross-dataset inference method to explore words and old emojis with semantics similar to new emojis.

4.1 Interpretation with High-attention Words

Attention scores are a well-studied interpretability method to identify important tokens for LMs to make the decision (Clark et al., 2019; Wang et al., 2021). To understand what words are specifically associated with newly created emojis, we design an emoji prediction task and extract high-attention words to reveal the emoji meaning.

4.1.1 Attention Calculation

Formally, we first construct an emoji classification dataset with the input space $x \in \mathcal{X}$ and the predefined emoji label y. To better distinguish the word association between new and old emojis, the tweet with label 0 means that the tweet contains the old emoji, and with label $1 \leq k \leq n$ means that the tweet contains the new emoji k. Denote f_e as the fine-tuned model in the emoji classification dataset (specifically the Roberta model for our experiments) (Liu et al., 2019). For each input tweet $x_i \in \mathcal{X}$ with tokens $\{t_i^1, \cdots, t_i^m\}$, where m is the token number, we adopt f_e in the input i and obtain the attention scores $\{a_i^1, \cdots, a_i^m\}$ and the emoji prediction $f_e(i)$. Since for Roberta model, the embeddings of the [CLS] token in the last layer are used to make the prediction, we compute the scores a_i^m as the average attention scores of the mth token to the [CLS] token across different heads. For the overall attention scores $\overline{a_{kt}}$ of the token t for the emoji k, we extract the sentences with prediction $f_e(x_i) = k$ and average the attention scores of the token t in these extracted sentences, which can be formulated as

$$\overline{a_{kt}} = \frac{\sum_{x_i \in \mathcal{X}} \mathbb{1}(f_e(x_i) = k) \cdot \sum_{j=1}^{m} (a_i^j \cdot \mathbb{1}(t_i^j = t))}{\sum_{x_i \in \mathcal{X}} \mathbb{1}(f_e(x_i) = k) \cdot \sum_{j=1}^{m} \mathbb{1}(t_i^j = t)}$$

where $\mathbb{1}$ is the indicator function. With the overall attention scores, for each emoji k, we can extract the keywords with the highest attention scores $\overline{a_{kt}}$ to understand the semantics of the new emojis.

4.1.2 Experiment Setup and Results

To verify the effectiveness of the attention method, we experiment on 6 emojis from Emoji 13.0: 🥲 (smiling face with tear), 🥷 (ninja), 🪄 (magic wand), 🤌 (pinched fingers), 🪙 (coin), 🫂 (people hugging). We extract the tweets with emojis from April 2022 to May 2022 and construct a balanced dataset with a total of 50,000 tweets and 7 labels (label 0 for old emojis and label 1 to 6 for new emojis). We fine-tune the Roberta

Emoji	Top 10 Attention Score Words	Top 3 Inference Score Emojis
🥲	same, its, so, me, **no**, **why**, this, **please**, **oh**, i	😔 (16.7), 👀 (8.5), 😩 (6.1)
🥷	days, account, **ninja**, both, **assassins**, who, **samurai**, coin, **gaming**, website	✨ (13.3), 👀 (13.2), 😊 (10.2)
🪄	**magic**, **wish**, **magician**, **wizard**, follow, hours, tweet, special, **light**, recent	✨ (22.6), 💥 (10.1), 🔥 (9.9)
🤌	this, puff, the, that, **kiss**, another, you, **art**, **perfect**, **please**	😍 (10.2), 🔥 (7.8), 💯 (6.1)
🪙	start, **billion**, **coins**, **token**, **coin**, **money**, proof, gob, hours, follow	💥 (25.1), 🎉 (16.4), 🔥 (10.5)
🫂	**thanks**, **hugs**, my, **hug**, **good**, you, **hugging**, **dear**, **friend**, **happy**	🥺 (11.2), 💙 (8.9), ✨ (8.3)

Table 2: Words with high attention scores and old emojis with high inference scores for new emojis from Emoji 13.0. High attention scores suggest words with similar semantics as emojis and indicate the application setting of emojis. Emojis with high inference scores represent the old emojis with similar meaning to the new emojis.

model pretrained on tweets (Barbieri et al., 2020) (`twitter-roberta-base` [2]) on the emoji prediction dataset with the split 8:1:1 and obtain the test accuracy 66.98%. We present the top 10 words with highest attention scores in the second column of Table 2 and the words recognized by the authors with similar semantics are highlighted in bold.

Table 2 demonstrates that attention scores are effective in identifying words semantically similar to specific emojis. These words not only mirror the primary meaning of the emojis but also reveal their application scenarios and extended interpretations. For entity-related emojis such as 🥷, 🪄, and 🪙, their high-attention words extend beyond their direct symbolism. For instance, the word "gaming" associated with 🥷 highlights its frequent use in the context of action video games. Similarly, the term "token" linked with 🪙 (coin) suggests its application in representing Bitcoin tokens, indicating a broader usage beyond its conventional meaning.

Regarding sentiment-related emojis, high-attention score words encapsulate the sentiments embedded in the emojis. For instance, the word "perfect" associated with 🥲 implies positive sentiment, while "why" and "no" linked to 🤌 suggests negative sentiments. The qualitative results in Table 2 demonstrate the effectiveness of using the attention mechanism to probe the application setting and the extensive meaning of the new emojis.

[2] `https://huggingface.co/cardiffnlp/twitter-roberta-base`

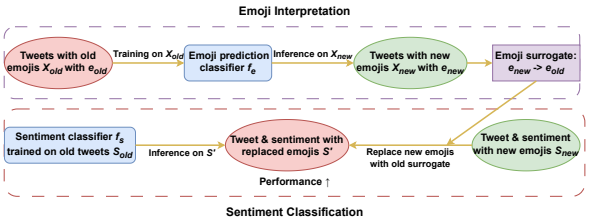


Figure 5: The upper half shows the framework of using old emojis to interpret new emojis and the lower half presents the pipeline of replacing new emojis in the sentiment classification dataset to old similar emojis to enhance the prediction results.

4.2 Interpret Emojis with Old Emojis

The words with high-attention scores can not fully capture the sentiments inherent in emojis. For example, the degree of negative sentiment conveyed by 🥲, or the specific sentiment associated with 🥷 and 🪙, remains ambiguous. Besides exploring words with similar semantics, we employ the rich and complex sentiments encoded in old emojis to interpret the newly created emojis.

We utilize LMs to conduct cross-version inference to explore old emojis with semantics similar to a new targeted emoji. If the semantics and syntactic features of the text containing two emojis are similar, the semantics of two emojis are also similar. Our method is to first fine-tune the LMs on an emoji classification dataset with old emojis as the labels, and we use the fine-tuned LMs to do the inference on tweets containing new emojis. If LMs predict tweets with a new emoji to contain another old emoji, it means that two emoji share a similar text context distribution, indicting similar semantics. The pipeline of the interpretation framework is shown in the upper half of Figure 5.

4.2.1 Cross-Version Analysis

We construct an emoji classification dataset with the input tweets $\mathcal{X}_{old}$ and the pre-defined emoji labels from old emojis $\{e^0_{old}, e^1_{old}, \cdots, e^n_{old}\}$. Pre-trained LM f_e is fine-tuned in tweet collection $\mathcal{X}_{old}$. We construct another tweet dataset $\mathcal{X}_{new}$, where each tweet contains the new emojis $\{e^0_{new}, e^1_{new}, \cdots, e^m_{new}\}$ and ask the fine-tuned LM f to do the inference on each tweet $x_{new} \in \mathcal{X}_{new}$. The prediction on the tweet x_{new} containing the emoji y_{new} is: $f_e(x_{new}) = \text{argmax}_{e_{old}}\, p(e_{old}|x_{new})$.

The semantic similarity between an old emoji e_{old} and a new emoji e_{new} can then be quantified as the proportion of predictions equal to e_{old}:

$$score(e_{old}, e_{new}) = \frac{\sum_{x_{new}} \mathbb{1}(y_{new}=e_{new}, f_e(x_{new})=e_{old})}{\sum_{x_{new}} \mathbb{1}(y_{new}=e_{new})}. \quad (1)$$

We sort the old emojis for each new emoji by the calculated inference $score(\cdot)$ values and for each new emoji, extract three old emojis with the highest inference scores, named as emoji surrogates.

4.2.2 Experiment Setup and Results

We first construct $\mathcal{X}_{old}$ with tweets from April 2020 to May 2020, when Emoji 13.0 does not appear on Twitter. We extract 10,000 tweets for every emoji in top 32 old emojis from April to May and another 10,000 tweets without any emoji. We form a training dataset $\mathcal{X}_{old}$ with 330,000 tweets and 33 labels (32 emojis + no emoji). The pre-trained Roberta model f_e is fine-tuned in $\mathcal{X}_{old}$ and we use f_e to infer on the test dataset constructed with tweets from 2022 containing the emojis from Emoji 13.0 in Table 2. We calculate the inference score between the old and new emojis as Equation 1 and present the top 3 similar ones for each new emoji, as well as the inference score in the third column in Table 2.

Table 2 reveals that preexisting emojis effectively reflect the semantic content, particularly the sentiment dimension, of new emojis. For sentiment-related emojis such as 🥲, older emojis such as 😔 (pensive face), 🥺 (pleading face) and 😩 (weary face) encapsulate the blend of sadness and begging inherent in the new emoji. Furthermore, in the case of entity-related emojis, while it may be challenging to pinpoint analogous emojis that precisely capture an entity's characteristics, it is feasible to discern the underlying sentiments these emojis convey. For example, the 🥷 (ninja) emoji, in relation to existing emojis such as ✨ (sparkles), 👀 (eyes), and 😊 (smiling face with smiling eyes), reveals the positive sentiment in 🥷 in usage.

In the next section, we show how cross-version analysis serves downstream tasks. We find that simply substituting new emojis with old emoji surrogates can significantly increase the accuracy in the sentiment classification task of fine-tuned LMs.

4.3 Replace Emojis for Sentiment Prediction

Sentiment classification is a well-studied task and is widely used in deployed systems. Emojis, which encode rich sentiments, have been shown to be important for enhancing the performance in the sentiment classification (Chen et al., 2019). The current state-of-the-art (SoTA) method for the sentiment classification on tweets is to fine-tune the pre-trained LMs on a training dataset (Barbieri et al., 2020). However, the newly-created emojis not in both fine-tuning and pre-training data raise a challenge for this method, which may prevent the model from making precise predictions.

Instead of fine-tuning models on the data with new emojis again, requiring a high computational cost, we propose a method to directly substitute the new emoji with old emoji surrogates without parameter updates. The sentiments encoded in emoji surrogates could compensate for the loss of semantics in the unknown emojis. We present the emoji substitution process in the lower half of Figure 5.

4.3.1 Methodology

Suppose that we have a sentiment classification dataset $\mathcal{S}_{new}$ that contains new emojis and each input has the sentiment label y. Given a sentiment classifier f_s fine-tuned in $\mathcal{S}_{old}$, where the new emojis are not in $\mathcal{S}_{old}$, for each tweet $s_i = [t_i^1, t_i^2, \cdots, e_{new}, \cdots, t_i^m] \in \mathcal{S}_{new}$, we replace the new emoji e_{new} in tweet s_i with the old emoji surrogates (top 3 old emojis sorted by the inference scores) and obtain a dataset $\mathcal{S}'$ with $s_i^{'} = [t_i^1, t_i^2, \cdots, e_{old}^1, e_{old}^2, e_{old}^3, \cdots, t_i^m]$. Then we feed the tweet $s_i^{'}$ to the classifier f_s to get the sentiment prediction, which is $f_s(s_i) = \text{argmax}_y\, p(y|s_i^{'})$.

4.3.2 Experiment Setup

Existing sentiment classification datasets with emojis did not contain the newly created emojis. Therefore, we collect tweets from 2020 and 2022, respectively, as the dataset with old and new emojis. Due to the superiority of LLMs over human annotators (Gilardi et al., 2023), we utilize GPT3.5 to annotate the label for tweets (Ouyang et al., 2022). Given a tweet input s, and $\mathcal{Y} = \{positive, neutral, negative\}$, ChatGPT is asked to choose a label. We repeat the annotation process twice with temperature 0.7 and keep the examples with labels agreed by two LLM annotators.

We first randomly collect tweets from April 2020 to May 2020 and then use ChatGPT to label the sentiment of each tweet as $\mathcal{S}_{old}$. We first split $\mathcal{S}_{old}$ into training, validation, and test dataset by 8:1:1 with 35,388 tweets. We fine-tune the Roberta classifier (`twitter-roberta-base`) on $\mathcal{S}_{old}$ as f_s, pre-trained on tweets before August 2019 (Barbieri et al., 2020). The fine-tuned model f_s can achieve an accuracy of 67.52% in test data from 2020. We repeat the collection and labeling process in tweets from April 2022 to May 2022 to form $\mathcal{S}_{new}$.

Emoji 13.0 (sentimental)					
emoji	surrogates	# test	ori acc	replaced acc	word acc
🙂	😔👀😩	23,080	63.43	**63.90**	59.87
🤨	🤔👀😔	1,908	48.69	63.87	**67.26**
🫂	🥺💙✨	6,580	80.32	**90.90**	85.88
👌	😍🔥💯	6,972	79.75	**82.80**	67.76
Emoji 13.0 (entity)					
🗼	👍🔥💥	614	22.05	18.10	**23.43**
🥷	✨👀😊	6,649	55.03	**56.86**	54.06
🎱	💥🎉🔥	3,209	52.32	**61.92**	59.58
🫐	💥👀✨	922	65.72	**67.25**	68.66

Table 3: Results of emoji replacement method on the sentiment prediction task for emojis from Emoji 13.0. The fine-tuned models are pre-trained on data without new emojis. *ori acc*, *replaced acc*, and *word acc* represent the accuracy on original test data, replacing new emojis to emoji surrogates, and replacing new emojis with emoji names, respectively.

Accuracy of ChatGPT in labeling sentiment can be verified using existing data. We randomly select 1,000 tweets from SemEval-2015 Task 10 (Rosenthal et al., 2015) and the accuracy of ChatGPT on sentiment prediction is 77.29%. The average agreement between human annotators and the gold standard annotations in the test dataset is 75.70%. Therefore, ChatGPT annotations are reliable due to the comparable agreement to the human annotators.

We randomly select 8 emojis from Emoji 13.0, and 4 of them are sentiment-related, and the others are entity-related. We collect the tweets from $\mathcal{S}_{new}$ with the selected emojis. We prepare a baseline method to substitute the new emoji with the emoji name (words) to represent the emoji semantics. We present the accuracy on the original tweets, on the emoji replacement tweets, and on the word replacement tweets for each specific emoji in Table 3.

4.3.3 Results

From Table 3, when replacing the new emoji with old emojis with similar semantics, we can observe a notable performance increase for the model f_s, with a relative improvement of 4.94%, 6.92% on emojis from Emoji 13.0 (sentimental) and Emoji 13.0 (entity), respectively. The relative improvement on the entity-related emojis is more significant, which indicates that the explore old emoji surrogates precisely suggest the sentiment embedded in the entity-related emojis, leading to better results in the sentiment prediction. Compared to replacing the new emoji with the emoji name, our emoji replacement method can outperform in most cases, and the possible reason could be attributed to words from the emoji name, which can not fully characterize the complex sentiments for emojis.

We repeat the experiments on the Roberta model with new emojis in the pre-trained corpus (`twitter-roberta-base-2022-154m`, pretrained on tweets before December 2022) (Loureiro et al., 2023), and present the results in Table 9 in Appendix. Compared with models pre-trained on tweets without new emojis, the average improvement of replacing new emojis becomes smaller or even negative, which follows our expectation that the fine-tuned model f_s has learned the semantics of new emojis in the pre-training data and the substitution of emojis causes loss of information.

5 Implications

Our research reveals the patterns of emoji diffusion and proposes a framework to understand the semantics of new emojis. For model developers, our emoji replacement framework can provide an effective way to increase the generalization of the fine-tuned LLMs on texts containing new emojis without further parameter updates, which can also be extended to other NLP tasks. For future emoji researchers, our work provides a pipeline to explore the pattern of new emoji diffusion and proposes a framework to utilize open source LMs to interpret the emoji semantics and usage scenarios.

Moreover, our work provides an insight of the diffusion patterns on a new language unit in social networks, which can inspire the future researchers to generalize our diffusion exploration pipeline and interpretation framework on other new tokens, such as memes and viral hashtags. Along with our work, we expect that the exploration study on the language unit diffusion can reveal the evolution mechanism of the content in digital communities

6 Conclusion

In this work, we explore the pattern of emoji diffusion by studying recently created emojis. We show that the external event can influence the short-period frequency and that some emojis may have an evolution in the sentiments when adopted by more users. Then, we propose a framework to use the words or existing emojis to interpret the new emojis. Finally, we show the effectiveness of our framework by replacing new emojis with similar old emojis to improve the sentiment prediction task. The promising results indicate that our framework for interpreting emojis can be used to increase the generalizability of trained models.

7 Limitations

There are three limitations in our work. First, both our exploration of emoji diffusion and the following semantic analysis regard a type of emoji as a whole at a macro level, but it is not clear how an emoji diffuses by the interaction between users due to the lack of complete network data for Twitter. Considering the current API policy of the Twitter platform, we leave that part to future work by exploring other online communication datasets, such as Slack.

Second, when analyzing the characteristics of new emojis, in this paper, we only focus on the semantic features of emojis, also the focus of the previous work, but we find that the visual features could also be an important feature for emoji diffusion and may also influence the emoji semantics. For example, after one year of diffusion, the number of tweets with 🤍 (white heart) is 10 times higher than the tweets with 🤎 (brown heart). These visual features play an important role in exploring emoji usage and downstream emoji applications. For future work, it could be interesting to use these visual features, combining the text features to predict the popularity of emojis.

Third, in addition to sentiment classification, the downstream tasks with emojis also include hashtag prediction, gender prediction, and GitHub user dropout prediction (Chen et al., 2018b; Zhou et al., 2023; Lu et al., 2023). Future work could extend our interpretation framework to improve the fine-tuned models on these well-defined tasks without updating the model parameters.

References

Mohamed Ahmed, Stella Spagna, Felipe Huici, and Saverio Niccolini. 2013. A peek into the future: Predicting the evolution of popularity in user generated content. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 607–616.

Wei Ai, Xuan Lu, Xuanzhe Liu, Ning Wang, Gang Huang, and Qiaozhu Mei. 2017. Untangling emoji popularity through semantic embeddings. In *ICWSM 2017*.

Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. 2011. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74.

Francesco Barbieri, Jose Camacho-Collados, Leonardo Neves, and Luis Espinosa-Anke. 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. *arXiv preprint arXiv:2010.12421*.

Yuxiao Chen, Jianbo Yuan, Quanzeng You, and Jiebo Luo. 2018a. Twitter sentiment analysis via bi-sense emoji embedding and attention-based lstm. *ACM-MM 2018*.

Zhenpeng Chen, Xuan Lu, Wei Ai, Huoran Li, Qiaozhu Mei, and Xuanzhe Liu. 2018b. Through a gender lens. *WWW 2018*.

Zhenpeng Chen, Sheng Shen, Ziniu Hu, Xuan Lu, Qiaozhu Mei, and Xuanzhe Liu. 2019. Emoji-Powered representation learning for Cross-Lingual sentiment classification. In *WWW 2019*.

Hsiu-Hua Cheng. 2017. The antecedents of creative article diffusion on blogs: Integrating innovation diffusion theory and social network theory. *Online Information Review*, 41(1):70–84.

Shu-Chuan Chu and Yoojung Kim. 2011. Determinants of consumer engagement in electronic word-of-mouth (ewom) in social networking sites. *International journal of Advertising*, 30(1):47–75.

Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D Manning. 2019. What does bert look at? an analysis of bert's attention. *arXiv preprint arXiv:1906.04341*.

Israel Cohen, Yiteng Huang, Jingdong Chen, Jacob Benesty, Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. *Noise reduction in speech processing*, pages 1–4.

Riley Crane and Didier Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653.

Evandro Cunha, Gabriel Magno, Giovanni Comarela, Virgilio Almeida, Marcos André Gonçalves, and Fabricio Benevenuto. 2011. Analyzing the dynamic evolution of hashtags on twitter: a language-based approach. In *Proceedings of the workshop on language in social media (LSM 2011)*, pages 58–65.

Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel. 2016. emoji2vec: Learning emoji representations from their description. *arXiv preprint*.

Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *EMNLP*.

Yunhe Feng, Zheng Lu, Wenjun Zhou, Zhibo Wang, and Qing Cao. 2020. New emoji requests from twitter users: when, where, why, and what we can do about them. *ACM Transactions on Social Computing*, 3(2):1–25.

Jing Ge. 2019. Emoji sequence use in enacting personal identity. In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19, page 426–438, New York, NY, USA. Association for Computing Machinery.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. Chatgpt outperforms crowd-workers for text-annotation tasks. *arXiv preprint arXiv:2303.15056*.

Jack Grieve, Andrea Nini, and Diansheng Guo. 2018. Mapping lexical innovation on american social media. *Journal of English Linguistics*, 46(4):293–319.

Colin ME Halverson, Claire E Donnelly, Michael Weiner, and Joy L Lee. 2023. Content analysis of emoji and emoticon use in clinical texting systems. *JAMA Network Open*, 6(6):e2318140–e2318140.

Jan Hauke and Tomasz Kossowski. 2011. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, 30(2):87–93.

Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *ICWSM 2014*.

Michael Johann and Lars Bülow. 2019. One does not simply create a meme: Conditions for the diffusion of internet memes. *International Journal of Communication*, 13:23.

Mayank Kejriwal, Qile Wang, Hongyu Li, and Lu Wang. 2021. An empirical study of emoji usage on twitter in linguistic and national contexts. *Online Social Networks and Media*, 24:100149.

Daniel James Kershaw. 2018. *Language change and evolution in online social networks*. Lancaster University (United Kingdom).

Anna Sophie Kümpel, Veronika Karnowski, and Till Keyling. 2015. News sharing in social media: A review of current research on news sharing users, content, and networks. *Social media+ society*, 1(2):2056305115610141.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint*.

Yinxia Lou, Yue Zhang, Fei Li, Tao Qian, and Donghong Ji. 2020. Emoji-based sentiment analysis using attention networks. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*

Daniel Loureiro, Kiamehr Rezaee, Talayeh Riahi, Francesco Barbieri, Leonardo Neves, Luis Espinosa Anke, and Jose Camacho-Collados. 2023. Tweet insights: A visualization platform to extract temporal insights from twitter. *arXiv preprint arXiv:2308.02142*.

Xuan Lu, Wei Ai, Zhenpeng Chen, Yanbin Cao, and Qiaozhu Mei. 2022. Emojis predict dropouts of remote workers: An empirical study of emoji usage on github. *PloS one*.

Xuan Lu, Wei Ai, Xuanzhe Liu, Qian Li, Ning Wang, Gang Huang, and Qiaozhu Mei. 2016. Learning from the ubiquitous language: An empirical analysis of emoji usage of smartphone users. In *UbiComp 2016*.

Xuan Lu, Wei Ai, Yixin Wang, and Qiaozhu Mei. 2023. Team resilience under shock: An empirical analysis of github repositories during early covid-19 pandemic. In *ICWSM 2023*.

Xuan Lu, Yanbin Cao, Zhenpeng Chen, and Xuanzhe Liu. 2018. A first look at emoji usage on github: An empirical study. *arXiv preprint*.

Long Ma, Chei Sian Lee, and Dion Hoe-Lian Goh. 2014. Understanding news sharing in social media: An explanation from the diffusion of innovations theory. *Online information review*, 38(5):598–615.

OpenAI. 2023. Gpt-4 technical report.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.

Everett M Rogers, Arvind Singhal, and Margaret M Quinlan. 2014. Diffusion of innovations. In *An integrated approach to communication theory and research*, pages 432–448. Routledge.

Shiyue Rong, Weisheng Wang, Umme Ayda Mannan, Eduardo Santana de Almeida, Shurui Zhou, and Iftekhar Ahmed. 2022. An empirical study of emoji use in software development communication. *Information and Software Technology*.

Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 451–463.

Brian H Spitzberg. 2014. Toward a model of meme diffusion (m3d). *Communication Theory*, 24(3):311–339.

Erin M Steffes and Lawrence E Burgee. 2009. Social ties and online word of mouth. *Internet research*, 19(1):42–59.

Tianlu Wang, Rohit Sridhar, Diyi Yang, and Xuezhi Wang. 2021. Identifying and mitigating spurious correlations for improving robustness in nlp models. *arXiv preprint arXiv:2110.07736*.

Lei Yang, Tao Sun, Ming Zhang, and Qiaozhu Mei. 2012. We know what @you #tag: Does the dual role affect hashtag adoption? In *WWW 2012*.

Jerrold H Zar. 2005. Spearman rank correlation. *Encyclopedia of biostatistics*, 7.

Fan Zhou, Xovee Xu, Goce Trajcevski, and Kunpeng Zhang. 2021. A survey of information cascade analysis: Models, predictions, and recent advances. *ACM Computing Surveys (CSUR)*, 54(2):1–36.

Yuhang Zhou and Wei Ai. 2022. #emoji: A study on the association between emojis and hashtags on twitter. In *ICWSM 2022*.

Yuhang Zhou, Xuan Lu, Ge Gao, Qiaozhu Mei, and Wei Ai. 2023. Emoji promotes developer participation and issue resolution on github. *arXiv preprint arXiv:2308.16360*.

Yuhang Zhou, Paiheng Xu, Xiyao Wang, Xuan Lu, Ge Gao, and Wei Ai. 2024. Emojis decoded: Leveraging chatgpt for enhanced understanding in social media communications. *arXiv preprint arXiv:2402.01681*.

Appendix

A Details of Influencing Factor Analysis

A.1 Influence of Community Size on Emoji Popularity

We first identify the early stage of Emoji 11.0, Emoji 12.0, and Emoji 13.0 as July 2018, September 2019, and November 2020, two months after the first appearance, and set the late stage as the month one year later than the early stage. We count the number of co-occurred hashtags of each emoji in the early stage and the number of emojis in the late stage. We present the top 10 mostly used emojis in the late stage from Emoji 13.0 as well as their top 3 popular co-occurred hashtags in Table 4 and the count of emojis and hashtags in the early and late stages of emoji diffusion. For the co-occurred hashtags for emojis in Emoji 11.0 and 12.0, we show them in Table 5.

From Table 4, we can observe that there is a pattern in which, for popular emojis in the future, the co-occurred hashtags are also popular in the early adopted stage. Since hashtag number and emoji number are in difference scales in our dataset, we measure the Spearman rank correlation coefficient between hashtag number in the early stage and emoji number in the late stage for the top 10 new emojis from Emoji 11.0, 12.0, and 13.0, respectively (Zar, 2005).

The Spearman's correlation of co-occurred hashtag number and emoji number is 0.580, 0.530, and 0.370 ($p - value$ = 0.08, 0.11, 0.29) for the top 10 emojis of Emoji 11.0, 12.0 and 13.0, respectively. It indicates that emojis cooccurring with the popular hashtags tend to be used more in the future, and larger communities promote the emoji diffusion. In addition to the characteristics of early adopters and the diffusion network, we are also interested in the influence of the inner characteristics of emojis on emoji diffusion.

A.2 Influence of Emoji Semantics on Emoji Popularity

We expect that extracted words with similar semantics as the emoji should be common on the Twitter platform, so our prompt for querying GPT-4 is composed as follows: *Show me five common single words on Twitter with similar semantics to this emoji: {emoji}*, where *{emoji}* is the new emoji from Emoji 11.0, 12.0, 13.0, such as 🥰 (smiling face with hearts). We count the average word number in the early stage of new emojis (the same month as in Section A.1) and the emoji number in the late stage (one year later after the early stage). Since we find that the emoji and word count are on the same scale, we calculate the Spearman's rank correlation coefficient and the Pearson correlation coefficient for the logarithm of these two numbers as the measurement of association (Zar, 2005; Cohen et al., 2009). We present the word and emoji number of the top 10 popular emojis from Emoji 13.0 in Table 6 and for similar words for emojis in Emoji 11.0 and 12.0, we show them in Table 7.

Emoji	Co-occurred hashtag	# hashtag (early)	# emoji (late)
🥲	PLTPINKMONDAY, PLTPINKSUNDAY, PLTCyberMonday	79982.2	67,103
🫵	ATINYDAY, HeartbreakWeather, NiallHoran	18806.0	11,583
🫂	EveryVoteCounts, StayHome, StaySafe	7172.9	7,829
🪙	Bitcoin, gold, DeFi	10267.2	5,900
🪄	MissguidedCyberTreat, PLTPINKMONDAY, ibelieveinfairys	11501.1	5,355
🥷	syurabaikhati, DerApotheker, wolfpac	1128.8	4,510
🪴	GreenIndiaChallenge, SupportSmallStreamers, RRRMovie	46272.1	3,014
😉	Election2020, PLTPINKMONDAY, TREASURE	83920.8	2,497
🫓	HappyThanksgivingEve, Design, Emoji	866.6	2,113
🪧	MerzmenschPresents, LatentVoices, JukeBox	3343.3	1,666

Table 4: Co-occurred hashtags for top 10 popular emojis in Emoji 13.0 in the late stage. # hashtag (early) represents the hashtag number in the early stage of emoji diffusion, two months after the first appearance of emojis. # emoji (late) shows the emoji number in the late stage, one year after the early stage.

A.3 Supplementary Results and Prompt Details

Emoji	Co-occurred hashtag	# hashtag (early)	# emoji (late)
🥺	SaveShadowhunters, EXO, BTSARMY,	37122.3	101,667
🥰	TeenChoice, SouhilaBenLachhab, Cover	70924.1	56,449
😔	Prediction, Democrats, Obamacare	5121.8	29,549
😡	Heatwave, SummerSkinSafety, WorldEmojiDay	693.8	14,650
😍	Ethereum, cryptocurrency, eth	22075.3	13,070
😷	TRuMP, WorldCup2018, UFC231	11416.0	2,137
🙇	Directive, ItsACelebration, TeamNGH	76.1	616
🥊	MAGA, CRO, Boxing	8831.0	539
🧁	Cakes, baking, weekendreads	1979.1	435
💅	TMay, Chequers, UK	1736.25	339

(a) Co-occurred hashtags for emojis in Emoji 11.0

Emoji	Co-occurred hashtag	# hashtag (early)	# emoji (late)
🤍	J9, 3YearsWithCBX, StreamLYTLM	448.50	85,981
🕴	Tawan_V, TeamGalaxy, withGalaxy	144.20	13,473
😢	13ReasonsWhy3, 13ReasonsWhy, DeleteFacebook	291.60	13,127
🤎	thestripperoficial, camren, natiese	2.0	9,321
🤦	StarTrekDiscovery, Friends, RossGeller	390.0	4,433
🩸	PeriodEmoji, PeriodStigma, PeriodPoverty	12.0	3,648
🍆	onlyfans, camgirl, cammodel	793.0	2,729
📏	AccessATE, InstructionalDesign, CADET	9.70	2,603
🟣	wolvsden	7.0	2,393
🥾	-	0	1,936

(b) Co-occurred hashtags for emojis in Emoji 12.0.

Table 5: Co-occurred hashtags for top 10 popular emojis in Emoji 11.0 and 12.0. # hashtag (early) represents the hashtag number in the early stage of emoji diffusion, two months after the first appearance of emojis. # emoji (late) shows the emoji number in the late stage, one year after the early stage.

Emoji	Words with similar semantics	# word (early)	# emoji (late)
😭	bittersweet, emotional, touched, relieved, grateful	1401.8	67,103
🤌	gesture, expressive, Italian, emphasis, talkative	190.8	11,583
🫂	hug, comfort, support, embrace, togetherness	3714.6	7,829
🪙	money, currency, cash, change, gold	5999.8	5,900
🪄	magic, enchantment, spell, wizardry, mystical	609.6	5,355
🥷	stealthy, mysterious, skilled, warrior, covert	159.4	4,510
🪴	green, leafy, indoor, botanical, decorative	726.6	3,014
🥸	incognito, hidden, undercover, sneaky, disguised	403.8	2,497
🫓	bread, flat, pita, naan, food	674.0	2,113
🪧	sign, protest, message, board, banner	1311.0	1,666

Table 6: Words with similar semantics (from ChatGPT) for top 10 popular emojis in the late stage in Emoji 13.0. # word (early) represents the word number in the early stage of emoji diffusion, two months after the first appearance of emojis. # emoji (late) shows the emoji number in the late stage, one year after the early stage.

Emoji	Words with similar semantics	# word (early)	# emoji (late)
🥺	please, sorry, help, sad, desperate	12973.8	101,667
🥰	love, happy, adorable, blissful, sweet	33005.4	56,449
😵‍💫	dizzy, confused, woozy, drunk, lightheaded	1242.0	29,549
🥵	hot, sweaty, exhausted, overwhelmed, burning	2578.2	14,650
🥳	celebrating, party, woohoo, ecstatic, jubilant	1470.0	13,070
🥶	cold, freezing, shivering, frosty, icy	591.6	2,137
🧸	cute, soft, cuddly, plush, adorable	4112.6	616
🧨	loud, explosive, bang, pop, fireworks	1273.2	539
🧁	sweet, delicious, cute, frosting, treat	5303.2	435
🦶	step, walk, run, sole, toe	2263.0	339

(a) Words with similar semantics for emojis in Emoji 11.0.

Emoji	Words with similar semantics	# word (early)	# emoji (late)
🤍	pure, love, clear, sincere, peace	21063.2	85,981
🧍	stand, upright, wait, solo, idle	4793.4	13,473
🥱	tired, sleepy, bored, exhausted, drowsy	1676.2	13,127
🤎	warmth, earthy, comfort, stable, rich	496.6	9,321
🤏	small, little, slight, precise, minimal	3843.8	4,433
🩸	blood, bleed, drip, red, donate	1376.2	3,648
🪐	saturn, space, cosmic, orbit, celestial	398.6	2,729
🦯	blind, aid, navigate, mobility, independence	266.2	2,603
🟣	purple, geometric, round, violet, circle	842.6	2,393
🦦	otter, playful, aquatic, furry, adorable	487.4	1,936

(b) Words with similar semantics for emojis in Emoji 12.0.

Table 7: Words with similar semantics (from ChatGPT) for top 10 popular emojis in Emoji 11.0 and 12.0. # word (early) represents the word number in the early stage of emoji diffusion, two months after the first appearance of emojis. # emoji (late) shows the emoji number in the late stage, one year after the early stage.

Time Period	Top 10 PMI Sentimental Words	Score Avg.
2018-10	cry sad please miss hate sorry idk wish bad stop	-0.198
2019-10	sobbing protect cry prettiest pls precious hug cutest sad heart	0.221

Table 8: Top 10 associated words with positive or negative sentiments of emoji 🥺 in October 2018 and October 2019. Red and blue highlight positive and negative words, respectively. The darker the background color, the more sentimental the words.

Emoji 13.x (sentimental)				
emoji	surrogates	# test	ori acc	replaced acc
🥲	😔🥺😩	23,080	**66.57**	60.58
🤠	🤔👀😔	1,908	57.08	**66.18**
🫂	🥺💙✨	6,580	86.99	**91.57**
🫵	😍🔥💯	6,972	82.77	**83.79**
🫨	😳✨😈	6,206	**56.86**	55.15
❤️‍🩹	✨🥺😔	6,160	84.10	**88.44**
😩	😈💦😟	6,390	**51.75**	40.94
🫠	✨😈👀	1,808	48.12	**56.19**
Emoji 13.x (entity)				
🗼	👍🔥💥	614	**22.39**	17.57
🌋	✨👀😊	6,649	58.02	**59.82**
🪨	💥🎉🔥	3,209	63.78	**66.84**
🫧	💥👀✨	922	**69.95**	69.31

Table 9: Results of emoji replacement method on the sentiment prediction task for selected emojis from Emoji 13.0 and 13.1 on models pre-trained on data before December 2022, including new emojis in the pre-training data.