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

Yuhang Zhou
University of Maryland
College Park, USA
tonyzhou@umd.edu

Xuan Lu
University of Arizona
Tucson, USA
luxuan@arizona.edu

Wei Ai
University of Maryland
College Park, USA
aiwei@umd.edu

Abstract

The frequent introduction of new emojis in each Unicode release creates a dynamic shift in social media content, providing a unique opportunity to explore the evolution of digital language. Analyzing a large dataset of sampled English tweets, we examine how newly released emojis gain popularity and evolve in meaning. We find that community size of early adopters and emoji semantics are positively correlated with 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 downstream text 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 study-

Proceedings of the 7th Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), pages 293–306
July 3, 2026 ©2026 Association for Computational Linguistics

ing the evolution of language in social 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 and irony detection 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 and irony detection task.

2 Related Work

Our work builds upon three research areas: emoji understanding and applications, innovation diffusion on social media, and temporal effects in machine learning models.

Emoji Understanding and Applications. The functions and applications of emojis are widely studied, including their role in conveying sentiment (Ai et al., 2017), irony (Hu et al., 2017), topics (Lu et al., 2016), and identity (Ge, 2019). Emojis have been incorporated into various downstream NLP tasks, notably sentiment analysis (Chen et al., 2019; Felbo et al., 2017; Eisner et al., 2016) and irony detection (Hayati et al., 2019). While the emoji development process (Lu et al., 2018) and requests (Feng et al., 2020) have been examined, the diffusion dynamics and evolving interpretations of *recently introduced* emojis remain largely unexplored. Furthermore, existing methods often fail to generalize effectively to these new, unseen emojis.

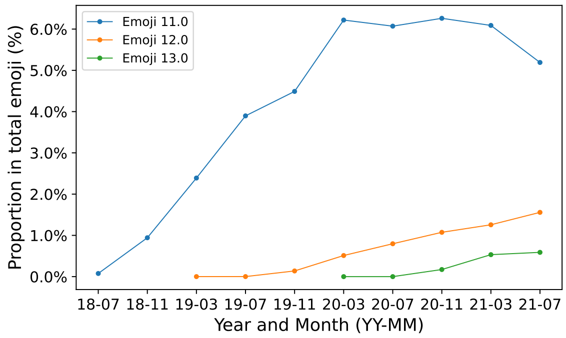


Figure 1: Frequency trends per emoji version over three years post-release. Y-axis: proportion of total emoji usage from each version (shown every four months).

Innovation Diffusion on Social Media. Research has explored the diffusion patterns of various online content like hashtags, memes, and news (Ma et al., 2014; Johann and Bülow, 2019; Kümpel et al., 2015; Zhou et al., 2021), including studies on general language innovation (Grieve et al., 2018; Kershaw, 2018). We argue that the standardized, versioned nature of Unicode emojis presents a unique, controlled setting to study language unit adoption. Therefore, this work specifically investigates the diffusion of new emojis through the lens of innovation diffusion theory (Rogers et al., 2014; Ma et al., 2014), examining factors influencing their uptake and popularity.

Temporal Effects on Models. Language evolution on social media causes a temporal shift, degrading the performance of models trained on older data when applied to newer text (Röttger and Pierrehumbert, 2021; Huang and Paul, 2018; Agarwal and Nenkova, 2022). Newly introduced emojis, appearing as out-of-vocabulary tokens to models trained prior to their release, exacerbate this issue. While methods like continual training can adapt models (Röttger and Pierrehumbert, 2021; Ke and Liu, 2022; Su et al., 2022), they incur significant computational costs. We propose a novel emoji substitution approach as a computationally efficient alternative to help models interpret new emojis without requiring retraining.

3 Emoji Diffusion: Frequency and Semantics

We explore the diffusion of new emojis (Unicode 11.0-13.0, released 2018-2020) on Twitter from May 2018 to May 2022, focusing on usage frequency and semantic evolution, particularly through the lens of sentiment context. Our dataset comprises English tweets collected via the Twitter API[1].

3.1 Emoji Usage Frequency Dynamics

Newly released emojis generally exhibit increasing usage over the subsequent two years (Figure 1), though adoption rates vary dramatically both between versions and within versions. The popularity often follows a power-law distribution, similar to established emojis (Lu et al., 2016), with top emojis being orders of magnitude more frequent than others (Figure 2). Adoption speeds also vary; for example, 🥺 initially lagged behind others in Emoji 11.0 but later surged in popularity, suggesting diffusion across different community boundaries over time. We also observe that older emoji versions (e.g., Emoji 11.0) can eventually plateau or decrease in usage, potentially as users adopt newer alternatives.

While overall usage grows, short-term frequency can be influenced by external events. Fine-grained weekly analysis revealed temporary spikes in usage coinciding with events like New Year's Day and Valentine's Day, confirmed by examining associated keywords during those periods (details in Appendix A).

3.2 Semantic Evolution during Diffusion

Given increasing adoption, we investigate if emoji meanings adapt during diffusion. As a proxy for semantic context, we analyze the sentiment of tweets containing new emojis using Vader (Hutto and Gilbert, 2014), averaging scores over two-month periods.

For most top emojis studied (from Emoji 11.0), the average sentiment score remained relatively constant, suggesting stable interpretations (Figure 3). However, a notable exception is 🥺 (pleading face), whose average sentiment score progressively increased (became more positive) during the first year post-release.

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

This shift suggests an evolution in usage. Initially, 🥺 appeared frequently in negative contexts, but over time, its usage in positive contexts grew significantly (see Appendix Figure 6 for score distributions). Analysis of highly associated words (via PMI) supports this: early associated sentimental words included negative terms like ‘*sad*‘ and ‘*hate*‘, while a year later, positive terms like ‘*prettiest*‘ and ‘*cutest*‘ became prominent (Table 1). This demonstrates that the perceived meaning and application context of an emoji can evolve as it diffuses through the user population.

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

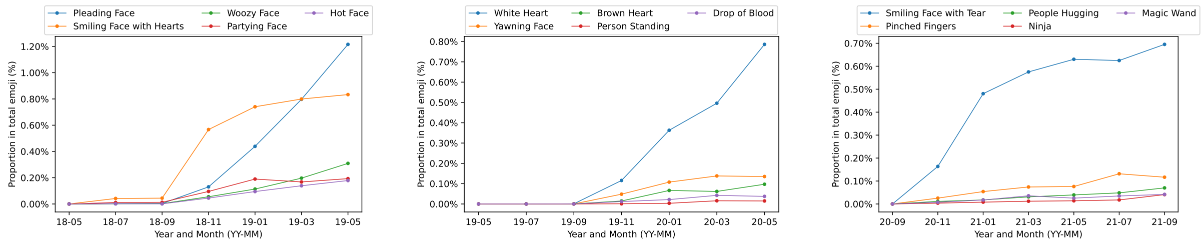
Table 1: Top 10 associated words with positive or negative sentiments of emoji 🥺 in October 2018 and October 2019. Red and blue highlight positive and negative words, respectively. The darker the background color, the more sentimental the words.

4 Influencing Factors of Emoji Diffusion

The observed popularity discrepancies (Section 3) raise the question of what influences the diffusion process. Viewing emojis as digital innovations, we apply Rogers' diffusion of innovation theory (Rogers et al., 2014), examining factors related to the diffusion network (early adopters) and the innovation itself (emoji semantics). We defined early and late stages as two months and fourteen months post-release, respectively.

Diffusion theory suggests larger early adopter communities accelerate spread (Ma et al., 2014; Steffes and Burgee, 2009). We proxied community size by the popularity of hashtags co-occurring with new emojis early on (Yang et al., 2012; Zhou and Ai, 2022) (details in Tables 5 and 6 in Appendix B.1). We found positive Spearman correlations between early hashtag counts and late emoji counts for the top 10 emojis across versions (e.g., $\rho = 0.580, p = 0.08$ for Emoji 11.0; $\rho = 0.530, p = 0.11$ for 12.0), indicating a tendency for larger initial communities to predict greater eventual popularity.

The innovation's characteristics also matter (Rogers et al., 2014). We proxied semantic



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

Figure 2: 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.

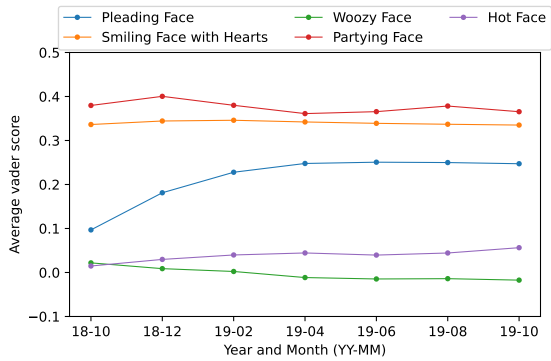


Figure 3: Average Vader scores of the tweets containing the top 5 popular emojis in Emoji 11.0 from October 2018 to October 2019.

popularity using the early-stage frequency of semantically similar words generated by GPT-4 (OpenAI, 2023) (details in Tables 7 and 8 in Appendix B.2). We found significant positive correlation for Emoji 12.0 (Spearman $\rho = 0.780, p < 0.01$; Pearson $r = 0.812$) and positive tendency for 11.0, suggesting emojis representing frequently discussed concepts tend to become more popular. The weaker correlation for 13.0 might reflect limitations in keyword proxies for certain emoji semantics. Overall, both early community size and semantic relevance show associations with emoji diffusion success, aligning with innovation diffusion principles.

5 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 in Emoji 11.0 to 13.0, it relies on the stereotypical associations to infer emojis' semantics and lacks the understanding based on the application scenario, which cannot capture the semantic evolution during diffusion (Zhou et al., 2024). Secondly, the pre-training data of ChatGPT may not cover the recently-released emojis or emojis not included in the Unicode organization. For example, GPT4 generates hallucinations when prompting it to explain the semantics of (broken) and (onlooker), specifically used on the WeChat app (Zhou et al., 2024).

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 used LMs in interpretation work (Lin et al., 2023; Zhou et al., 2023; Romanou et al., 2023) and relied on the attention mechanism to explore the inner association of the fine-tuning data (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.

5.1 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 interpret the emoji's meaning.

5.1.1 Attention Calculation

Formally, we first construct an emoji classification dataset with the input space $x \in \mathcal{X}$ and the pre-defined output space $y \in \mathcal{Y} = \{0, 1, \cdots, n\}$, where x is a tweet excluding the emoji and y represents the emoji label. 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 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 $\{t_i^1, t_i^2, \cdots, t_i^m\}$, where m is the token number, we adopt f in the input i and obtain the attention scores $\{a_i^1, a_i^2, \cdots, a_i^m\}$ and the emoji prediction $f(i)$. 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 m^{th} 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 $f(x_i) = k$ and average the attention scores of the token t in these extracted sentences, which can be formulated as

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

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.

5.1.2 Experiment Setup and Results

To verify the effectiveness of the attention method, we first extracted tweets containing emojis from April 2022 to May 2022 using the Twitter API. We randomly selected six emojis from Emoji 13.0, each appearing in more than 1,000 tweets: 🥲 (smiling face with tears), 🥷 (ninja), 🪄 (magic wand), 🤌 (pinched fingers), 🪙 (coin), and 🫂 (people hugging). A balanced dataset was constructed with a total of 50,000 tweets, divided into 7 labels (label 0 for old emojis and labels 1 to 6 for the new emojis), ensuring each emoji had a sufficient number of tweets for accurate fine-tuning.

We fine-tuned the RoBERTa model pre-trained on tweets (Barbieri et al., 2020) (`twitter-roberta-base`[2]) on the emoji prediction dataset with an 8:1:1 train-validation-test split, achieving a training accuracy of 78.29% and a test accuracy of 66.98%. The top 10 words with the highest attention scores are presented in the second column of Table 2, with words recognized by the authors as having similar semantics highlighted in bold.

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

Emoji	Top 10 Attention Score Words	Top 3 Inference Score Old 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 similar inference scores for emoji inference of the new emojis from Emoji 13.0. High-attention words suggest words with similar semantics and the application scenario of emojis.

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, the 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 🥲 suggest 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.

5.2 Interpret Emojis with Old Emojis

Words with high attention scores do 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 infer-

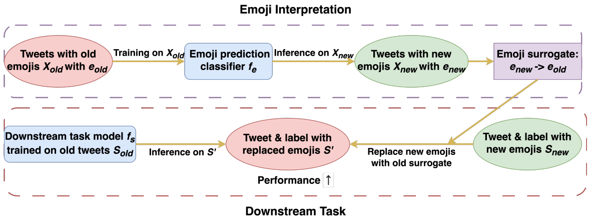


Figure 4: The upper half shows the framework of using old emojis to interpret new emojis and the lower half presents the pipeline of replacing new emojis in the downstream prediction task dataset to old similar emojis to enhance the results.

ence 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 4.

5.2.1 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_e 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:

$$f_e(x_{new}) = \underset{e_{old}}{\operatorname{argmax}}\, p(e_{old}|x_{new}).$$

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}:

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

We sort the old emojis for each new emoji by the calculated $simi(\cdot)$ values and extract the top three old emojis with the highest similarity values for each new emoji.

5.2.2 Experiment Setup and Results

We first constructed $\mathcal{X}_{old}$ using tweets from April 2020 to May 2020, before Emoji 13.0 appeared on Twitter. We extracted 10,000 tweets for each of the top 32 emojis from April to May, along with an additional 10,000 tweets without any emojis. This resulted in a training dataset, $\mathcal{X}_{old}$, comprising 330,000 tweets and 33 emoji labels (32 emojis plus a "no emoji" label). The pre-trained RoBERTa model, f_e, was fine-tuned on $\mathcal{X}_{old}$, and the fine-tuned model was then used to infer on the test dataset constructed with tweets from 2022 containing the emojis from Emoji 13.0, as described in Section 5.1. We calculated the semantic similarity values between the old and new emojis using Equation 1 and presented the top 3 similar ones for each new emoji, along with their similarity values, in the third column of 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 will demonstrate how this associated emoji analysis can be applied to downstream tasks. We find that simply substituting new emojis with their similar old emojis, as emoji surrogates, significantly increases the accuracy in the sentiment classification and irony detection task of fine-tuned language models.

5.3 Emoji Substitution in Downstream Tasks

Sentiment classification and irony detection are two well-studied natural language processing (NLP) tasks widely used in many deployed systems. Emojis, as non-verbal tokens rich in semantics, have been shown to be important for training effective machine learning models for sentiment classification or irony detection (Chen et al., 2019; Hayati et al., 2019).

The current state-of-the-art (SoTA) method for text classification tasks on tweets involves fine-tuning pre-trained language models on a training dataset. However, given a fine-tuned language model on a tweet dataset from 2020, testing on tweets from 2022 containing the new emojis from Emoji 13.0 may lead to imprecise predictions due to the absence of these new emojis in the fine-tuning data.

To address this issue, instead of fine-tuning models on 2022 tweets again, which requires high computational cost, we propose a method to directly substitute the new emojis with old emoji surrogates known by the fine-tuned language models. We expect the semantics encoded in similar old emojis (emoji surrogates) to compensate for the loss of semantics in the unknown new emojis. We illustrate the emoji substitution process in the downstream classification tasks in the lower half of Figure 4.

5.3.1 Emoji Substitution

Suppose that we have a tweet classification dataset $\mathcal{S}_{new}$ that contains the new emojis and that each tweet is labeled with the label $y \in \{0, 1\}$. Given a classifier f_s fine-tuned in $\mathcal{S}_{old}$, where the new emojis are not in the vocabulary of the classifier f_s, for each tweet $s_i = [t_i^1, t_i^2, \cdots, e_{new}, \cdots, t_i^m] \in \mathcal{S}_{new}$, we can replace the new emoji e_{new} in tweet $s_i \in \mathcal{S}_{new}$ with the old emoji surrogates (top 3 old similar emojis sorted by the calculated similarity value in Section 5.2) 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 $s_i^{'}$ with the old emoji surrogates to the classifier f_s to get the model prediction, which is: $f_s(s_i) = \text{argmax}_{y \in \{0,1\}}\, p(y|s_i^{'})$.

5.3.2 Experiment Setup and Results

For the existing classification datasets for tweets, they were developed before 2018 and did not contain the newly created emojis for Emoji 13.0 or Emoji 14.0. Therefore, we collect tweets from 2020 and 2022, respectively, and ask LLM to annotate the labels. Due to the superiority of LLMs over human crowd-workers on the text annotation task (Gilardi et al., 2023; Zhou et al., 2024), we utilize ChatGPT to annotate the sentiment or irony label for tweets (Ouyang et al., 2022). Given a tweet input s, the ChatGPT model (GPT-3.5 for Emoji 13.0 and GPT-4o for Emoji 14.0), a sentiment label set {positive, neutral, negative}, an irony label set {irony, non-irony}, we ask ChatGPT to annotate the tweet with the given sentiment and irony label sets. We repeat the LLM annotation process twice with the temperature 0.7 and only keep the examples and labels agreed by two LLM annotators.

We first 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 and downsample the training set to make a balanced set with 35,388 tweets. We fine-tune the Roberta classifier (`twitter-roberta-base`) pre-trained on tweets before August 2019 as f_s (Barbieri et al., 2020). We repeat the collection and labeling process on tweets from April 2022 to May 2022 to form $\mathcal{S}_{new}$. In the randomly sampled test dataset, the fine-tuned model f_s can achieve an accuracy of 67.52% in tweets from 2020 and 67.65% in tweets from 2022 in the sentiment classification task, which means that the two-year language evolution is not significant to greatly affect the overall performance of the classifier.

To validate ChatGPT's reliability for tweet annotation, we evaluated its performance on the SemEval-2015 Task 10 sentiment classification test set (Rosenthal et al., 2015). ChatGPT achieved 77.29% agreement with the gold standard labels. This accuracy compares favorably to the reported average human agreement of 75.70% obtained during the dataset's construction. Given this comparable performance to human annotators on a standard benchmark, we consider ChatGPT sufficiently reliable for annotating tweets in our study.

To evaluate the effectiveness of our emoji replacement method for tweets containing new emojis, we randomly selected 16 popular emojis from Emoji 13.0 and Emoji 14.0, each appearing more than 300 times in $\mathcal{S}_{new}$. These included 8 sentiment-related emojis and 8 entity-related emojis. We collected a total of 91,066 tweets containing the selected emojis and used the fine-tuned language model f_s to make predictions for each tweet. Next, we applied the proposed emoji replacement method, substituting the new emojis with existing emoji surrogates, and asked f_s to re-predict. Additionally, we prepared two baseline methods for replacing new emojis with text: one replaced the new emoji with its corresponding emoji name (words representing the emoji's meaning), while the other replaced the emoji with a description generated by ChatGPT, leveraging its detailed understanding of emoji semantics (Zhou et al., 2024). The generated descriptions of ChatGPT for each emoji and the prompt are shown in Table 9 in the Appendix. We computed the prediction accuracy

Emoji 13.0 (sentimental)						
emoji	surrogates	# test	original acc	replaced acc	name acc	description acc
🥲	😔🥺😩	23,080	64.4 ± 1.38	$\mathbf{66.4 \pm 3.53}$	60.9 ± 1.49	58.5 ± 7.77
🥸	🤔👀😔	1,908	55.8 ± 10.0	64.9 ± 1.42	$\mathbf{65.9 \pm 1.97}$	64.1 ± 1.07
🫂	🥺💙✨	6,580	82.4 ± 2.01	$\mathbf{90.4 \pm 1.22}$	86.4 ± 1.63	85.2 ± 3.92
🤌	😍🔥💯	6,972	80.8 ± 1.53	$\mathbf{83.6 \pm 1.05}$	65.7 ± 2.89	64.8 ± 2.85
Emoji 13.0 (entity)						
🪜	👍🔥💥	614	23.9 ± 2.34	20.7 ± 2.53	24.4 ± 1.32	$\mathbf{28.3 \pm 2.06}$
🥷	✨👀😊	6,649	56.1 ± 1.53	$\mathbf{58.4 \pm 2.19}$	56.0 ± 2.74	56.8 ± 2.50
🪙	💥🎉🔥	3,209	56.0 ± 5.22	$\mathbf{61.3 \pm 0.88}$	59.2 ± 0.49	56.3 ± 0.59
🫐	💥👀✨	922	66.5 ± 1.11	67.4 ± 0.82	$\mathbf{68.1 \pm 0.85}$	61.1 ± 5.56
Emoji 14.0 (sentimental)						
🫠	😔✨🥺	9305	$\mathbf{78.7 \pm 0.36}$	77.2 ± 0.88	70.9 ± 5.22	70.5 ± 4.48
🥹	🥺✨💖	9328	80.0 ± 0.92	$\mathbf{82.5 \pm 0.49}$	77.5 ± 3.37	61.7 ± 0.84
🫶	✨💖🥺	9374	76.7 ± 0.07	$\mathbf{81.9 \pm 0.18}$	80.2 ± 2.49	81.2 ± 1.15
🫡	💯✨👀	9333	$\mathbf{80.4 \pm 0.53}$	79.9 ± 1.31	76.7 ± 1.84	78.2 ± 0.49
Emoji 14.0 (entity)						
🪷	✨💥💯	1702	86.2 ± 0.16	$\mathbf{87.7 \pm 0.25}$	86.7 ± 0.75	81.4 ± 1.38
🪩	✨🎉🔥	1325	83.3 ± 0.32	$\mathbf{85.0 \pm 0.48}$	82.2 ± 1.28	79.0 ± 3.15
🪸	💥✨🔥	383	89.2 ± 1.67	$\mathbf{89.4 \pm 1.66}$	86.9 ± 1.10	88.4 ± 0.86
🪺	💥✨💯	382	$\mathbf{93.3 \pm 0.18}$	92.7 ± 0.74	90.8 ± 0.00	89.0 ± 1.41

Table 3: Sentiment prediction accuracy using emoji replacement for Emoji 13.0/14.0. Model pretrained pre-Aug 2019 (new emojis excluded). Accuracy columns compare: original text, replacement with similar old emojis, emoji names, or GPT descriptions.

for original tweets, emoji-replaced tweets, name-replaced tweets, and description-replaced tweets for each emoji. The results of the sentiment classification task are presented in Table 3, and those of the irony detection task are presented in Table 10 in the Appendix.

From Table 3, we observe a notable performance improvement for the model f_s fine-tuned on tweets without new emojis when replacing the new emojis with older emojis of similar semantics. Specifically, the model exhibits a relative improvement of 5.02% for sentimental emojis and 4.71% for entity-related emojis from Emoji 13.0, as well as a relative improvement of 1.78% and 1.51% for sentimental and entity-related emojis from Emoji 14.0, respectively. When compared to replacing new emojis with text (names or generated descriptions), our proposed method outperforms in 11/16 cases. This is likely because the words from emoji names or ChatGPT-generated descriptions fail to fully capture the complex sentiments conveyed by the emojis. Furthermore, as shown in Table 10 in the Appendix, a similar pattern emerges in the irony detection task, where our emoji replacement method achieves the best performance in 13/16 cases. These findings demonstrate the effectiveness and robustness of our proposed method in addressing out-of-vocabulary emoji tokens in downstream text classification tasks.

We also repeat the experiments on the Roberta model with new emojis in the pretrained corpus (`twitter-roberta-base-2022-154m`, pretrained on tweets before December 2022) (Loureiro et al., 2023), and present the results of sentiment classification in Table 11 in the Appendix. Compared with Roberta model pretrained on tweets before 2020 (`twitter-roberta-base`), the average relative improvement of replacing new emojis with old 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 pretraining data and the substitution of emojis causes loss of information.

6 Implications

Our research reveals the patterns of emoji diffusion and proposes a framework to understand the semantics of new emojis. We expect that our exploration of emoji diffusion can benefit the emoji designer by considering the influencing factors to design more attractive emojis in future works. 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 into the diffusion patterns of a new language unit, which can inspire future researchers to generalize our diffusion exploration pipeline and interpretation framework on other new non-verbal tokens, such as memes and viral hashtags. Along with our work, we expect that the exploratory study on the language unit diffusion can reveal the evolution mechanism of the content in digital communities.

7 Conclusion

In this work, we analyzed the diffusion patterns of recently created emojis, finding external events influence short-term frequency and sentiment meanings can evolve during adoption. Early adopter community size and emoji semantics correlate with eventual popularity. We proposed an interpretation framework using associated words and similar existing emojis. Its effectiveness was demonstrated by substituting new emojis with interpreted surrogates, improving model performance on sentiment and irony detection tasks. This framework offers a way to enhance model generalization to new emojis in downstream applications.

8 Limitations

There are two limitations in our work. First, the sample size for our correlation measurement in Section 4 is small and such a small sample size is difficult to obtain the significant correlation (p-value< 0.05). We select 10 emojis from each version to analyze the factors that influence the emoji diffusion, since the emoji number is a long-tailed distribution, where the emoji numbers after the top 10 are close. We measure the association across three versions of emojis to make our conclusion more generalizable.

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.

References

Oshin Agarwal and Ani Nenkova. 2022. Temporal effects on pre-trained models for language processing tasks. *Transactions of the Association for Computational Linguistics*, 10:904–921.

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

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*.

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*.

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.

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.

Shirley Anugrah Hayati, Aditi Chaudhary, Naoki Otani, and Alan W Black. 2019. What a sunny day: Toward emoji-sensitive irony detection. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 212–216.

Tianran Hu, Han Guo, Hao Sun, Thuy-vy Nguyen, and Jiebo Luo. 2017. Spice up your chat: the intentions and sentiment effects of using emojis. In *ICWSM 2017*.

Xiaolei Huang and Michael Paul. 2018. Examining temporality in document classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 694–699.

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.

Zixuan Ke and Bing Liu. 2022. Continual learning of natural language processing tasks: A survey. *arXiv preprint arXiv:2211.12701*.

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.

Victoria Lin, Louis-Philippe Morency, and Eli Ben-Michael. 2023. Text-transport: Toward learning causal effects of natural language. *arXiv preprint arXiv:2310.20697*.

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*.

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, 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, 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.

Angelika Romanou, Syrielle Montariol, Debjit Paul, Leo Laugier, Karl Aberer, and Antoine Bosselut. 2023. Crab: Assessing the strength of causal relationships between real-world events. *arXiv preprint arXiv:2311.04284*.

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.

Paul Röttger and Janet B Pierrehumbert. 2021. Temporal adaptation of bert and performance on downstream document classification: Insights from social media. *arXiv preprint arXiv:2104.08116*.

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

Zhaochen Su, Zecheng Tang, Xinyan Guan, Juntao Li, Lijun Wu, and Min Zhang. 2022. Improving temporal generalization of pre-trained language models with lexical semantic change. *arXiv preprint arXiv:2210.17127*.

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, Paiheng Xu, Xiaoyu Liu, Bang An, Wei Ai, and Furong Huang. 2023. Explore spurious correlations at the concept level in language models for text classification. *arXiv preprint arXiv:2311.08648*.

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*.

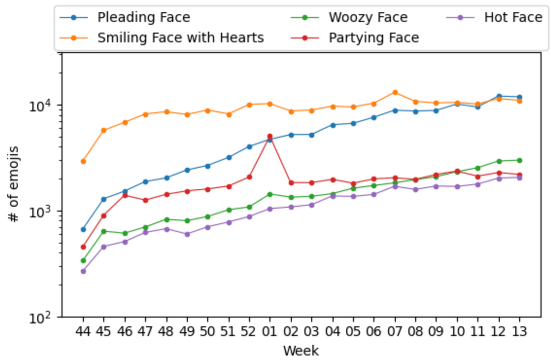


Figure 5: Frequency trends by week of the top 5 popular emojis in Emoji 11.0 from the end of 2018 to the start of 2019.

Emoji	Time period	Top 10 PMI words
🥳	51st week, 2018	birthday, happy, hope, great, days, we, day, **year**, amazing, christmas
🥳	1st week, 2019	happy, **new**, **year**, birthday, **2019**, **happynewyear**, may, everyone, **years**, **2018**
🥰	51st week, 2018	amazing, you, wait, love, thank, thanks, see, beautiful, heart, so
🥰	7th week, 2019	looking, **valentines**, tomorrow, **valentine**, ever, amazing, pretty, sweet, beautiful, you

Table 4: Highly-associated words (quantified by PMI value) of emoji 🥳 and 🥰 in different weeks. The semantics of the associated words coincide with the external events happened at that week.

Appendix

A Weekly Frequency Fluctuations and Event Correlation

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 5. 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 4 and highlight the words related to external events, as recognized by the authors.

From Table 4, 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.

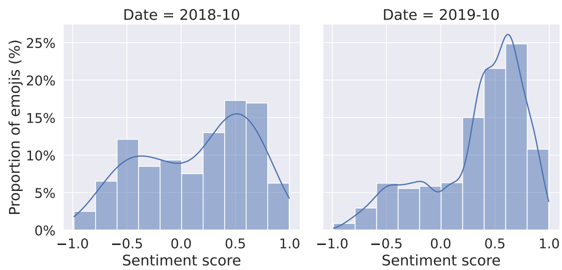


Figure 6: Distribution of Vader scores of tweets containing emoji 🥺 in October 2018 and October 2019. During one year period, 🥺 concentrate more on the positive tweets.

B Details of Influencing Factor Analysis

B.1 Influence of Community Size on Emoji Popularity

We first identify the early stages of Emoji 11.0, Emoji 12.0, and Emoji 13.0 as July 2018, September 2019, and November 2020, respectively, two months after their initial release, and set the late stage as one year after the early stage. We count the number of co-occurring hashtags for each emoji in the early stage and the number of emojis in the late stage. We present the top 10 most used emojis in the late stage from Emoji 13.0, along with their top 3 popular co-occurring hashtags, in Table 5 and the count of emojis and hashtags in the early and late stages of emoji diffusion. The co-occurring hash-

Emoji	Co-occurred hashtags	# 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 5: 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.

tags for emojis in Emoji 11.0 and 12.0 are shown in Table 6.

B.2 Influence of Emoji Semantics on Emoji Popularity

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 a new emoji from Emoji 11.0, 12.0, or 13.0, such as 🥰 (smiling face with hearts). We count the average number of words in the early stage of new emojis (the same month as in Section B.1) and the emoji number in the late stage (one year after the early stage). Since we find that the emoji and word counts are on the same scale, we calculate the Spearman 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 numbers for the top 10 popular emojis from Emoji 13.0 in Table 7, and for the similar words for emojis in Emoji 11.0 and 12.0, we show them in Table 8 in the Appendix.

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 6: 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 7: 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 8: 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.

Emoji	ChatGPT Description
Emoji 13.0	
🥲	Bittersweet smile
🥸	Trying to act clever
🫂	Warm embrace
🤌	Perfectly done
🪜	Step up
🥷	Stealthy move
🪙	Shiny coin
🫐	Juicy blueberry
Emoji 14.0	
🫠	Feeling overwhelmed
🥹	Teary gratitude
🫶	Heartfelt gesture
🫡	Respectfully noted
🪷	Serene lotus
🪩	Disco vibe
🪸	Coral reef
🪺	Cozy nest

Table 9: ChatGPT-generated descriptions for emojis from Unicode Emoji 13.0 and 14.0. The prompt used for generating descriptions is: *Please use a few words to replace the emoji: {emoji}, preserving its meaning and ensuring the text reads naturally. Only give me the most suitable substitution.*

Emoji 13.0 (sentimental)						
emoji	surrogates	# test	original acc	replaced acc	name acc	description acc
🥲	😔🥺😩	23,080	72.0 ± 2.87	$\mathbf{74.6 \pm 1.68}$	74.3 ± 0.45	73.1 ± 0.15
🥸	🤔👀😔	1,908	56.7 ± 4.67	$\mathbf{69.7 \pm 7.85}$	67.0 ± 5.15	68.4 ± 4.96
🫂	🥺💙✨	6,580	95.0 ± 0.81	$\mathbf{95.8 \pm 0.06}$	93.9 ± 0.46	95.5 ± 0.08
🤌	😍🔥💯	6,972	85.7 ± 0.38	$\mathbf{86.8 \pm 0.16}$	85.8 ± 0.13	86.2 ± 0.32
Emoji 13.0 (entity)						
🪜	👍🔥💥	614	96.8 ± 0.45	$\mathbf{97.1 \pm 0.21}$	96.9 ± 0.31	96.9 ± 0.22
🥷	✨👀😊	6,649	97.0 ± 0.05	97.1 ± 0.15	97.1 ± 0.02	$\mathbf{97.2 \pm 0.02}$
🪙	💥🎉🔥	3,209	97.0 ± 0.06	97.1 ± 0.09	97.2 ± 0.06	$\mathbf{97.3 \pm 0.25}$
🫐	💥👀✨	922	97.6 ± 0.08	$\mathbf{97.7 \pm 0.07}$	97.4 ± 0.12	97.5 ± 0.04
Emoji 14.0 (sentimental)						
🫠	😔✨🥺	9305	78.2 ± 0.53	$\mathbf{78.8 \pm 0.52}$	76.2 ± 0.07	77.5 ± 0.30
🥹	🥺✨💖	9328	93.6 ± 0.46	$\mathbf{94.0 \pm 0.03}$	92.0 ± 0.88	93.4 ± 0.54
🫶	✨💖🥺	9374	96.2 ± 0.19	$\mathbf{96.5 \pm 0.01}$	96.1 ± 0.35	96.4 ± 0.13
🫡	💯✨👀	9333	84.4 ± 0.84	$\mathbf{84.5 \pm 0.83}$	76.2 ± 5.78	84.1 ± 0.37
Emoji 14.0 (entity)						
🪷	✨💥💯	1702	97.1 ± 0.24	$\mathbf{97.2 \pm 0.08}$	96.7 ± 0.31	97.0 ± 0.39
🪩	✨🎉🔥	1325	95.5 ± 0.38	95.6 ± 0.24	95.1 ± 0.19	$\mathbf{95.7 \pm 0.20}$
🪸	💥✨🔥	383	98.3 ± 0.14	$\mathbf{98.7 \pm 0.14}$	97.9 ± 0.14	97.6 ± 0.00
🪺	💥✨💯	382	96.0 ± 0.41	$\mathbf{96.1 \pm 0.28}$	95.2 ± 0.27	95.3 ± 0.25

Table 10: Results of the emoji replacement method on irony detection tasks for selected emojis from Emoji 13.0 and Emoji 14.0, using models pretrained on data prior to August 2019 (excluding new emojis in the pre-training data). All values are reported in percentage.

Emoji 13.0 (sentimental)				
emoji	surrogates	# test	ori acc	replaced acc
🥲	😔🥺😩	23,080	$\mathbf{66.57 \pm 0.03}$	60.63 ± 0.06
🥸	🤔👀😔	1,908	58.22 ± 1.61	$\mathbf{68.10 \pm 2.71}$
🫂	🥺💙✨	6,580	83.86 ± 4.42	$\mathbf{91.33 \pm 0.35}$
🤌	😍🔥💯	6,972	83.14 ± 0.52	$\mathbf{83.90 \pm 0.16}$
Emoji 13.0 (entity)				
🪜	👍🔥💥	614	$\mathbf{29.36 \pm 9.85}$	24.60 ± 9.94
🥷	✨👀😊	6,649	$\mathbf{58.69 \pm 0.95}$	57.45 ± 3.36
🪙	💥🎉🔥	3,209	63.05 ± 1.03	$\mathbf{66.55 \pm 0.42}$
🫐	💥👀✨	922	67.95 ± 2.83	$\mathbf{68.17 \pm 1.62}$

Table 11: Results of emoji replacement method on the sentiment prediction task for selected emojis from Emoji 13.0 on models pre-trained on data before December 2022, including new emojis in the pre-training data.