# Exploring (Dis-)Similarities in Emoji-Emotion Association on Twitter and Weibo


Mingyang Li
University of Pennsylvania
Philadelphia, PA, USA
myli@seas.upenn.edu

Sharath Chandra Guntuku
University of Pennsylvania
Philadelphia, PA, USA
sharathg@sas.upenn.edu

Vinit Jakhetiya
Indian Institute of Technology
Jammu, J&K, India
vinit.jakhetiya@iitjammu.ac.in

Lyle H. Ungar
University of Pennsylvania
Philadelphia, PA, USA
ungar@cis.upenn.edu



## ABSTRACT

Emojis have gained widespread acceptance, globally and cross-culturally. However, Emoji use may also be nuanced due to differences across cultures, which can play a significant role in shaping emotional life. In this paper, we a) present a methodology to learn latent emotional components of Emojis, b) compare Emoji-Emotion associations across cultures, and c) discuss how they may reflect emotion expression in these platforms. Specifically, we learn vector space embeddings with more than 100 million posts from China (Sina Weibo) and the United States (Twitter), quantify the association of Emojis with 8 basic emotions, demonstrate correlation between visual cues and emotional valence, and discuss pairwise similarities between emotions. Our proposed Emoji-Emotion visualization pipeline for uncovering latent emotional components can potentially be used for downstream applications such as sentiment analysis and personalized text recommendations.



## CCS CONCEPTS

• **Human-centered computing → Social networks**.

## KEYWORDS

Emoji; Emotions; Weibo; Twitter; China; United States

**ACM Reference Format:**

Mingyang Li, Sharath Chandra Guntuku, Vinit Jakhetiya, and Lyle H. Ungar. 2019. Exploring (Dis-)Similarities in Emoji-Emotion Association on Twitter and Weibo. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308560.3316546


## 1 INTRODUCTION

Across countries, interpretations of Emojis may be nuanced due to cultural differences, which can play significant role in shaping emotional life [17]. Same Emojis can be differently associated with emotions to people from different cultures [39]. If quantitatively analyzed, such differences can inform research for cross-cultural linguists, personalized recommender systems for travelers, and downstream applications for creative designers who employ Emojis in their products. However, related quantitative research often focuses on usage statistics within individual countries, such as the *SwiftKey* Emoji Report [38], and cross-cultural comparison in Emoji-Emotion similarities is still a growing field of research.

In this paper, we present a methodology to uncover latent emotional components associated with Emojis using more than 100 million posts from China (Sina Weibo) and the United States (Twitter). These platforms are the closest analogs in both countries considering the demographics and similarities in degrees of affective and informative content [11, 23, 25]. Starting with learning vectorial representations for tokens in each platform, we quantify similarities of Emojis with 8 basic emotions. Immediately, this enables us to infer correlation between visual cues and emotional valence, as demonstrated in Figure 1. Furthermore, Emoji-Emotion associations can be compared across distinctive languages, enabling more possibilities in cross-cultural research.

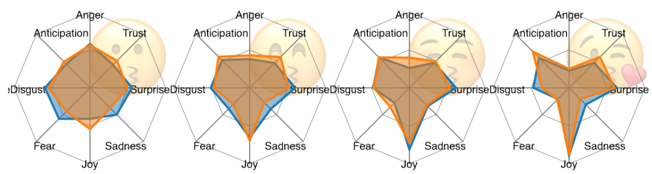


**Figure 1: Associations of kissing face Emojis to 8 basic emotions. On each radar chart (orange: Sina Weibo; blue: Twitter), Center points, inner octagons, and outer octagons indicate −100%, 0, and 100% similarities, respectively. With more visual cues (curly eyes, blushes, and a heart mark) added to the Smiley Emoji, the emotional valence increases.**

## 2 BACKGROUND AND RELATED WORK

*Weibo and Twitter Platforms.* Social media posts on Twitter and Weibo are predictive of several traits, including users' demographics [36, 45], personality [13, 21], location [35, 46], psychological stress [12, 22], and mental health [15, 40]. In this paper, we study latent emotional components in Emojis using Emoji-Emotion associations across the US, represented by Twitter posts, and in China by Weibo posts.



*WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA*



ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3316546

*Differences in Emotion Expressions and Perceptions.* There are similarities as well as differences in how people of different cultural backgrounds express and interpret emotions. First, evolutionary and biological processes generate universal expressions and perceptions of emotions [1, 42]. For example, facial expressions is one of such universal channels that convey emotions across populations [8]. On the other hand, culture can play a significant role in shaping emotional life. Specifically, different cultures may value different types of emotions (e.g., Americans value excitement while Asians prefer calm) [41], and there are different emotional display rules across cultures [27]. In general, psychological research reveals both cultural similarities and differences in emotion expressions and perceptions [9]. Prior works also suggest that culture plays a key role in predicting perceptions of affect [14, 16, 47]

*Lexical Approach to Measuring Emotions.* Psychologists have long been employing curated lexicons in their researches [18]. For emotion analysis, the NRC Word-Emotion Association Lexicon (EmoLex, or simply NRC) is a popular dictionary consisting of 14, 182 unigrams mapped to 10 'Affect Categories' [30]. Twitter is one of the platforms where EmoLex has been shown to be helpful in predicting sentiment analysis [6]. Prior work has also compared Twitter to Weibo using lexical approach in emotion usage [37].

However, cross-platform, and specifically cross-cultural, differences in Emoji-Emotion associations are less explored. Analyzing use of Emojis between the US and Chinese contexts provides an opportunity to assess a plethora of behaviors related to emotion expression and, arguably, emotional symbols that are culturally embedded.

*Prior Studies on Emojis.* Prior Emoji-related psychological research mainly focuses on (1) rendering-specific interpretations to Emojis, (2) Emoji-assisted sentiment annotation, and (3) embedding-derived insights of Emoji use. In the former topic, researchers systematically studied Emoji misinterpretation based upon rendering used by different operating systems (Android, iOS and Windows) [29]. Cross-platform mapping to correct the misinterpretation by applying it to the task of sentiment analysis was also proposed [31]. In sentiment-focused studies, a linguistic approach is often employed. For example, researchers studied causal inference to test whether adopting Emojis would cause individual Twitter users to employ fewer emoticons in their text [32] and also examined the effect of diversity of Emoji set in learning representations of emotional content in texts [10]. With the growing popularity and promise of word embedding algorithms, Emojis' semantic association with different word tokens using distributional semantics was carried out [2]. Meanwhile, there were also studies focused on how Unicode descriptions could indicate sentimental components in the Emojis [7]. Later, EmojiNet, the largest machine-readable Emoji sense inventory that links 2, 389 Unicode Emoji representations to their English meanings extracted from the Web was released [44]. Researchers also studied the non-compositional nature of multi-Emoji expressions using frequency distribution of the words and sentiment analysis of the tweets containing them [24].

# 3 METHODS

## 3.1 Data Collection

This study received approval from authors' Institutional Review Board (IRB). We obtained Twitter data from a 10% archive released by the *TrendMiner* project [33], which exploited a streaming API from Twitter. Since Weibo lacks a streaming interface (as Twitter) for downloading random samples over time, we queried for all posts from individual users. The list of users were crawled using a breadth-first search strategy beginning with random users. On both platforms, we limited our analysis to posts created in the year of 2014 (to avoid potential confounds in the adoption of skin tones introduced in 2015). We obtained approximately 31.8 million posts on Twitter and 136 million posts on Weibo.

## 3.2 Pre-processing

*Geo-location:* On Twitter, the coordinates and tweet country location (whichever was available) was used to geo-locate tweets. On Weibo, user's self-identified profile location was used to identify the geo-location of messages.

*Language Filtering:* To remove the confounds of bilingualism, we filtered posts by the languages in which they were composed. Language used in each post is detected using a pre-trained *fastText* model released by Facebook [4, 19, 20]. Non-English tweets in the Twitter corpus and non-Mandarin posts on Weibo are removed. Particularly in the Weibo dataset, traditional Chinese characters were converted to Simplified Chinese using *hanziconv* Python package [1] to conform with the EmoLex lexicon used in later sections. We also remove any direct re-tweets (indicated by '`RT @USERNAME:`' on Twitter and '`@USERNAME//`' on Weibo).

*Tokenizing:* Twitter text was tokenized using *Social Tokenizer* bundled with *ekphrasis* [3], a text processing pipeline designed for social networks. Weibo posts were segmented using *Jieba*[2] considering its ability to discover new words and Internet slang, which is particularly important for a highly colloquial corpus like Sina Weibo. Using *ekphrasis*, URLs, email addresses, percentages, currency amounts, phone numbers, user names, emoticons and time-and-dates were normalized with meta-tokens such as '`<url>`', '`<email>`', '`<user>`' etc.

## 3.3 Training Embedding Models

To study the lexical semantics on both platforms, *fastText* models were trained on each corpus. These models were trained for 10 epochs with learning rates initialized at .025 and allowed to drop till $10^{-4}$. The dimension of learned token vectors was chosen to be 100 based on previous work. To counter side effects due to the randomized initialization in the *fastText* algorithm, each model was trained for 5 times independently with identical parameters. The 10 models are referred to as $\{m_{t,1}, m_{t,2}, ..., m_{t,5}, m_{w,1}, m_{w,2}, ..., m_{w,5}\}$, where $c \in \{t, w\}$ refers to Twitter and Weibo, respectively. Each model can provide a vectorial representation to a token of interest.

To provide an intuitive view on the vectors obtained, we compressed the vectorial representations of Emojis from $m_{t,1}$ and

[1] https://pypi.org/project/hanziconv/
[2] https://github.com/fxsjy/jieba

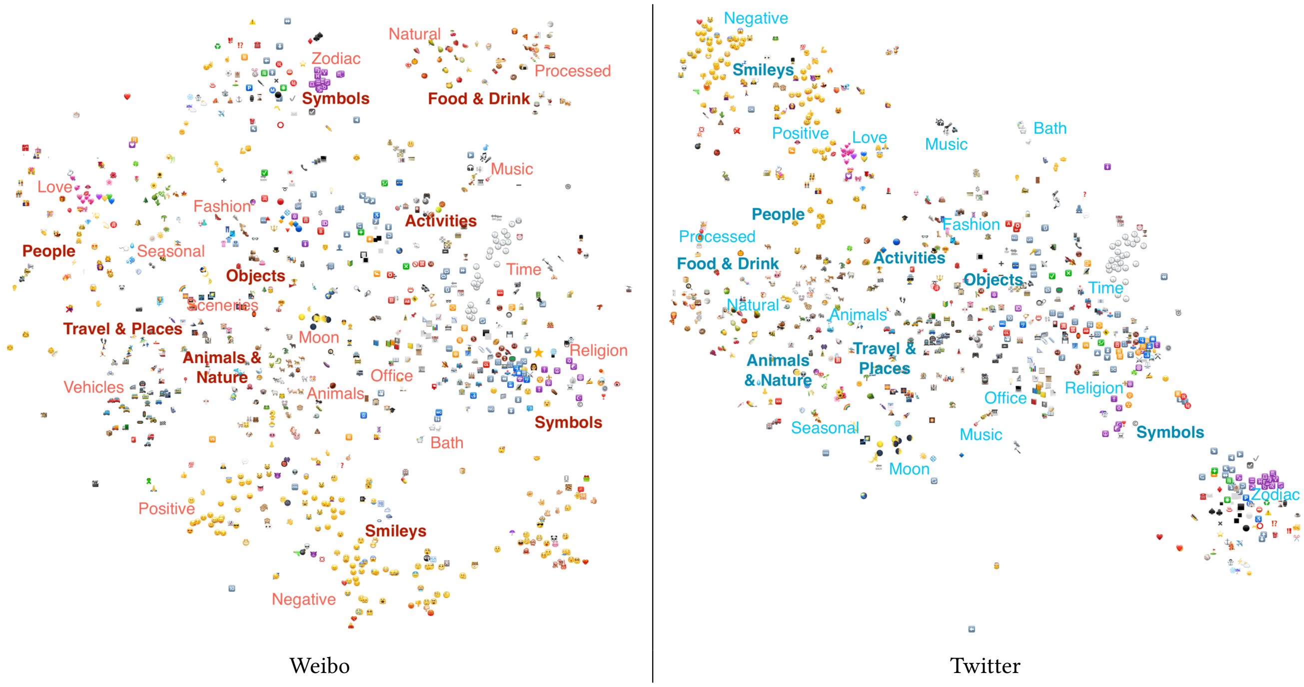


**Figure 2: t-SNE visualization of the learned vectors for Emojis on both platforms. Weibo is shown on the left, and Twitter is put on the right. Textual labels are manually added afterwards for clarity. Dark labels are names of Unicode categories, while light-colored labels are manually created classifications that does not necessarily map to Unicode categories.**

$m_{w,1}$ into two dimensions using the t-SNE algorithm [26]. The two-dimensional coordinates are plotted in Figure 2 side by side. Naturally-forming clusters are labeled in-place for clarity.

On both platforms, Emojis tend to cluster by semantic usage. Two patterns are worth noticing: 1) coherent to how bag-of-words model works [28], tokens with limited possible usages appear closer (e.g. clock faces, zodiac signs, and moon phases), while less topic-specific Emojis tend to spread away (e.g. smileys, objects, and symbols). 2) clusters resemble the categories of Emojis defined by the Unicode Consortium[3]. For example, the clusters formed by clock-face Emojis may be resulted from the common usage of the sentence structure, 'Let's meet at [clock-face Emoji here]'. Similarly, the cluster of transportation tools may be a result of many 'Let's go there by [transportation Emoji]' sentences. Despite that t-SNE visualizations can be misleading at times [43], these clearly separated clusters here suggest that our *fastText* training has been successful in capturing usages of Emojis.

### 3.4 Projecting Emojis onto Emotions

Within each model, $m_{c,i}$ where corpus $c \in \{t, w\}$ and instance $i = 1, ..., 5$, for each of the 8 basic emotions in the EmoLex, $T \in$ {Anger, Joy, ..., Trust}, all tokens associated with this emotion are averaged into a (model-specific) 'emotion vector' $\vec{T}_{c,i}$. Since all EmoLex tokens are verbal, which can render as a bias when their corresponding vectors are compared with those of the Emojis (which are non-verbal tokens), the average vector of all 8 emotions are then subtracted from each of the 8 emotion vectors, essentially zeroing out the center of the polyhedron defined by the 8 emotions. The average of all Emoji vectors is also subtracted from each of the Emojis for similar reason. For each pair of emotion ,$\vec{T}_{c,i}$, and Emoji, $\vec{J}_{c,i}$, in this model, their cosine similarity, denoted by $s_{J,T,c,i}$, is computed. The similarities between all Emojis and all emotions are then assembled into a 'similarity matrix' $\mathbf{s}_{c,i}$. The 5 similarity matrices across the 5 instances are then averaged for stability. The result is denoted by $\mathbf{s}_c$.

The similarity matrix $\mathbf{s}_c$ can be interpreted as follows. Each row of the matrix, $\vec{s}_{J,c}$, specifies the similarities between the Emoji $J$ to each of the 8 emotions. Every column of the matrix, $\vec{s}_{T,c}$, contains the similarities between the emotion $T$ to each of the Emojis. In other for the matrices to be comparable, we consider only the Emojis appeared in both platforms.

Data and source code to replicate the results in this paper are available at https://github.com/tslmy/EMOJI2019.

## 4 RESULTS AND DISCUSSION

### 4.1 How are Emojis associated with different Emotions?

For each Emoji $J$, its 2 similarity vectors (described in the previous section), $\vec{s}_{J,\mathrm{w}}$ and $\vec{s}_{J,\mathrm{t}}$, are plotted as 2 overlapping polygons. This radar chart, Figure 3, compares the (dis-)similarities between several Emojis (in every Unicode category) to each of the 8 basic emotions.

In the radar charts, center points, inner octagons, and outer octagons indicate −100%, 0, and 100% similarities, respectively. For

[3] https://unicode.org/emoji/charts/full-emoji-list.html

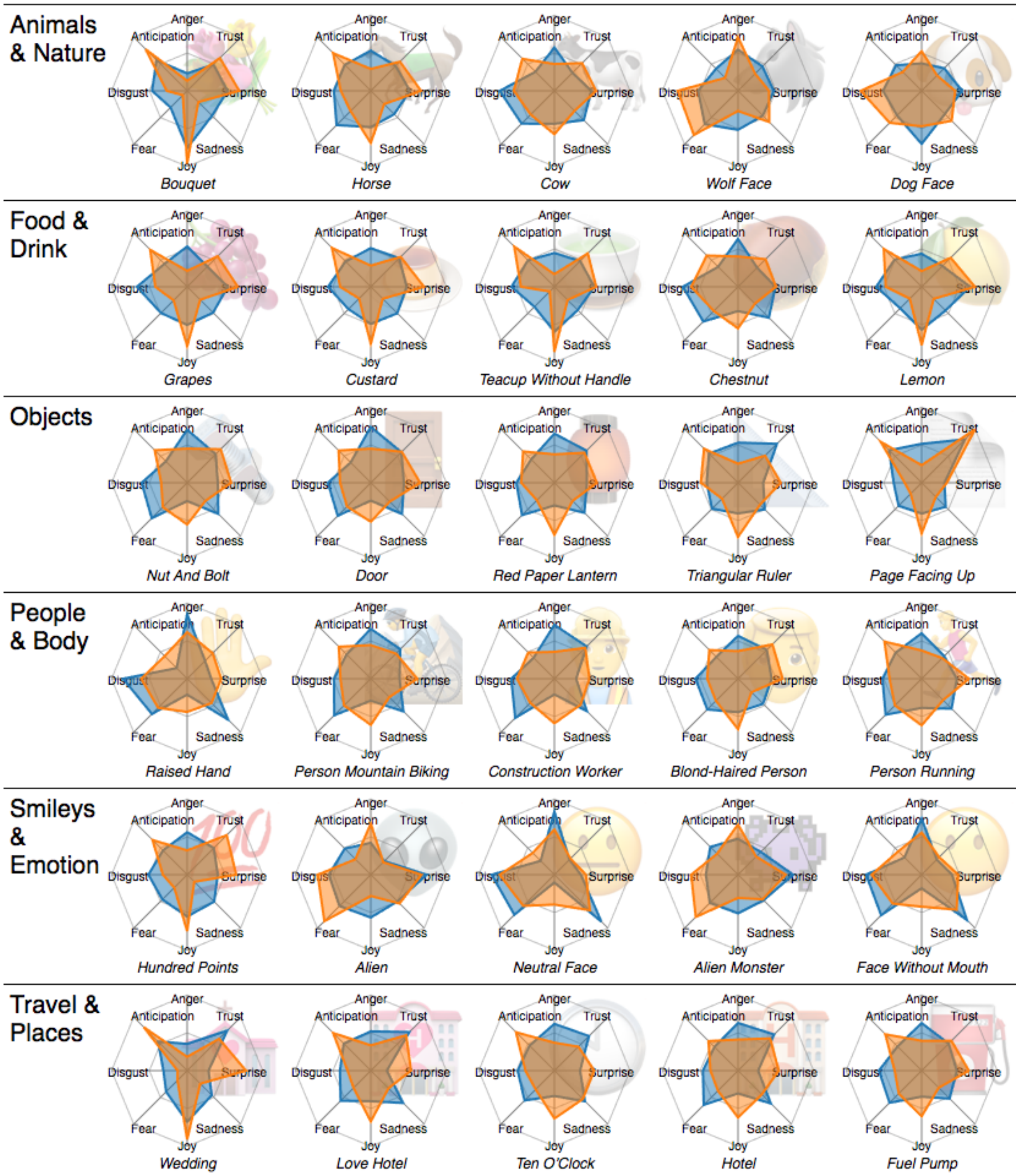


**Figure 3: Radar chart comparing Emoji-Emotion similarities across the two platforms (orange: Weibo; blue: Twitter). From left to right, Emojis are sorted in increasing order of cross-cultural similarity (i.e., decreasing sum of absolute difference across platforms over the 8 emotions). Center points, inner octagons, and outer octagons indicate $-100\%$, $0$, and $100\%$ similarities, respectively.**

example, the Running Man Emoji 🏃 has near-zero similarities to all 8 emotions, indicating the potential emotionally neutral usage this Emoji.

Another counter-intuitively neutral Emoji is the Kissing Face Emoji 😗 featuring '3' shaped mouth and dotted eyes. It is interesting to compare this Emoji with the following Emojis: Kissing Face With Smiling Eyes 😙, Kissing Face With Closed Eyes 😚, and

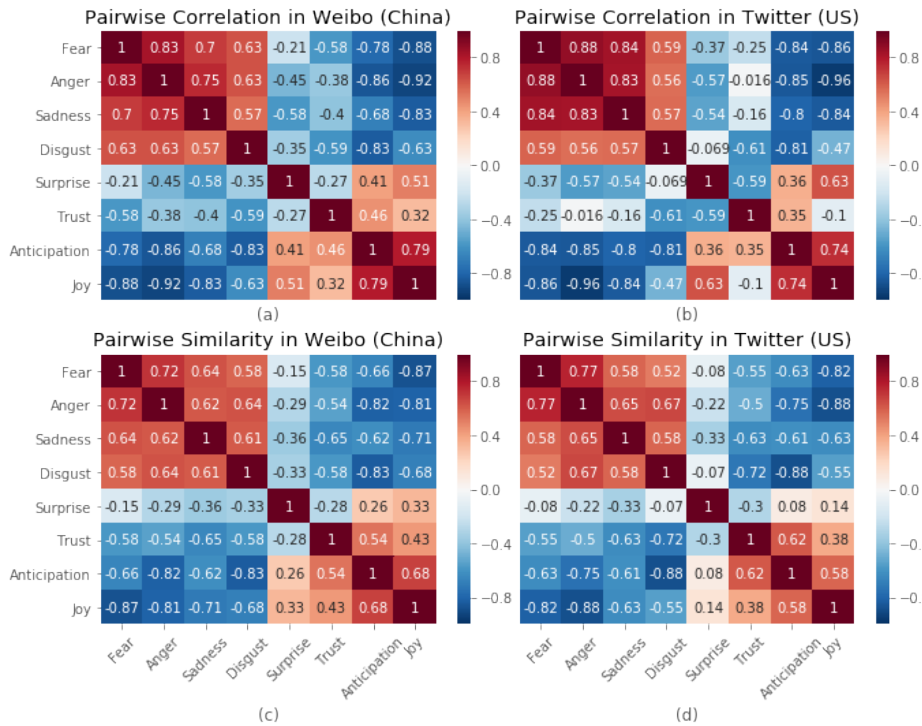


**Figure 4: Pairwise correlations of emotions based on Emoji-Emotion cosine similarities in (a) Weibo and in (b) Twitter. Pairwise cosine similarities of emotions based on NRC emotion vectors in (c) Weibo and in (d) Twitter.**

Face Blowing a Kiss 😘, as shown in Figure 1. With graphical features added, namely smiling eyes in the second, rosy cheeks in the third, and finally a heart in the forth, the radar charts demonstrate increasing positive valences in the emotions.

Besides comparing a set of Emojis with some similar features, the radar charts also enable us to compare emotional association of each Emoji across platforms, and potentially cultures. Each row in Figure 3 are Emojis segregated based on Unicode categories. Emojis that tend to be associated with negative feelings have large similarity values with Anger, Disgust, Fear, and Sadness, with a negative value in Joy and in Trust. Emojis categorized to be associated with positive feelings, on the other hand, usually have large positive value in Joy, negative values in Anger, Fear, and Sadness, and near-zero similarities with Anticipation, Surprise, and Disgust. This suggests recognizable normative and also culture specific patterns of Emoji-Emotion associations.

## 4.2 Pairwise Correlations of Emoji-Emotion Association

To further quantify the differences in Emoji-Emotion associations, we examined the pairwise correlations between Emoji-Emotion similarities on Weibo and Twitter. For each platform $c$, the Spearman correlation coefficient (SCCs) matrix of the 8 emotions' similarity vectors, $\vec{s}_{T,c}$, is computed. The two matrices are shown as heatmaps (Figure 4 (a) and (b)) to demonstrate pairwise correlations of emotions in each platform.

From the correlation matrix showing the average between Weibo and Twitter (Figure 4 (c) and (d)), we can observe that negative emotions, namely Fear, Anger, Sadness, and Disgust, are highly correlated with each other. Among the other 4 emotions, Anticipation and Joy are highly correlated, but Trust and Surprise are rather uncorrelated. While negative emotions seem to be universal, Surprise and Trust potentially have cultural nuances. This is particularly interesting when compared to prior research on the recognition scores of the corresponding facial expressions among Americans and Chinese [34].

The most variance correlated pair of emotions is Joy and Trust. While Joy is often (0.32) correlated with Trust in Weibo, they are inversely correlated (−0.10) in Twitter. This suggests that Weibo users treat trustworthiness as an enjoyable personal trait more than Twitter users do, coherent to a claim in [5].

Furthermore, the Surprise-Trust pair is more than twice as negatively correlated to Twitter users than to Weibo users. This may suggest that Twitter users expect more predictable behavior (fewer surprises) from trustworthy people than Weibo users do. Similarly, the heatmaps also suggest that Weibo users may not enjoy surprises as much as Twitter users do.

It is noted that that the correlations computed over Emoji-Emotion similarities are based on emotions expressed only in posts containing Emojis. We compare them pairwise cosine similarities between the 8 NRC emotion vectors. The pairwise similarity matrix between NRC emotions on Twitter and on Weibo are plotted in Figures 4

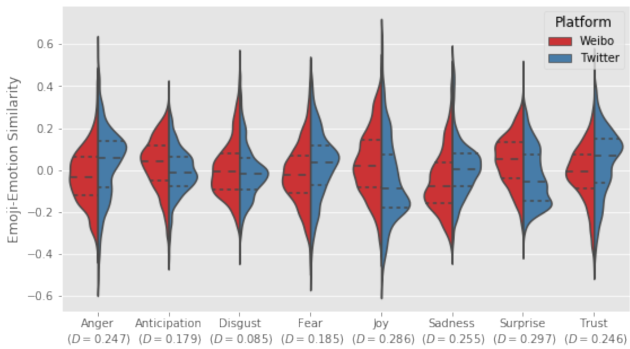


**Figure 5: Violin chart showing distributions of Emoji-Emotion similarities for each emotion. Left halves and right halves show distributions on Weibo and on Twitter, respectively. Quartiles of the distributions are marked with dashed lines. A bandwidth of .2 is chosen to smooth out the distribution when plotting.**

(c) and (d) respectively. Figures 4 (c) and (d) closely resembled Figures 4 (a) and (b), indicating that Emoji-Emotion associations also correlated with NRC Emotion similarities on both platforms.

## 4.3 Distributions of Emoji-Emotion Similarities for Each Emotion

Besides analyzing Emoji-Emotion similarities on a per-Emoji basis and by pairs of NRC emotions, an essential statistic is the distribution of these values. For each emotion $T$, its 2 similarity vectors, $\vec{s}_{T,w}$ and $\vec{s}_{T,t}$, are plotted as two sides of a violin chart, Figure 5. This chart compares how many Emojis the two platforms' users found similar to each of the 8 basic emotions.

An emotion that has a larger mean correlation to Emojis on Twitter than that on Weibo is more directly expressed by Twitter users, because this implies that Twitter users find many Emojis more relevant to the said emotion. In this sense, Twitter users express Anger, Fear, Sadness, and Trust more directly than Weibo users do, who more readily display Surprise, Joy, and Anticipation.

We report the Kolmogorov-Smirnov $D$ statistic for each emotion in Figure 5. With a confidence level of 0.001 and identical numbers of Emojis considered (843), any emotion with a $D$-value exceeding 0.090 can be recognized as having distinct distributions of similarities to Emojis. All emotions except Disgust fit this criterion.

## 5 CONCLUSION

In this paper, we presented a method to learn latent emotional components of Emojis by learning vectorial representations of tokens on large-scale social media corpus from China (Weibo) and US (Twitter). The universality of the Unicode character set, to which modern Emojis belong, enabled us to compare Emojis across cultures, even if the selected languages share no common character.

Comparing latent emotional components of similar Emojis enables us to infer correlations between visual cues and emotion categories. As an example, we demonstrated in Figure 1 how the addition of visual features can increase the positive valence associated to kissing-face Emojis. In a similar visualization, we presented most differently used Emojis across platforms in Figure 3. Further, we studied pairwise correlations between emotions on each platform. The heatmaps (a) and (b) in Figure 4 implied that Twitter and Weibo users have different views towards Surprise and Trust, at least in terms of Emojis associated with them. The fact that (a) and (b) resembled the pairwise similarity matrices between emotions, (c) and (d), suggested that the correlations are a valid proxy for emotion analysis, strengthening our point. Finally, we studied the distributions of Emoji-Emotion similarities across platforms (Figure 5). We concluded that Twitter users and Weibo users indeed take different set of Emojis as suitable to each emotion with the exception of Disgust.

Though the current work is preliminary and insights are correlational in nature, future research can potentially infer insights into specific large scale psychological and cognitive phenomena by comparing the latent emotional contexts behind Emojis across Weibo and Twitter. Further, it would be interesting to study user behavior on Chinese users on Twitter and English users on Weibo.

## REFERENCES


[1] Richard P Bagozzi, Nancy Wong, and Youjae Yi. 1999. The role of culture and gender in the relationship between positive and negative affect. *Cognition & Emotion* 13, 6 (1999), 641–672.

[2] Francesco Barbieri, Francesco Ronzano, and Horacio Saggion. 2016. What does this Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis.. In *LREC*.

[3] Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis. 2017. DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, Vancouver, Canada, 747–754.

[4] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. *CoRR* abs/1607.04606 (2016). arXiv:1607.04606 http://arxiv.org/abs/1607.04606

[5] Michael Harris Bond. 2010. *The Oxford handbook of Chinese psychology.* Oxford Library of Psychology.

[6] Felipe Bravo-Marquez, Marcelo Mendoza, and Barbara Poblete. 2013. Combining strengths, emotions and polarities for boosting Twitter sentiment analysis. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining.* ACM, 2.

[7] Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel. 2016. emoji2vec: Learning emoji representations from their description. *arXiv preprint arXiv:1609.08359* (2016).

[8] Paul Ekman. 1993. Facial expression and emotion. *American psychologist* 48, 4 (1993), 384.

[9] Hillary Anger Elfenbein and Nalini Ambady. 2003. Universals and cultural differences in recognizing emotions. *Current directions in psychological science* 12, 5 (2003), 159–164.

[10] 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. *arXiv preprint arXiv:1708.00524* (2017).

[11] Qi Gao, Fabian Abel, Geert-Jan Houben, and Yong Yu. 2012. A comparative study of usersâĂŽ microblogging behavior on Sina Weibo and Twitter. In *International Conference on User Modeling, Adaptation, and Personalization.* Springer, 88–101.

[12] Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes Eichstaedt, and Lyle Ungar. 2018. Understanding and Measuring Psychological Stress using Social Media. *arXiv preprint arXiv:1811.07430* (2018).

[13] Sharath Chandra Guntuku, Weisi Lin, Jordan Carpenter, Wee Keong Ng, Lyle H Ungar, and Daniel Preoţiuc-Pietro. 2017. Studying personality through the content of posted and liked images on Twitter. In *Proceedings of the 2017 ACM on web science conference.* ACM, 223–227.

[14] Sharath Chandra Guntuku, Weisi Lin, Michael James Scott, and Gheorghita Ghinea. 2015. Modelling the influence of personality and culture on affect and enjoyment in multimedia. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII).* IEEE, 236–242.

[15] Sharath Chandra Guntuku, J Russell Ramsay, Raina M Merchant, and Lyle H Ungar. 2017. Language of ADHD in adults on social media. *Journal of attention*

*disorders* (2017), 1087054717738083.
[16] Sharath Chandra Guntuku, Michael James Scott, Huan Yang, Gheorghita Ghinea, and Weisi Lin. 2015. The CP-QAE-I: A video dataset for exploring the effect of personality and culture on perceived quality and affect in multimedia. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–7.
[17] Rachael E Jack, Oliver GB Garrod, Hui Yu, Roberto Caldara, and Philippe G Schyns. 2012. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences* 109, 19 (2012), 7241–7244.
[18] Oliver P John, Alois Angleitner, and Fritz Ostendorf. 1988. The lexical approach to personality: A historical review of trait taxonomic research. *European journal of Personality* 2, 3 (1988), 171–203.
[19] Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hervé Jégou, and Tomas Mikolov. 2016. FastText.zip: Compressing text classification models. *CoRR* abs/1612.03651 (2016). arXiv:1612.03651 http://arxiv.org/abs/1612.03651
[20] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of Tricks for Efficient Text Classification. *CoRR* abs/1607.01759 (2016). arXiv:1607.01759 http://arxiv.org/abs/1607.01759
[21] Lin Li, Ang Li, Bibo Hao, Zengda Guan, and Tingshao Zhu. 2014. Predicting active users' personality based on micro-blogging behaviors. *PloS one* 9, 1 (2014), e84997.
[22] Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Qi Li, Jie Huang, Lianhong Cai, and Ling Feng. 2014. User-level psychological stress detection from social media using deep neural network. In *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 507–516.
[23] Xialing Lin, Kenneth A Lachlan, and Patric R Spence. 2016. Exploring extreme events on social media: A comparison of user reposting/retweeting behaviors on twitter and weibo. *Computers in Human Behavior* 65 (2016), 576–581.
[24] Rebeca Padilla López and Fabienne Cap. 2017. Did you ever read about Frogs drinking Coffee? Investigating the Compositionality of Multi-Emoji Expressions. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 113–117.
[25] Lin Ma. 2013. Electronic word-of-mouth on microblogs: A cross-cultural content analysis of Twitter and Weibo. *Intercultural Communication Studies* 22, 3 (2013).
[26] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, Nov (2008), 2579–2605. http://www.jmlr.org/papers/v9/vandermaaten08a.html
[27] David Matsumoto. 1990. Cultural similarities and differences in display rules. *Motivation and emotion* 14, 3 (1990), 195–214.
[28] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
[29] Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. Blissfully happyâĂİ or âĂIJready to fightâĂİ: Varying Interpretations of Emoji. *Proceedings of ICWSM* 2016 (2016).
[30] Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. 29, 3 (2013), 436–465.
[31] Fred Morstatter, Kai Shu, Suhang Wang, and Huan Liu. 2017. Cross-Platform Emoji Interpretation: Analysis, a Solution, and Applications. *arXiv preprint arXiv:1709.04969* (2017).
[32] Umashanthi Pavalanathan and Jacob Eisenstein. 2015. Emoticons vs. Emojis on Twitter: A Causal Inference Approach. *CoRR* abs/1510.08480 (2015). arXiv:1510.08480 http://arxiv.org/abs/1510.08480
[33] Daniel Preotiuc-Pietro, Sina Samangooei, Trevor Cohn, Nicholas Gibbins, and Mahesan Niranjan. 2012. Trendminer: An architecture for real time analysis of social media text. In *Sixth International AAAI Conference on Weblogs and Social Media*.
[34] James A Russell. 1994. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin* 115, 1 (1994), 102.
[35] Bahar Salehi, Dirk Hovy, Eduard Hovy, and Anders Søgaard. 2017. Huntsville, hospitals, and hockey teams: Names can reveal your location. In *Proceedings of the 3rd Workshop on Noisy User-generated Text*. 116–121.
[36] Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar, and Hansen Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1146–1151.
[37] Kaisong Song, Shi Feng, Wei Gao, Daling Wang, Ge Yu, and Kam-Fai Wong. 2015. Personalized sentiment classification based on latent individuality of microblog users. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
[38] SwiftKey. 2015. SwiftKey Emoji Report | The United States | Data. https://www.scribd.com/doc/262594751/SwiftKey-Emoji-Report
[39] Kohske Takahashi, Takanori Oishi, and Masaki Shimada. 2017. Is ☺️ Smiling? Cross-Cultural Study on Recognition of EmoticonâĂŹs Emotion. *Journal of Cross-Cultural Psychology* 48, 10 (2017), 1578–1586. https://doi.org/10.1177/0022022117734372 arXiv:https://doi.org/10.1177/0022022117734372
[40] Xianyun Tian, Philip Batterham, Shuang Song, Xiaoxu Yao, and Guang Yu. 2018. Characterizing depression issues on Sina Weibo. *International journal of environmental research and public health* 15, 4 (2018), 764.
[41] Jeanne L Tsai, Brian Knutson, and Helene H Fung. 2006. Cultural variation in affect valuation. *Journal of personality and social psychology* 90, 2 (2006), 288.
[42] Yukiko Uchida and Shinobu Kitayama. 2009. Happiness and unhappiness in east and west: themes and variations. *Emotion* 9, 4 (2009), 441.
[43] Martin Wattenberg, Fernanda Viégas, and Ian Johnson. 2016. How to use t-sne effectively. *Distill* 1, 10 (2016), e2.
[44] Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, and Derek Doran. 2017. Emojinet: An open service and api for emoji sense discovery. In *Eleventh International AAAI Conference on Web and Social Media*.
[45] Wanru Zhang, Andrew Caines, Dimitrios Alikaniotis, and Paula Buttery. 2016. Predicting Author Age from Weibo Microblog Posts.. In *LREC*.
[46] Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, and Xing Xie. 2015. You are where you go: Inferring demographic attributes from location check-ins. In *Proceedings of the eighth ACM international conference on web search and data mining*. ACM, 295–304.
[47] Yi Zhu, Sharath Chandra Guntuku, Weisi Lin, Gheorghita Ghinea, and Judith A Redi. 2018. Measuring Individual Video QoE: A Survey, and Proposal for Future Directions Using Social Media. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 14, 2s (2018), 30.