Language-based Valence and Arousal Expressions between the United States and China: a Cross-Cultural Examination

Anonymous ACL submission

Abstract

Although affective expressions of individuals have been extensively studied using social media, research has primarily focused on the Western context. There are substantial differences among cultures that contribute to their affective expressions. This paper examines the differences between Twitter (X) in the United States and Sina Weibo posts in China on two primary dimensions of affect - valence and arousal. We study the difference in the functional relationship between arousal and valence (so-called V-shaped) among individuals in the US and China and explore the associated content differences. Furthermore, we correlate word usage and topics in both platforms to interpret their differences. We observe that for Twitter users, the variation in emotional intensity is less distinct between negative and positive emotions compared to Weibo users, and there is a sharper escalation in arousal corresponding with heightened emotions. From language features, we discover that affective expressions are associated with personal life and feelings on Twitter, while on Weibo such discussions are about socio-political topics in the society. These results suggest a West-East difference in the V-shaped relationship between valence and arousal of affective expressions on social media influenced by content differences. Our findings have implications for applications and theories related to cultural differences in affective expressions.

1 Introduction

Subjective expression of affect (how we feel) plays a crucial role in understanding learning outcomes in individuals (Hourihan et al., 2017), their perceptions (Gorn et al., 2001), well-being (Xu et al., 2015), and mental and physical health (Cohen and Pressman, 2006). Multiple theoretical and empirical works have therefore studied the underlying dimensions of affect and their relationship. While there are several models of affective structure, Russell's 2-dimensional circumplex model is one of the popular models (Russell, 1980), where orthogonal valence and arousal are considered its most fundamental dimensions (Yik et al., 1999). Valence is the measure of pleasure (or displeasure) and arousal of activation or sense of energy.

Figure 1: Example posts from Twitter and Weibo dataset. Users from the two platforms express affect differently, closely related to cultural background.

Understanding the relationship between valence and arousal is interesting from empirical, psychometric, and theoretical perspectives. While previous studies suggest several models to describe the valence arousal relationship (Ortony et al., 1990; Lang, 1994), where valence and arousal are correspondingly the horizontal and vertical axis in the circumplex, representing the polarity and intensity of affective states, a "V-shaped" relationship of arousal as a function of valence is one of the most widely tested and accepted (Kuppens et al., 2013; Cacioppo and Gardner, 1999). Arousal is shown to be directly related to the intensity of positive or negative valence with a positivity offset and a negativity bias, with varying levels of cross-cultural support (Kuppens et al., 2017a).

While affective structure and relationship are considered universal, most previous works studying the valence arousal relationship are based exclusively on Western samples and do not consider cross-cultural heterogeneity (Tsai et al., 2006). Culture can have a big impact on one's emotional life. Different cultures value emotions (ideal affect) and have different emotional display standards (Matsumoto, 1990). For example, Americans appreciate enthusiasm (high arousal) to express high positive valence, while Asians more often choose quiet (low arousal) (Tsai et al., 2006). Past research also suggests that culture has an important influence in predicting affect perceptions, which can reflect differences in cultural expectations (Guntuku et al., 2015b; Zhu et al., 2018; Guntuku et al., 2015a). Furthermore, the few studies that revealed significant evidence of the affective differences across Western and Eastern cultures were based on small samples recruited in lab-based environments (Kuppens et al., 2017b). Asking participants to recall recent events and report their emotions can contain measurement bias and other concerns relating to self-report (Tarrant et al., 1993; Winograd and Neisser, 2006). To counter this, there is a call for more research going beyond self-reports to behaviors (Baumeister et al., 2007). The strength of our approach is to examine the generalizability of past cross-cultural findings using actual behaviors of language expressions.

This paper has two contributions. First, we compare the functional relationship between valence and arousal in public affective expression using large-scale social media data, going beyond self-report surveys. Second, we use language samples collected from the United States (in the West) and China (in the East) to elucidate the differences and similarities in the relationship between valence and arousal across cultures by mining ecological expressions from individual timelines on Twitter (X) and Sina Weibo and use thematic content analysis to understand the language markers contributing to such differences and similarities.

Social media data are known to capture individuals' feelings in a naturalistic and ecological setting and have been shown to reliably estimate well-being (Jaidka et al., 2020), sentiment (Preoţiuc-Pietro et al., 2016), psychological traits such as personality (Schwartz et al., 2013), and health (Merchant et al., 2019). The variations in content and user behavior on Weibo and Twitter have been explored in various scenarios (Ma, 2013; Lin et al., 2016b). Despite the difficulties of working with a non-random, non-representative sample of social media users, posts can reveal a variety of psychological qualities and consequences, including users' demographics (Sap et al., 2014; Zhang et al., 2016), personality (Li et al., 2014; Quercia et al., 2011), location (Salehi et al., 2017; Zhong et al., 2015), stress levels (Guntuku et al., 2019b; Lin et al., 2016a), and mental health (Guntuku et al., 2019d; Tian et al., 2018) on both platforms. Examples of posts from the two platforms are shown in Figure 1.

2 Methods

2.1 Data Collection

Our data consist of public messages posted on Weibo and Twitter. To collect Twitter data, we used the survey platform Qualtrics, which included demographic questions such as gender and age. Participants shared their Twitter handles after completing the survey. Users were compensated for participating, and we received informed consent to access their Twitter posts. There were 3,113 users in the Twitter dataset, with around 3.6 million posts until 2016.

Weibo, unlike Twitter, does not offer an API tool for obtaining random samples over time. So, starting with a random set of individuals from a public dataset (Guntuku et al., 2019a), Weibo posts were gathered using a breadth-first search method on users. We obtained over 29 million posts from 2014 from 859,054 people on Weibo. Gender and age were collected from self-reported demographic information on their Weibo profile. Subsetting to users with more than 500 words and with a reasonable age (<100 years) and gender, the dataset consisted of 668,257 Weibo posts from 8,731 users. 500 words were found to be the minimum threshold to obtain reliable psychological estimates from individuals' language (Eichstaedt et al., 2021; Jaidka et al., 2018).

Based on the gender and age distribution of Weibo and Twitter users, we built propensity-score-based matched samples, resulting in 2,191 users each on both platforms with at least 500 words. These matched users had 2.4 million posts on Twitter and 177,042 posts on Weibo. In our matched dataset 67.1% self-reported as being female and 32.9% as male, and the mean age was 26.9 (s.d. 8.8). On Twitter, there were on average 15.6 (s.d. 2.8) words per user and Weibo had 57.3 (s.d. 15.4) words per user.

2.2 Language Preprocessing

To eliminate the confounds of bilingualism (Fishman, 1980), we retain only English posts on Twitter and Mandarin posts on Weibo by using langid (Lui and Baldwin, 2012). Re-tweets are also removed from both datasets ('RT @USERNAME:' on Twitter and '@USERNAME//' on Weibo). Weibo posts were split into tokens using THULAC (Li and Sun, 2009) while Twitter posts were segmented using happierfuntokenizing (DLATK/happierfuntokenizing, 2017) due to their ability to handle emoticons and other social media slang. To eliminate uncommonly used words (outliers), we filtered words with different frequency thresholds for each platform. Words used by fewer than 0.1% of the total posts on Twitter and 0.5% on Weibo were removed from the analysis. Most words are seldom used in language, as they follow a Zipfian distribution. By removing these words, we ensure that the language insights from our research can be generalized to out-of-sample cases.

2.3 Feature Extraction

We extract two open-vocabulary features from Twitter and Weibo: n-grams and topics. We extract contiguous sequences of one or two words (1-2 grams, Kern et al., 2014; Andrew Schwartz et al., 2013) with pointwise mutual information (PMI = 3; Church and Hanks, 1990). This resulted in unique unigrams and bigrams set of 10,477 for Weibo and 12,798 for Twitter. We extracted the normalized distribution of the n-grams for each post in the Weibo and Twitter datasets. We then used 2000 topics generated using Latent Dirichlet Allocation (LDA, Blei et al., 2003) as the second feature to represent users' language in our Twitter and Weibo datasets. We utilized topics generated on much larger datasets to favor high diversity and coverage. 2,000 English topics generated a corpus of approximately 18 million Facebook updates with alpha set to 0.30 to favor fewer topics per document. 2,000 Mandarin topics were generated on the entire set of 29 million Weibo posts with similar parameters set in Mallet (Andrew Schwartz et al., 2013). Inherently, each topic is realized as a set of words with probabilities. Every post is thus scored in terms of its probability of containing each of the 2000 topics, $p(topic, post)$, which is derived from their probability of containing a word, $p(word|post)$, and the probability of the words being in given topics, $p(topic|word)$.

2.4 Valence-Arousal Measurement

The circumplex and vector models of emotion have been broadly used for representing affective states (Russell, 1980; Bradley et al., 1992). In these two-dimensional models, valence is the x-axis, expressing pleasantness and unpleasantness, attractiveness and aversiveness, joy, and sorrow (Frijda, 1986). Arousal is the y-axis, describing the degree of wakefulness, boredom, excitement, and calm. These models allow any affective state, emotion, word, or expression to be represented as a point in the space, regardless of the difference in language, country, or culture. We measure valence and arousal using a validated data-driven lexicon generated based on the circumplex model in both English and Mandarin. We used NRC Valence, Arousal, and Dominance (NRC-VAD) Lexicon (Mohammad, 2018) for Twitter data and its translated version for Weibo data. NRC-VAD consists of valence and arousal weights for more than 20,000 words in English and shows a "boomerang" relationship between two dimensions: extremely positive or negative valence is usually paired with high arousal, while calmness matches low arousal. We subtract 0.5 from all scores to make them zero-centered.

Multilinguality is another reason to choose NRC-VAD as our valence-arousal measurement lexicon. There are over 100 languages available for NRC-VAD by translating English terms using Google Translate (August 2022), and the authors claim that most affective norms are stable across languages. Since an original-translated term pair has the same scores, this lexicon avoids the annotator agreement and scale-matching issue, which are common problems using two different lexica over two languages.

2.5 Valence-Arousal Relationship Models

Kuppens et al., 2013 showed six possible functional relationships between valence and arousal. These models are independence (Model 1), Linear Relation (Model 2), Symmetric V-Shaped Relation (Model 3), and Asymmetric V-Shaped Relations, including asymmetric interception (Model 4), asymmetric slope (Model 5), and asymmetric interception and slope (Model 6). The models' functional representations are shown below:

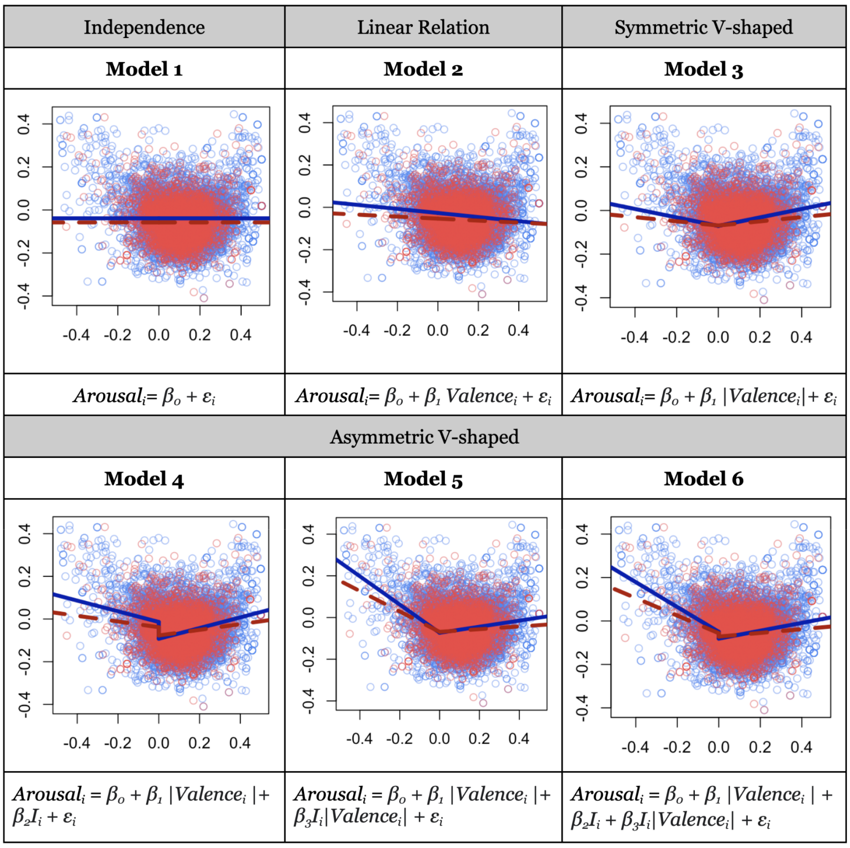

Figure 2: Scatter plots of valence (x-axis) and arousal (y-axis) of Twitter (blue) and Weibo (red) posts. The lines of best fit for each model's function are appended to each plot (Twitter: solid line, Weibo: dashed line). Each model is tested with a within-person intercept and slope.

$$A_i = \begin{cases} \beta_0 + \epsilon_i & \text{(Model 1)} \\ \beta_0 + \beta_1 V_i + \epsilon_i & \text{(Model 2)} \\ \beta_0 + \beta_1 |V_i| + \epsilon_i & \text{(Model 3)} \\ \beta_0 + \beta_1 |V_i| + \beta_2 I_i + \epsilon_i & \text{(Model 4)} \\ \beta_0 + \beta_1 |V_i| + \beta_3 |V_i| + \epsilon_i & \text{(Model 5)} \\ \beta_0 + \beta_1 |V_i| + \beta_2 I_i + \beta_3 |V_i| + \epsilon_i & \text{(Model 6)} \end{cases}$$

Where A_i and V_i are short for $Arousal_i$ and $Valence_i$, arousal and valence scores for the post i, I_i denotes a dummy variable that indicates whether $Valence_i$ is positive($I_i = 1$) or negative($I_i = 0$). Each model is tested with a within-person intercept and slope.

We use Akaike Information Criterion (AIC; Bozdogan, 1987) and Bayesian Information Criterion (BIC; Schwarz, 1978) for model selection. AIC is defined as:

$$AIC = 2 \cdot k - 2 \cdot \ln(\hat{L}) \quad (1)$$

BIC has the following format:

$$BIC = -2 \cdot \ln(\hat{L}) + k \cdot \ln(N) \quad (2)$$

where $\hat{L}$ is the maximized value of the likelihood function of the model, k is the number of parameters, and N is the number of observations. One advantage of using BIC is that it can be used to approximate posterior probability for each model:

$$Pr(model_m|data) = \frac{exp(-0.5BIC_m)}{\sum exp(-0.5BIC_m)} \quad (3)$$

While applying the six models, we use mixed effects models to fit the datasets. We assume there is a fixed relationship between valence and arousal across all posts, while the average level of arousal may vary from user to user. The regression models can correctly represent the relationship between the two variables by setting within-person differences as the random effect.

2.6 Differential Language Analysis

We use ordinary least squares (OLS) regression to identify significant associations between the two feature sets, namely n-grams and topics, and valence and arousal. We use feature sets (n-grams or topics) as input, and valence and arousal as outputs

Dataset	Model	AIC	BIC	PostP
Twitter	**Model 1**	-3.752×10^6	-3.752×10^6	0
	Model 2	-3.790×10^6	-3.790×10^6	0
	Model 3	-3.833×10^6	-3.833×10^6	0
	Model 4	-4.001×10^6	-4.001×10^6	0
	Model 5	-4.048×10^6	-4.048×10^6	0
	Model 6	$\mathbf{-4.060 \times 10^6}$	$\mathbf{-4.060 \times 10^6}$	**1**
Weibo	**Model 1**	-4.155×10^5	-4.155×10^5	0
	Model 2	-4.162×10^5	-4.161×10^5	0
	Model 3	-4.172×10^5	-4.171×10^5	0
	Model 4	-4.219×10^5	-4.219×10^5	0
	Model 5	-4.247×10^5	-4.246×10^5	0
	Model 6	$\mathbf{-4.251 \times 10^5}$	$\mathbf{-4.250 \times 10^5}$	**1**

Table 1: Results of fitting 6 different models on Twitter and Weibo dataset. AIC is Akaike Information Criterion, BIC is Bayesian Information Criterion (the lower the better fit), PostP is posterior probability.

and built one OLS model per feature set. We calculated correlation coefficients of each feature dimension and utilized Benjamini-Hochberg p-correction (Benjamini and Hochberg, 1995) to correct for multiple comparisons and used $p < .05$ for indicating meaningful correlations.

3 Results

Warning: The following section contains swear words.

3.1 Valence-Arousal Relation Models

Among the different models we tested across Twitter and Weibo data, Model 6 with within-person intercept and slope best fit with the lowest AIC and BIC (Table 1). Within-person models were also significantly different from the models without within-person effects. The presence of an asymmetric V-shape in the data, including a negativity bias and negativity offset, was confirmed in the models on both Twitter and Weibo data. Compared with Weibo, Twitter shows a larger intercept gap (Twitter: $\beta_2 = -0.031$; Weibo: $\beta_2 = -0.016$). The intensity of emotion gets significantly stronger with higher positivity/negativity. This conclusion is consistent with both Twitter and Weibo, with the smallest BIC values in Model 6, characterized by a V shape (Twitter: $\beta_1 = 0.573$, Weibo: $\beta_1 = 0.404$) and negativity bias (Twitter: $\beta_3 = -0.392$, Weibo: $\beta_3 = -0.318$). The Twitter model has a steeper slope on both positive and negative valence compared to Weibo (Twitter: $\beta_1 = 0.573$, $\beta_3 = -0.392$; Weibo: $\beta_1 = 0.404$, $\beta_3 = -0.318$).

3.2 Differential Language Analysis

To uncover the content differences in emotional expression across cultures, we utilized differential language analysis to obtain the most correlated n-grams and topics in each platform. Figure 3 shows the top significantly correlated words and phrases with valence and arousal in both platforms. On the dimension of valence, Twitter users tended to use words conveying superlatives ('great', 'awesome', 'amazing') and festive celebrations ('birthday', 'Christmas', 'new', 'win') in expressing positive valence, while profanity ('shit', 'fuck'), negation ('hate', 'bad', 'wrong') and discomfort ('wait', 'tired', 'stop') were indicative of negative valence. Conversely, Weibo users commonly employed terms related to personal affect ('like', 'love', 'happiness') and emojis ('oh', 'heart') when expressing positive valence, whereas words indicative of negation ('no') and sorrow ('sad', 'cry') are prevalent in expressing negative valence. On the dimension of arousal, Twitter users expressed profanity ('shit', 'fuck') and interpersonal expressions ('awesome', 'amazing') for high arousal while using terms indicating low activities ('sleep', 'bed') and time-oriented description ('today', 'week', 'day', 'time') for low arousal. In contrast, Weibo users predominantly utilized positive emojis('steal-laugh', 'applaud') to convey high arousal, while employing affirmation ('yes'), negation ('no'), and sharing aspects of daily life ('home', 'sleep') to express low arousal.

Comparing the two DLA results for topics, Twitter users had relaxing weekend ('weekend', 'awesome', 'amazing', 'great', 'retreat'), celebration of events ('birthday', 'wishes', 'happy', 'present', 'wished'), luck and achievement ('win', 'won', 'contest', 'prize', 'lottery') for positive valence high arousal. Conversely, Weibo users discussed affectionate bonding ('love', 'hopeless', 'willing', 'protective', 'friendly', where hopeless means love in deep) to express their feelings, particularly in the context of festivals and celebrations ('new year', 'red envelope') and interests in celebrities and TV shows ('celebraties','singer'). For positive valence low arousal, Twitter users usually talked about relaxing routines ('day', 'today', 'good', 'chilled') and sleep ('night', 'sleep', 'tonight', 'rest', 'hoping'). Besides, Weibo users shared family reunion ('home', 'return', 'mother', 'family', 'new year') and savory cuisines ('dish', 'meat', 'delicious', 'soup', 'dish'). When ex-

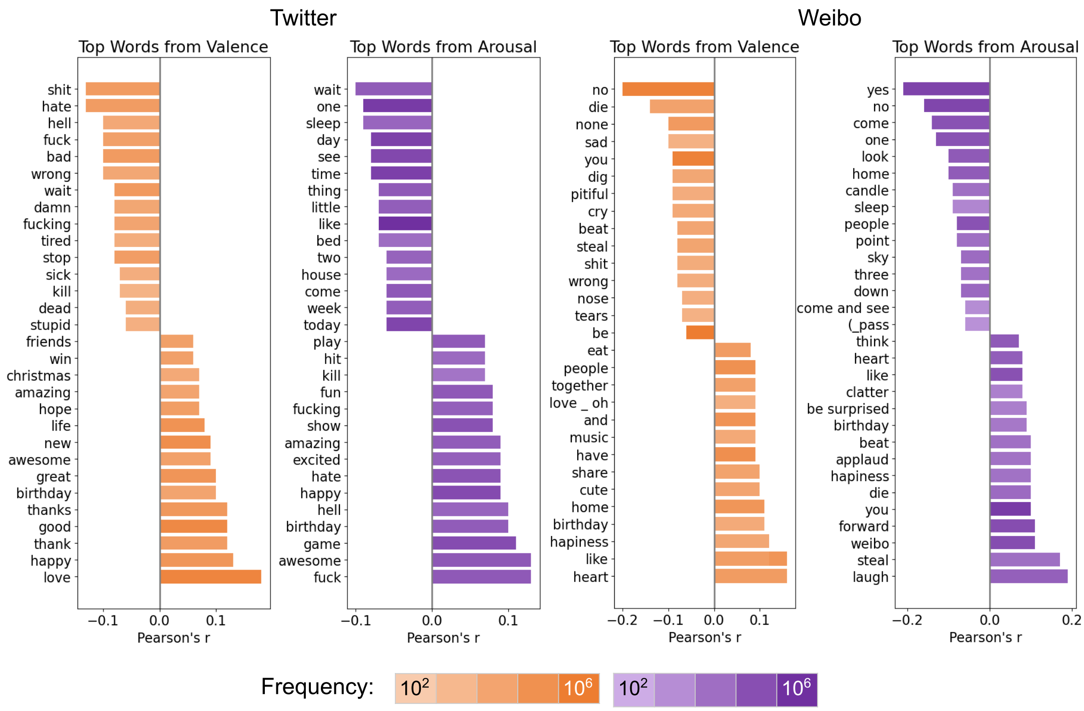

Figure 3: Words and phrases associated with valence and arousal on Twitter and Weibo (translated) from the top 15 phrases for effect strength (Pearson r), colored by frequency. Statistically significant ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected).

pressing strong negative feelings, Twitter users mainly used profanities ('fucking', 'fuck', 'shit', 'pissed', 'bullshit') to convey intense emotions, while Weibo users discussed law enforcement and criminal investigation ('police', 'crime', 'suspect', 'caught', 'case'). Additionally, Weibo discussions on negative high arousal included the use of emojis ('sweat') and negative emotions ('schocking', 'hurt', 'give up'). Concerning negative valence low arousal, Twitter users usually showed personal negative feelings like tiredness ('tired', 'sleepy', 'sleep', 'sooo', 'ugh'), engaged in discussions about daily activities ('hair', 'cut', 'short', 'haircut', 'cutting') and mentioned words related to time ('hour', 'minute'). Similarly, Weibo users also mentioned sleep ('sleep', 'awake', 'bed').

4 Discussion

This paper examined the functional relationship between valence and arousal based on large-scale social media texts across the United States and China. Our findings suggest that public affective expressions replicate the asymmetrical affective V-shaped relationship but with a negativity bias (negative feelings increase more strongly than positive feelings with increasing arousal) and negativity offset (feelings of arousal are higher at low negative valence levels than positive valence). In addition, the arousal and valence slope was steeper for Twitter users than for Weibo users.

One of the major findings in our study is that the American participants had stronger negativity bias and overall had higher arousal with higher positive and negative valence compared to Chinese participants. This is consistent with past findings on West-East distinction in emotional arousal: in Western or individualist culture, high-arousal emotions are valued and promoted more than low-arousal emotions, while in Eastern or collectivist culture, low-arousal emotions are valued more than high-arousal emotions (Lim, 2016). Even in traditional Asian medicine, there is an assumption that excessive emotional expression can be harmful and cause diseases, whether it is positive or negative emotions (Lim et al., 2008). Our findings confirmed that Chinese users on Weibo express lower arousal levels for both negative and positive emotions.

Content analyses of the findings suggested that Chinese participants displayed less high arousal positive affect emotional behavior than their American counterparts. This is consistent with past findings that there seems to be a general preference in the West for high-arousal positive states like

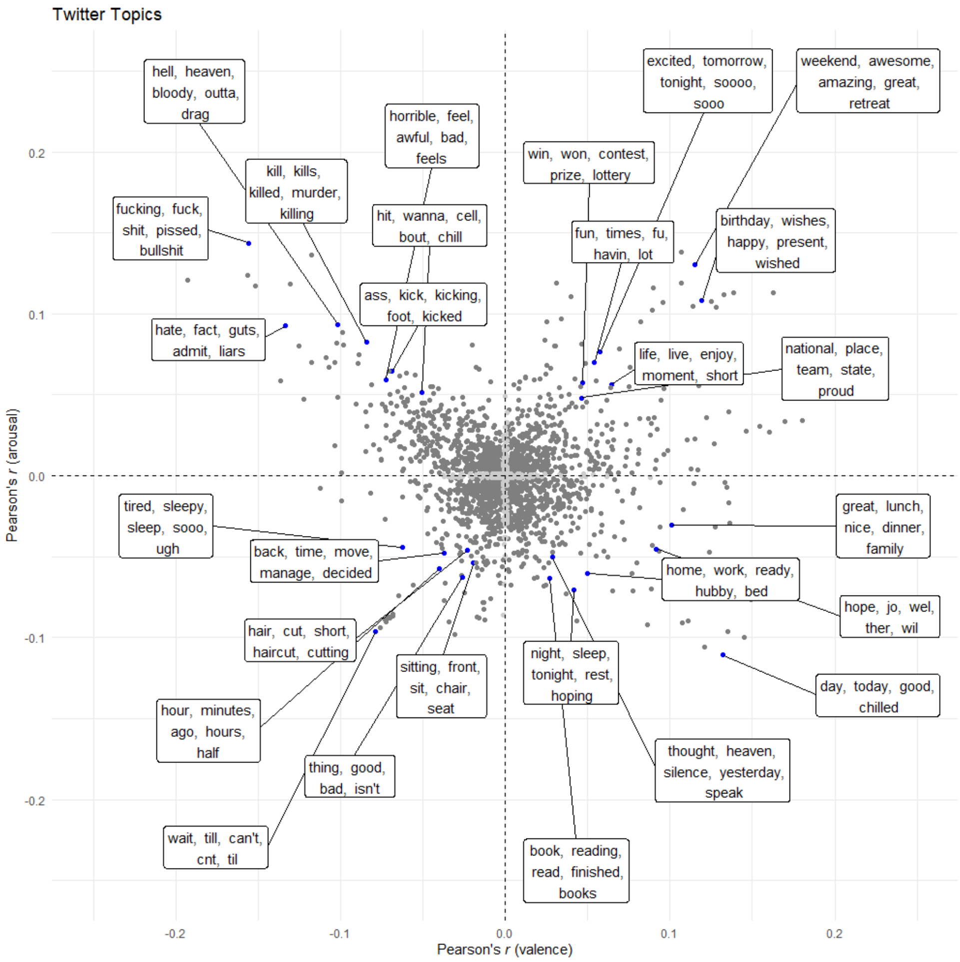

Figure 4: Topics associated with valence and arousal on Twitter, sorted by effect size (Pearson r). Each point is a topic, and statistically significant topics ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected) are shown in dark gray. The X-axis is the Pearson r with valence and the Y-axis with arousal. The top 5 words in each topic are shown.

excitement or enthusiasm (Sommers, 1984). At the same time, people in the East generally prefer low-arousal positive affective states like calm or peacefulness (Tsai, 2007). Moreover, we saw Twitter users using more explicit excitement-focused terms such as awesomeness, while Weibo users tended to express positive emotions more implicitly, e.g., emojis. This is consistent with findings that the communication style of East Asian language communities tends to be more indirect than that of their Western counterparts (Fong, 1998; Gudykunst et al., 1988; Neuliep, 2012).

Similarly, past literature suggests that high-arousal emotions serve as an effective means of influencing others in the West (Tsai, 2007), while low-arousal emotions serve as an effective means of adjusting and conforming to others in the East (Markus and Kitayama, 1991). We found that low arousal emotions in Weibo were used to influence others through sharing wisdom and conclusions about life and love. On the high arousal-high positive affect sphere, Twitter users celebrated more personal events, while Weibo users talked more excitedly about celebrities and current events. Therefore, it is likely that while the ideal affect preference translates into affective expressions about personal experiences in the East, discussion of media culture is exempt from such norms: for instance, while it may be frowned upon to act too excited about personal events, the same restrictions are not in place when expressing excitement about celebrities and cultural events. As such, our findings provide a novel insight into our understanding of norm differences in affective expression in East vs West.

Similarly, looking at the difference in negativity bias for Twitter and Weibo, while Twitter users use profanity primarily, Weibo users tend to use words with much lower intensity, confirming the as-

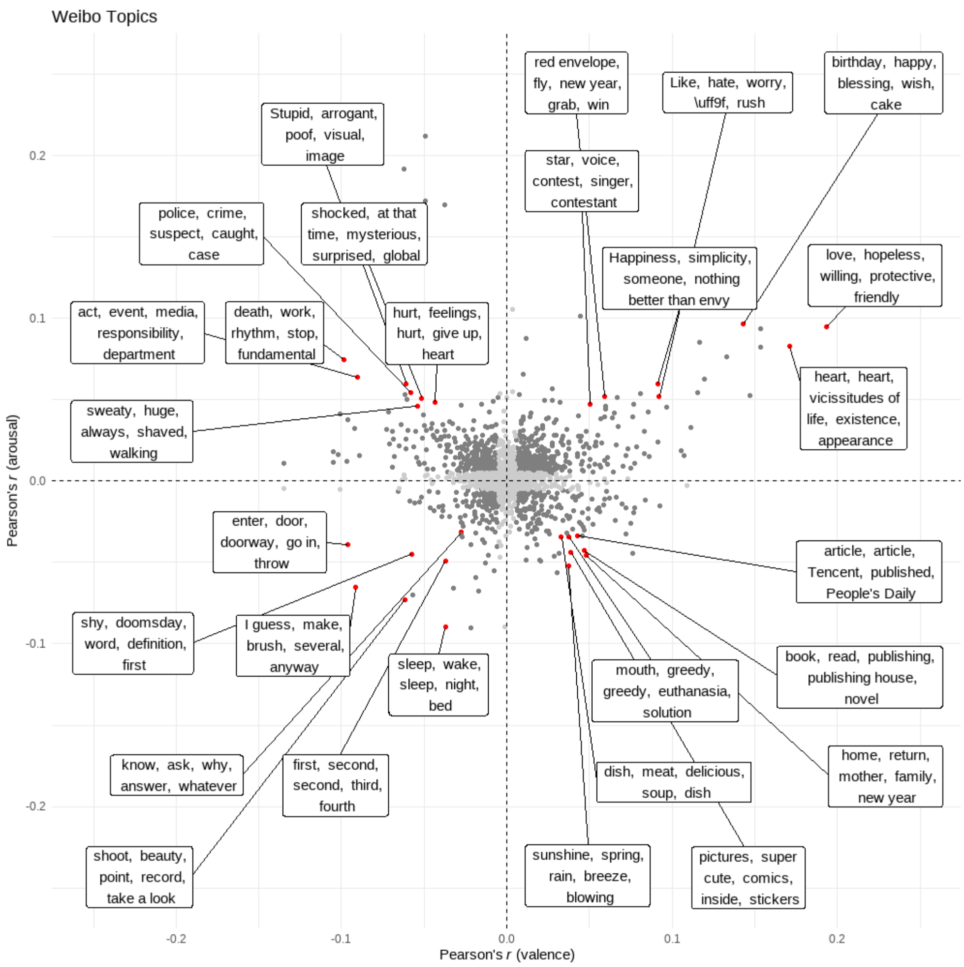

Figure 5: Topics associated with valence and arousal on Weibo, sorted by effect size (Pearson r). Each point is a topic and statistically significant topics ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected) are shown in dark gray. The X-axis is the Pearson r with valence and the Y-axis with arousal. English translations of the top 5 words in each topic are shown.

sumption that Chinese users try to avoid expressing extreme emotions.

One surprising finding in our study was that we did not find a positivity offset. We instead found a negativity offset for both American and Chinese participants. The theoretical explanation for the positivity offset (and negativity bias) comes from the Evaluative Space Model (ESM; Cacioppo et al., 1999; Norris et al., 2010), which proposed that positive and negative affect have different arousal functions and predicts greater positive than negative affect at low levels of affective input. The adaptive reason for the offset was hypothesized to encourage approaching novel stimuli in low-threat conditions. However, our finding suggests this may not translate to public affective behavior, particularly on social media. It suggests that people on both Twitter and Weibo are more likely to approach neutral stimuli in negative terms while simultaneously having stronger negative reactions to higher arousal events. Therefore, our studies elucidate how certain theories of affect may not explain affective behavior universally, partly because of the contexts not considered in said theories.

This study highlights the importance of studying public emotional behavior and how it is distinguished from self-reported findings. Our findings could confirm some theoretical assumptions in traditional self-report research by adding new empirical evidence when applied to public emotional behavior. Future research looking at individual self-reports and public behavior can help us understand what these differences can represent at the individual level.

5 Limitations

Even though Twitter and Weibo are comparable in usage (Li et al., 2020; Guntuku et al., 2019c) and have not been shown to have significant differences in predicting individual states (Gao et al., 2012), data from other platforms such as WeChat and RenRen in China and Facebook in the US have not been included in this study due to access constraints. Emojis are a significant contributor to affective expressions (Li et al., 2019); however, we did not include them in this study due to differences in encodings while collecting the data making it infeasible for us to parse them accurately. Further, social media users are non-representative of the general population, and the participants in this study are non-random and convenient samples.

6 Ethics

This study, focusing on the cultural differences in affective expressions between Twitter users in the United States and Sina Weibo users in China, raises several ethical considerations:

1. Data Privacy and Anonymity: The research analyzes social media posts from Twitter and Sina Weibo. It is important to ensure that individual users' privacy is respected. All data extracted from these platforms is anonymized by removing personally identifiable information.

2. Cultural Sensitivity and Bias: Given the cross-cultural nature of the study, it is critical to approach the analysis with cultural sensitivity. Researchers must be aware of and mitigate any biases arising from their cultural backgrounds or perspectives. This includes being mindful of how cultural contexts influence affective expressions and the interpretation thereof.

3. Representation and Generalization: Care should be taken to avoid over-generalizing the findings. The study's results are based on specific social media platforms and may not represent the broader United States and China populations.

References

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E P Seligman, and Lyle H Ungar. 2013. Personality, gender, and age in the language of social media: The Open-Vocabulary approach. *PLoS One*, 8(9):e73791.

Roy F Baumeister, Kathleen D Vohs, and David C Funder. 2007. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on psychological science*, 2(4):396–403.

Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Hamparsum Bozdogan. 1987. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370.

M M Bradley, M K Greenwald, M C Petry, and P J Lang. 1992. Remembering pictures: pleasure and arousal in memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, 18(2):379–390.

John T Cacioppo and Wendi L Gardner. 1999. EMOTION.

John T Cacioppo, Wendi L Gardner, and Gary G Berntson. 1999. The affect system has parallel and integrative processing components: Form follows function. *Journal of personality and Social Psychology*, 76(5):839.

Kenneth Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Comput. Linguist.*, 16(1):22–29.

Sheldon Cohen and Sarah D Pressman. 2006. Positive affect and health. *Curr. Dir. Psychol. Sci.*, 15(3):122–125.

Johannes C Eichstaedt, Margaret L Kern, David B Yaden, H A Schwartz, Salvatore Giorgi, Gregory Park, Courtney A Hagan, Victoria A Tobolsky, Laura K Smith, Anneke Buffone, Jonathan Iwry, Martin E P Seligman, and Lyle H Ungar. 2021. Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations.

Joshua A Fishman. 1980. Bilingualism and biculturism as individual and as societal phenomena.

Mary Fong. 1998. Chinese immigrants' perceptions of semantic dimensions of direct/indirect communication in intercultural compliment interactions with north americans. *Howard journal of Communication*, 9(3):245–262.

Nico H Frijda. 1986. *The Emotions*. Maison des Sciences de l'Homme.

Qi Gao, Fabian Abel, Geert-Jan Houben, and Yong Yu. 2012. A comparative study of users' microblogging behavior on sina weibo and twitter. In *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20, 2012. Proceedings 20*, pages 88–101. Springer.

Gerald Gorn, Michel Tuan Pham, and Leo Yatming Sin. 2001. When arousal influences ad evaluation and valence does not (and vice versa). *J. Consum. Psychol.*, 11(1):43–55.

William B Gudykunst, Stella Ting-Toomey, and Elizabeth Chua. 1988. *Culture and interpersonal communication.* Sage Publications, Inc.

S C Guntuku, M Li, L Tay, and L H Ungar. 2019a. Studying cultural differences in emoji usage across the east and the west. *Proceedings of the International*.

Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C Eichstaedt, and Lyle H Ungar. 2019b. Understanding and measuring psychological stress using social media. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 214–225.

Sharath Chandra Guntuku, Mingyang Li, Louis Tay, and Lyle H Ungar. 2019c. Studying cultural differences in emoji usage across the east and the west. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 226–235.

Sharath Chandra Guntuku, Weisi Lin, Michael James Scott, and Gheorghita Ghinea. 2015a. 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.

Sharath Chandra Guntuku, Daniel Preotiuc-Pietro, Johannes C Eichstaedt, and Lyle H Ungar. 2019d. What twitter profile and posted images reveal about depression and anxiety. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 236–246.

Sharath Chandra Guntuku, Michael James Scott, Huan Yang, Gheorghita Ghinea, and Weisi Lin. 2015b. The CP-QAE-I: A video dataset for exploring the effect of personality and culture on perceived quality and affect in multimedia.

Kathleen L Hourihan, Scott H Fraundorf, and Aaron S Benjamin. 2017. The influences of valence and arousal on judgments of learning and on recall. *Mem. Cognit.*, 45(1):121–136.

Kokil Jaidka, Salvatore Giorgi, H Andrew Schwartz, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2020. Estimating geographic subjective well-being from twitter: A comparison of dictionary and data-driven language methods. *Proc. Natl. Acad. Sci. U. S. A.*, 117(19):10165–10171.

Kokil Jaidka, Sharath Guntuku, and Lyle Ungar. 2018. Facebook versus twitter: Differences in Self-Disclosure and trait prediction. *ICWSM*, 12(1).

Margaret L Kern, Lea Waters, Alejandro Adler, and Mathew White. 2014. Assessing employee wellbeing in schools using a multifaceted approach: Associations with physical health, life satisfaction, and professional thriving. *Psychology*, 05(06):500–513.

Peter Kuppens, Francis Tuerlinckx, James A Russell, and Lisa Feldman Barrett. 2013. The relation between valence and arousal in subjective experience. *Psychol. Bull.*, 139(4):917–940.

Peter Kuppens, Francis Tuerlinckx, Michelle Yik, Peter Koval, Joachim Coosemans, Kevin J Zeng, and James A Russell. 2017a. The relation between valence and arousal in subjective experience varies with personality and culture. *Journal of personality*, 85(4):530–542.

Peter Kuppens, Francis Tuerlinckx, Michelle Yik, Peter Koval, Joachim Coosemans, Kevin J Zeng, and James A Russell. 2017b. The relation between valence and arousal in subjective experience varies with personality and culture. *J. Pers.*, 85(4):530–542.

P J Lang. 1994. The varieties of emotional experience: a meditation on James-Lange theory. *Psychol. Rev.*, 101(2):211–221.

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):e84997.

Mingyang Li, Sharath Guntuku, Vinit Jakhetiya, and Lyle Ungar. 2019. Exploring (dis-) similarities in emoji-emotion association on twitter and weibo. In *Companion proceedings of the 2019 world wide web conference*, pages 461–467.

Mingyang Li, Louis Hickman, Louis Tay, Lyle Ungar, and Sharath Chandra Guntuku. 2020. Studying politeness across cultures using english twitter and mandarin weibo. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–15.

Zhongguo Li and Maosong Sun. 2009. Punctuation as implicit annotations for chinese word segmentation. *Comput. Linguist.*, 35(4):505–512.

Nangyeon Lim. 2016. Cultural differences in emotion: differences in emotional arousal level between the east and the west. *Integrative medicine research*, 5(2):105–109.

Youn-kyung Lim, Justin Donaldson, Heekyoung Jung, Breanne Kunz, David Royer, Shruti Ramalingam, Sindhia Thirumaran, and Erik Stolterman. 2008. Emotional experience and interaction design. *Affect and emotion in human-computer interaction: From theory to applications*, pages 116–129.

H Lin, Jia Jia, Liqiang Nie, Guangyao Shen, and Tat-Seng Chua. 2016a. What does social media say about your stress? *undefined*.

Xialing Lin, Kenneth A Lachlan, and Patric R Spence. 2016b. Exploring extreme events on social media: A comparison of user reposting/retweeting behaviors on twitter and weibo.

Marco Lui and Timothy Baldwin. 2012. langid.py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 System Demonstrations*, pages 25–30.

Lin Ma. 2013. Electronic word-of-mouth on microblogs: A cross-cultural content analysis of twitter and weibo. *Intercultural Communication Studies*, 22(3).

Hazel R Markus and Shinobu Kitayama. 1991. Cultural variation in the self-concept. In *The self: Interdisciplinary approaches*, pages 18–48. Springer.

David Matsumoto. 1990. Cultural similarities and differences in display rules. *Motivation and emotion*, 14:195–214.

Raina M Merchant, David A Asch, Patrick Crutchley, Lyle H Ungar, Sharath C Guntuku, Johannes C Eichstaedt, Shawndra Hill, Kevin Padrez, Robert J Smith, and H Andrew Schwartz. 2019. Evaluating the predictability of medical conditions from social media posts. *PLoS One*, 14(6):e0215476.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Stroudsburg, PA, USA. Association for Computational Linguistics.

James W Neuliep. 2012. The relationship among intercultural communication apprehension, ethnocentrism, uncertainty reduction, and communication satisfaction during initial intercultural interaction: An extension of anxiety and uncertainty management (aum) theory. *Journal of Intercultural Communication Research*, 41(1):1–16.

Catherine J Norris, Jackie Gollan, Gary G Berntson, and John T Cacioppo. 2010. The current status of research on the structure of evaluative space. *Biological psychology*, 84(3):422–436.

Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. *The Cognitive Structure of Emotions*. Cambridge University Press.

Daniel Preoţiuc-Pietro, H Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. Modelling valence and arousal in facebook posts.

Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our twitter profiles, our selves: Predicting personality with twitter.

James A Russell. 1980. A circumplex model of affect.

Bahar Salehi, Dirk Hovy, Eduard Hovy, and Anders Søgaard. 2017. Huntsville, hospitals, and hockey teams: Names can reveal your location.

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.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E P Seligman, and Others. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One*, 8(9):e73791.

Gideon Schwarz. 1978. Estimating the dimension of a model. *aos*, 6(2):461–464.

Shula Sommers. 1984. Reported emotions and conventions of emotionality among college students. *Journal of Personality and Social Psychology*, 46(1):207.

Michael A Tarrant, Michael J Manfredo, Peter B Bayley, and Richard Hess. 1993. Effects of recall bias and nonresponse bias on Self-Report estimates of angling participation. *N. Am. J. Fish. Manage.*, 13(2):217–222.

Xianyun Tian, Philip Batterham, Shuang Song, Xiaoxu Yao, and Guang Yu. 2018. Characterizing depression issues on sina weibo. *Int. J. Environ. Res. Public Health*, 15(4).

Jeanne L Tsai. 2007. Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, 2(3):242–259.

Jeanne L Tsai, Brian Knutson, and Helene H Fung. 2006. Cultural variation in affect valuation. *J. Pers. Soc. Psychol.*, 90(2):288–307.

Eugene Winograd and Ulric Neisser. 2006. *Affect and Accuracy in Recall: Studies of 'Flashbulb' Memories*. Cambridge University Press.

Yuanyuan Xu, Yongju Yu, Yuanjun Xie, Li Peng, Botao Liu, Junrun Xie, Chen Bian, and Min Li. 2015. Positive affect promotes well-being and alleviates depression: The mediating effect of attentional bias. *Psychiatry Res.*, 228(3):482–487.

Michelle S M Yik, James A Russell, and Lisa Feldman Barrett. 1999. Structure of self-reported current affect: Integration and beyond. *J. Pers. Soc. Psychol.*, 77(3):600–619.

Wanru Zhang, Andrew Caines, Dimitrios Alikaniotis, and Paula Buttery. 2016. Predicting author age from weibo microblog posts. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2990–2997.

Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, and Xing Xie. 2015. You are where you go.

Yi Zhu, Sharath Chandra Guntuku, Weisi Lin, Gheorghita Ghinea, and Judith A Redi. 2018. Measuring individual video QoE.

Appendix

All the figures and tables from Weibo are translated into English with Google Translate. In the Appendix, we show the figures with original Chinese text. Figure 6 shows the top 15 phrases for effect strength with valence and arousal. Figure 7 shows the top topics associated with valence and arousal.

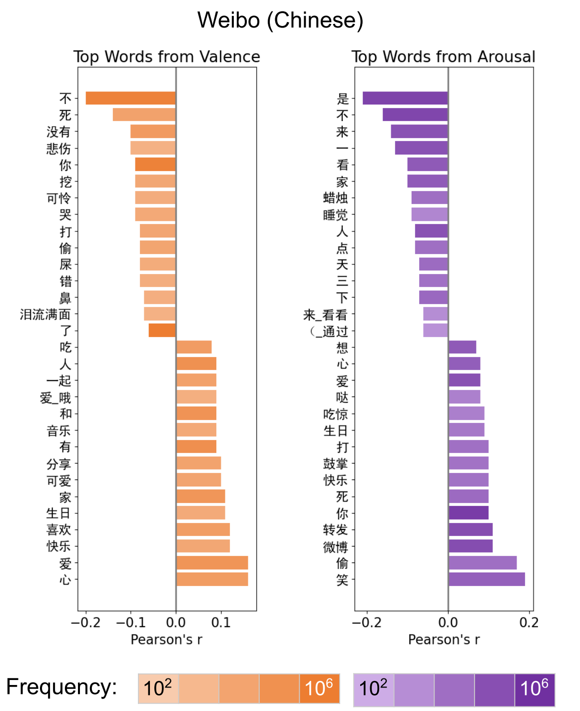

Figure 6: Words and phrases associated with valence and arousal on Weibo (Chinese) from the top 15 phrases for effect strength (Pearson r), colored by frequency. Statistically significant ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected).

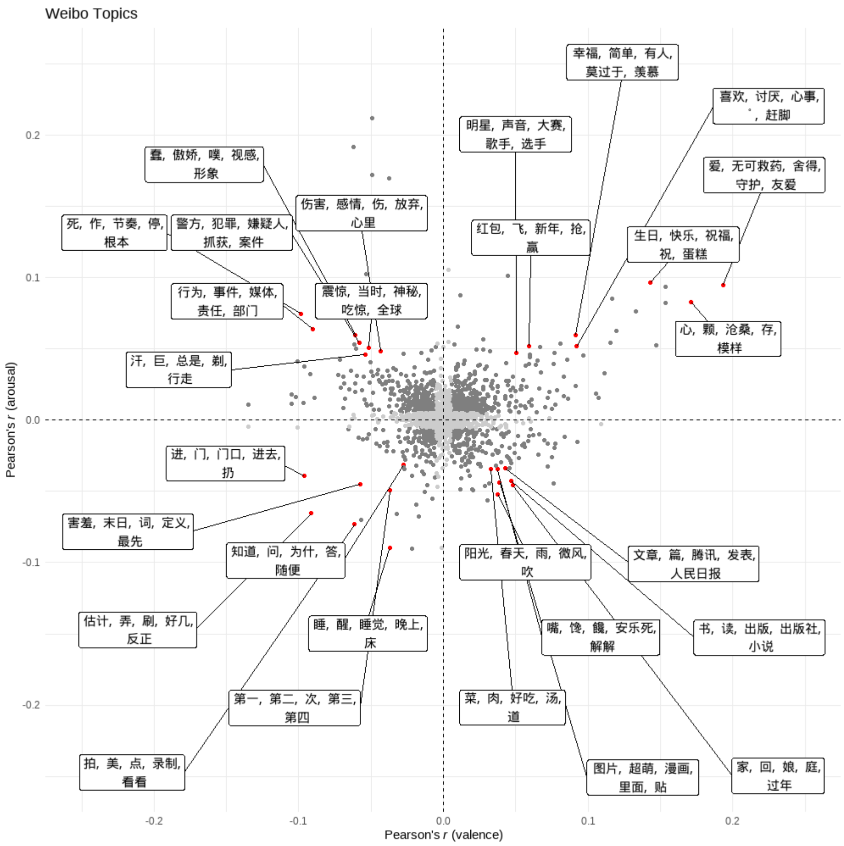

Figure 7: Topics associated with valence and arousal on Weibo (Chinese), sorted by effect size (Pearson r). Each point is a topic and statistically significant topics ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected) are shown in dark gray. The X-axis is the Pearson r with valence and the Y-axis with arousal. English translations of the top 5 words in each topic are shown.