Social Media Reveals Urban-Rural Differences in Stress across China

Jesse Cui,[1] **Tingdan Zhang,**[2] **Kokil Jaidka,**[3] **Dandan Pang,**[4] **Garrick Sherman,**[1] **Vinit Jakhetiya,**[5] **Lyle Ungar,**[1] **Sharath Chandra Guntuku**[1]

[1]University of Pennsylvania, [2]Beijing Normal University, [3]National University of Singapore, [4]University of Bern
[5]Indian Institute of Technology, Jammu
{jessecui, sharathg}@cis.upenn.edu

Abstract

Modeling differential stress expressions in urban and rural regions in China can provide a better understanding of the effects of urbanization on psychological well-being in a country that has rapidly grown economically in the last two decades. This paper studies linguistic differences in the experiences and expressions of stress in urban-rural China from Weibo posts from over 65,000 users across 329 counties using hierarchical mixed-effects models. We analyzed phrases, topical themes, and psycho-linguistic word choices in Weibo posts mentioning stress to better understand appraisal differences surrounding psychological stress in urban and rural communities in China; we then compared them with large-scale polls from Gallup. After controlling for socioeconomic and gender differences, we found that rural communities tend to express stress in emotional and personal themes such as relationships, health, and opportunity while users in urban areas express stress using relative, temporal, and external themes such as work, politics, and economics. These differences exist beyond controlling for GDP and urbanization, indicating a fundamentally different lifestyle between rural and urban residents in very specific environments, arguably having different sources of stress. We found corroborative trends in physical, financial, and social wellness with urbanization in Gallup polls.

Introduction

Digital trace data is increasingly used to better understand population-level health trends worldwide. An emerging body of research focuses on mining the language of self-expression on social media to understand the well-being and quality of life of the people living in cities (Quercia, Seaghdha, and Crowcroft 2012), regions (Rentfrow 2010), and countries (De Choudhury et al. 2017). In general this body of work has demonstrated that language expressions on social media can aid measurement of regional differences in social and cultural aspects of life.

Social media analyses on *stress*, the mental or emotional strain arising from difficult circumstances, have illustrated the predictive efficacy of supervised language models trained on *English* social media posts and are shown to be a valid and robust predictor of subjective mental health and well-being in communities (Bagroy, Kumaraguru, and
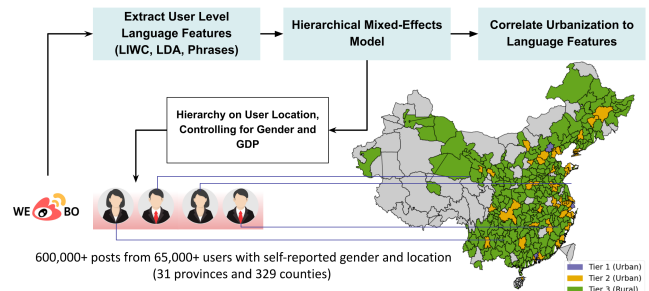
Figure 1: Hierarchical Model to Study Urban-Rural Differences in Stress across China

De Choudhury 2017; Guntuku et al. 2019a; Jaidka et al. 2020b).

However, there are two major gaps in this body of work. First, in focusing only on social media platforms and posts from an English-speaking audience, social media activity of non-English speakers and writers is less explored. With 1.3 billion speakers or over three times as many native speakers as English, Mandarin Chinese is one of the most spoken language worldwide. A relatively insular social media climate and a notable under-representation in the computational social science sphere makes China one of the, if not the most, important cultural context for understanding the relationship between social media language, mental health, and well-being in rapidly economically growing countries (Jackson and Wang 2013). While there are studies on language from a small number of individuals in China analyzing depression (Tian et al. 2018), suicidal ideation (Wang et al. 2018), and anxiety (Tian et al. 2017), psychological stress across regions informing public health is understudied. Second, little is understood about how the differences in social media usage relate to urban-rural experiences and socioeconomic differences.

Based on a body of literature on digital inequality (DiMaggio and Hargittai 2001; Ahmed et al. 2020) and the reinforcement hypothesis (Robinson et al. 2015), we can anticipate differences in how urban and rural residents *use* social media. Therefore, digital inequality could be a factor driving regional variations in the language of social media posts. Analyses that recognize the heterogeneity in social

media posts allow researchers to be more cognizant of demographic and social differences (Gilbert, Karahalios, and Sandvig 2010) in large scale analyses. The insights from this study can enable public health professionals to better understand nuanced differences in psychological stressors, especially in an understudied context.

Specifically, we address the following research questions in this study: 1) How do linguistic markers of psychological stressors on Weibo differ between urban and rural Chinese communities; and 2) How do large-scale polls on comorbidities of stress compare with linguistic markers on Weibo? We answered these questions by: 1) modeling the contribution of regional characteristics on individuals' language using hierarchical linear modeling (Figure 1); 2) analyzing social media users' discussion of psychological stressors in rural and urban areas in China using a) words and phrases ($n-$grams), b) topics generated from latent Dirichlet allocation (LDA), and c) psycholinguistic categories from the Chinese Linguistic Inquiry Word Count (LIWC) dictionary (Zhao et al. 2016); and 3) providing insights about the language markers of stressors on Weibo that are indicative of other comorbidities, using large scale survey data from Gallup Polls. We also discuss this study, by comparing and contrasting the findings, in the light of related work in the United States (Jaidka et al. 2020b).

Background

Stress and Urbanization

There are several potential reasons for urban and rural areas to report different well-being. For example, subjective perception of economic or class differences in urban areas lead to higher stress due to higher social comparisons (Knight and Song 1999). Rural communities are also known to be different than urban communities when it comes to how closely people feel connected with their family, friends, and neighbors (Beaudoin and Thorson 2004). Since social relationships help to buffer stress and anxiety (Helliwell and Putnam 2004), a lack in that domain could force an individual to cope with their stress in other ways, or experience lower well-being and higher stress as a trade-off.

Urban residents are more likely to face the environmental stressors associated with living in densely populated areas or in compact housing conditions (Evans and English 2002). Major concerns to rural areas include access to healthy food and quality medical care (Walker, Keane, and Burke 2010). On the other hand, studies have reported that an urban upbringing affects the region of the brain responsible for tasks involving negative affect processing and stress evaluation as compared to tasks involving regular cognitive processing (Lederbogen et al. 2011).

Prior studies suggest that urban-rural differences affect personal well-being and social attitudes of their residents. This has been examined in many Western contexts, such as the United Kingdom (White et al. 2013) and the United States (Rentfrow 2010). Findings from the United Kingdom (White et al. 2013) show that individuals in urban communities with greener surroundings have lower mental distress and higher well-being, even after controlling for indi-

vidual and regional covariates. Similar findings have been reported in the United States, where neuroticism is higher in the states on the east coast than on the west (Rentfrow 2010). Our study investigates whether, and how, these findings would generalize to the cultural and sociopolitical context of China.

Urban-Rural Divide in China

Since the 'Reform and Opening-up' in 1978, China's total economic output has made huge strides. However, due to urban-biased policies, the rapid growth of the Chinese economy also led to urban-rural inequality (Zhu, Yu, and He 2020). These are reflected in almost all aspects of daily life, such as the systems of household, education, and welfare (Partridge and Rickman 2008; Ward and Brown 2009). The unique 'hukou' system (i.e., house registration system) is a prominent factor for the large urban-rural divide in China as it restricts the ability of workers to move from poor rural areas to more productive urban regions (Liu 2005).

Compared to their rural counterparts, urban people appear to have a high level of compliance with China's one-child policy (Synder 2000). This leads to significantly higher ratio of boys to girls in rural areas in China. It is thus harder for rural males to find a wife. They often need to strictly follow the so-called 'Caili' culture (after engagement, a man or his family has to give the bride's family a betrothal gift to get married), resulting in financial strain (Jiang, Zhang, and Sánchez-Barricarte 2015). Therefore, we hypothesized stressors related to marriage and relationships would show up in the language of rural communities.

Weibo and Users' Traits

Sina Weibo is a widely used platform in China, similar to Twitter. On Weibo, users can communicate, express themselves, and project their identity (Bucher and Helmond 2017). The language of Weibo posts thus acts as an extensive resource for deriving insights into its users' health and well-being, which is often resource intensive to obtain with traditional surveys of nationwide samples or in areas which are politically or geographically inaccessible.

While social media users are not representative of national demographics, several psychological traits can be inferred from Weibo posts, including users' age and gender (Zhang et al. 2016), personality (Li et al. 2014), as well as individuals' mental health (Tian et al. 2018). Weibo has also been used to study cultural differences across China and other countries (Li et al. 2019; Guntuku et al. 2019b; Li et al. 2020). Weibo has more than 550 million registered users as of 2021, 50% of whom use it daily and more than 64,000 messages are posted every minute (Thomala 2021). In this study, we used Sina Weibo posts across China to measure the linguistic differences in psychological stressors across urban and rural communities.

Variation of Weibo Stress Language with Urbanization

Data

Weibo posts were obtained using a breadth-first search strategy on Weibo users from prior work (Guntuku et al. 2019b). A total of $296,932,162$ posts from $888,435$ users were collected between 2012-2019. User's self-identified profile locations were used to identify their county and province. We filtered posts by removing duplicate messages per user and only included Mandarin posts detected via *langid* (Lui and Baldwin 2012). We also removed users with fewer than 5 posts and only included counties with more than 100 posts. The selection of the stress keywords was determined independently by two native Chinese speakers. A joint list was generated, and differences were arbitrated. Considering the language used on social media is not standardized, relying on dictionaries was insufficient. For example, many people use "鸭梨" (which, literally, means white pear) to replace "压力" (which, literally, means stress) on social media as their pronunciations are similar and the former is netspeak. Therefore, we included words from both the thesaurus and expert curation based on different mentions of stressors on Sina Weibo. This resulted in 30 keywords. A list of the final keywords is provided in the Appendix. We sampled a random set of 100 posts with the keywords and found more than 90% of the posts to contain mentions of stressors. This study was considered exempt by University of Pennsylvania Institutional Review Board.

Descriptives: Using 30 stress-related words and phrases, we retrieved 641,262 posts from 67,027 users in 329 counties (also known as 县级市, county cities) that are spread across 22 provinces, 5 autonomous regions, and 4 municipalities. This data provided coverage for all provinces and 91% of counties in Mainland China. Our dataset consisted of 160,064 rural messages and 481,198 urban messages from 15,916 rural users and 51,111 urban users, respectively. 29,305 users were male, 37,722 users were female. We used the three-point system for labeling counties into urban (tiers 1 and 2) and rural (tier 3) categories (Chung and Lam 2004; Long 2016). Tier 1 counties include super-cities of China, namely Shanghai, Beijing, Guangzhou, and Shenzhen. Tier 2 counties are urbanized cities, and tier 3 counties are rural areas. We also replicated the language analyses for Weibo users at the province level, where we had access to non-binary urbanization outcome (percentage of individuals living in urban areas), and observed that the urban-rural stressor differences are consistent for both LIWC and latent Dirichlet allocation (LDA) language categories. Mandarin text along with translations for county and province level analyses are presented in Supplementary Material 1 and 2 respectively.

Post-stratification: We replicated the language analyses for Weibo users at the county level with post-stratified samples on gender, using province gender ratios from the 2015 China statistical yearbook (Statistics 2015), so that our sampled data from Weibo can match the offline population distribution (Jaidka et al. 2020a), and observed that the urban-rural stressor differences are consistent for both LIWC and latent Dirichlet allocation (LDA) language categories (results in Supplementary Material 3).

Language Features

Weibo posts were segmented with the *jieba* package in Python 3.4. We masked specific features such as user mentions and URLs into meta tags. We aggregated posts at the user level and extracted three different features: a) words and phrases (1-3grams), b) topics generated from latent Dirichlet allocation (LDA) (Blei, Ng, and Jordan 2003), and c) psycho-linguistic lexicon - Linguistic Inquiry Word Count (LIWC) (Pennebaker et al. 2015). We used a validated simplified Chinese version of LIWC (Zhao et al. 2016). These features have been used in several social media studies on health and well-being (Saha et al. 2019; Guntuku et al. 2020, 2021).

Words and Phrases: We used a bag of words and multi-word expressions ('phrases'), with counts normalized per user, as a linguistic dimension. These phrases ranged from one word to three consecutive words. Words and phrases that were not present in at least 10% of the users' posts were excluded from analyses to remove outliers. Further, we use a point-wise mutual information (PMI) (Bouma 2009), keeping phrases with a threshold above 3 to consider only phrases that occur with high probability, e.g., so_much_pressure (压力_很大). Word and phrase frequencies were divided by the user's total number of words, yielding relative frequencies of each. **Topics:** We created content-specific topics limited to the thematic scope of stress expressions on Weibo using a process that consists of three steps, similar to previous works (Van Loon et al. 2020; Zamani et al. 2020): a) tokenization and calculating co-locations; b) identifying words most associated with stressors, and c) the topic modeling process. We used a random set of 100,000 posts each containing the stressors keywords, and another set of 100,000 random posts not containing stressors keywords. After removing re-posts and posts with URLs to minimize spam and news posts, we tokenized them and extracted the relative frequency of single words and phrases (1-3 grams) in all posts. These n-grams were used as input features in logistic regression to identify words and phrases that were significantly differentially present in tweets containing stressors keywords compared to the control set. The outcome variable was set to 1 for posts with stressors keywords and 0 for those without. To test significance, we used $p < 0.05$ level after applying Benjamini-Hochberg correction (Benjamini and Hochberg 1995). Next, we filtered out standard stop words, the stress keywords, and any non-significant words. We then used latent Dirichlet allocation (LDA) (Blei, Ng, and Jordan 2003), a probabilistic generative model which utilizes Bayesian inference to identify the prevalence of different latent linguistic states, to identify 100 topics, with an $alpha$ level of .05. We then obtained the relative topic distribution for each user.

LIWC: We counted the tokens in all posts of each user that match the tokens in the LIWC dictionary based on the validated Mandarin version (Zhao et al. 2016). We then summed these token counts per user and normalized by the number of words posted by each user.

n-grams (Urban, β₁)

Mandarin	English
元	yuan
教育	education
吃饭	eat
穿	wear
啊 ！	what!
么	what
辛苦	hard
经济	economic
儿子	son
解决	solve
您	you
话	words
一点	a little
地方	locality
家人	family
得到	get
都 有	both have
休息	rest
手机	cell phone
力量	power

n-grams (Urban, β₁)

Mandarin	English
都 不	neither
真是	really
压力 很大	so much pressure
由	by
看看	take a look
要求	claim
医生	doctors
北京	beijing
一年	one year
恐惧	fear
受	receive
毫无	nothing
大学	the university
样子	look like
各种	various
找	find
跑	run
目前	currently
家里	home
活动	activity

n-grams (Rural, $-\beta_1$)

Mandarin	English
就算	even if
选	select
真	true
有 机会	have the opportunity
也 要	also want
每次	each
不了	no more
看不到	can not see
获得	obtain
便	then
即使	even if
比 你	than you
眼泪	tears
承诺	committed to
上班	go to work
依然	still
至少	at least
测试	test
并	and
昨天	yesterday

n-grams (Rural, $-\beta_1$)

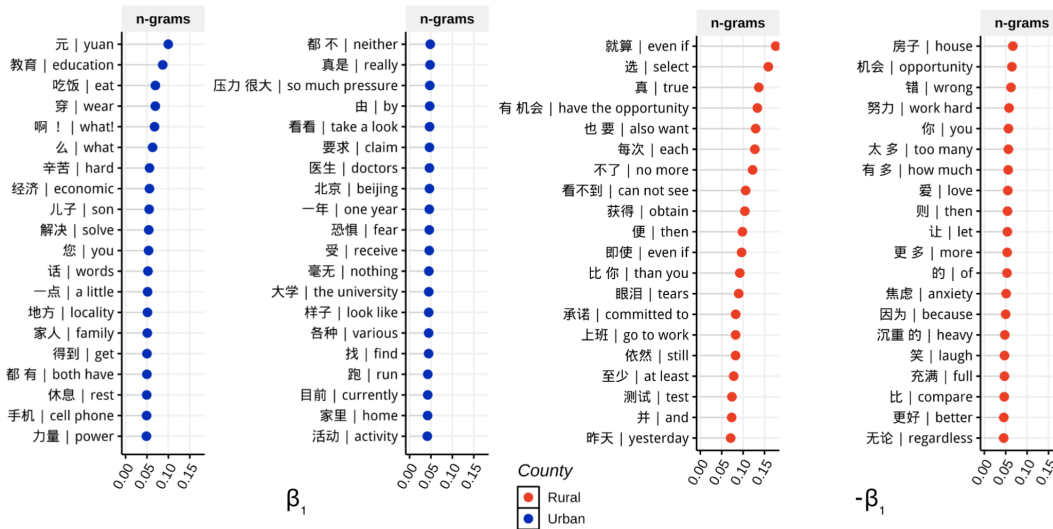| Mandarin	English
房子	house
机会	opportunity
错	wrong
努力	work hard
你	you
太 多	too many
有 多	how much
爱	love
则	then
让	let
更 多	more
的	of
焦虑	anxiety
因为	because
沉重 的	heavy
笑	laugh
充满	full
比	compare
更好	better
无论	regardless

County: Rural (red), Urban (blue)

Figure 2: Rural stress phrases (in red) and urban stress phrases (in blue) from the top 100 most frequently occurring phrases in each area, sorted by effect size (β_1). Statistically significant ($p < .05$, two-tailed t-test, Benjamini-Hochberg corrected). Mandarin phrases and corresponding English translations are shown.

Hierarchical Linear Model

An overview of the linear mixed-effects model used to analyze multi-level language data is shown in Figure 1. Level-1 of the multi-level model represents the user, the level at which we extract language features, and level-2 represents the location (county, in our case). These levels allow us to model the contribution of regional characteristics on individuals language. The use of a hierarchical mixed-effects regression modeling also allows for a robust control over location context of messages: accounting for shared variance, both within- and across-groups while not violating the independence assumption on observations (Woltman et al. 2012; Giorgi et al. 2021).

The hierarchical linear model we created treated each user-level language feature as the dependent variable and county features as the independent variable. We controlled for users' gender to account for linguistic differences with genders (Schwartz et al. 2013) and county's log Gross Domestic Product (GDP) from the 2015 China statistical yearbook (Statistics 2015) to account for its effects in relation to urbanization. The model was setup to study how being in an urban (as opposed to rural) area influences users' language attributes. We also analyzed the effect of adding a county's education level measured by the percentage of individuals who are college graduates as a control and did not find any significant difference in the results. The hierarchical model is defined in Eq. 1:

$$
\begin{aligned}
Feature_i = & \beta_1 * County_{IsUrban} \\
& + \beta_2 * County_{LogGDP} \\
& + \beta_3 * User_{Gender} + \phi + \epsilon
\end{aligned}
\tag{1}
$$

where ϕ is the random effect for one of the 329 counties and ϵ is the error term. We used Benjamini-Hochberg p-correction (Benjamini and Hochberg 1995) and report β_1

for all language features where correlations are significant at p<0.05.

For each result, we translated Mandarin words into English for wider readability, primarily using the Google Translate Cloud API. Several words were translated by dual-lingual Mandarin and English co-authors where we believed the Google API did not accurately translate the word.

Results

We present and discuss language features sorted by effect size of urban-rural differences (β_1) in this section. The list of all significant language features (with Mandarin words) - a) words and phrases, b) topics, and c) LIWC categories, along with effect sizes, 95% confidence intervals, and overall R^2, both conditional and marginal R^2, for the linear models is presented in Supplementary Material.

Words and Phrases: Figure 2 shows words and phrases associated with stress in rural (red) and urban (blue) counties. In the context of stress mentions, individuals in the rural areas predominantly used words and phrases such as 'even if', 'have the opportunity', 'no more', 'tears', 'love', 'anxiety', and 'work hard' whereas, individuals in urban areas mostly used non-personal words related to society at large, such as 'yuan', 'education', 'economic', 'power', 'the university', and 'doctors', with the exception of family terms such as 'son', 'family', and 'home'.

Topics: Figures 3 and 4 show rural and urban topics respectively sorted by their effect size. Rural topics surrounding stress centered around the major themes of emotions/experiences, discrepancy, dissatisfaction, achievement, and health/body. Specifically, rural stress topics contained words expressing dissatisfaction, fear, work ethic, insomnia/tiredness, health, discrepancy, power, and pursuit. Rural topics also indicate themes that are personal in nature, with many emotions, largely negative.

Emotions, Experiences, and Perceptions

effect: -0.054 (-0.073, -0.036) effect: -0.056 (-0.074, -0.039) effect: -0.049 (-0.068, -0.029)

Discrepancy

effect: -0.039 (-0.058, -0.019)

Dissatisfaction

effect: -0.056 (-0.076, -0.037) effect: -0.057 (-0.076, -0.037)

Achievement and Work Ethic

effect: -0.061 (-0.082, -0.039) effect: -0.059 (-0.078, -0.041) effect: -0.047 (-0.064, -0.030)

Health and Body

effect: -0.056 (-0.075, -0.036) effect: -0.045 (-0.061, -0.028) effect: -0.036 (-0.053, -0.019)

Figure 3: Topics associated with stress in rural areas. Coefficients represent the standardized effect of county urbanization to topic occurrence, significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected.

Politics, Governance, and Society

effect: 0.059 (0.043, 0.075) effect: 0.046 (0.031, 0.061) effect: 0.039 (0.023, 0.055)

Career

effect: 0.022 (0.009, 0.036)

Time and Relativity

effect: 0.061 (0.044, 0.079) effect: 0.022 (0.009, 0.034)

Finance and Economics

effect: 0.032 (0.018, 0.046) effect: 0.027 (0.014, 0.040) effect: 0.026 (0.013, 0.040)

Family and Support

effect: 0.022 (0.007, 0.036) effect: 0.018 (0.003, 0.032) effect: 0.015 (0.002, 0.028)
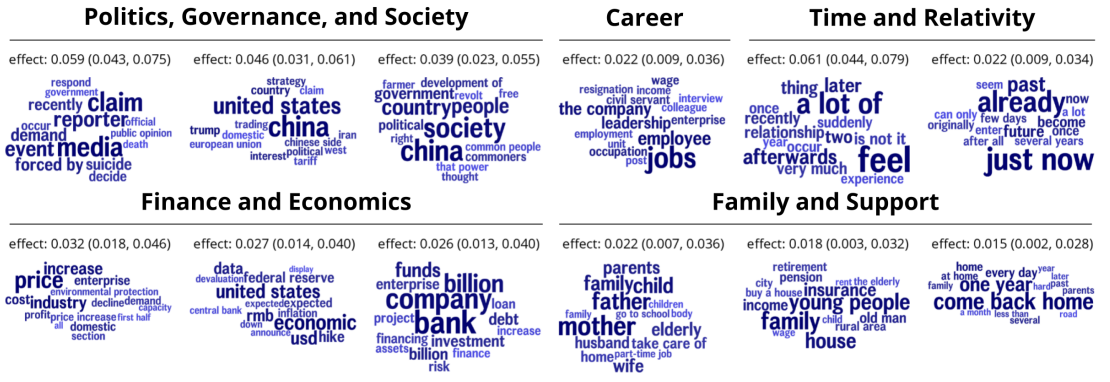
Figure 4: Topics associated with stress in urban areas. Coefficients represent the standardized effect of county urbanization to topic occurrence, significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected.

Urban topics surrounding stress centered around the major themes of economics/finance, politics/society, work/careers, family/support, and time-based terms. Urban topics are predominantly around financial, economic, and market related stressors. Politics, governance, reform, career, time-based, and family support topics also surround stress in urban areas. We observed that urban stress, in contrast to rural stress, focused more on larger economic and financial issues. Also, themes around urban stress seemed to be collective, such as stress on the financial markets, stress on the economy, or stress on the people at large.

We also pulled the messages, which have the most prevalence of a topic, for each of the topics to observe qualitative examples. Rural messages included topics of (1) *body and health*: '饮食习惯不良，使青春痘恶化，生活不规律。压力过大...' – translating to 'Poor eating habits make acne worse; irregular life; too much stress...'; (2) *love and relationships*: '经营爱情或是经营婚姻都让人疲惫！' – translating to 'Maintaining love or maintaining marriage makes people tired!'; and (3) *achievement*: '压力带来动力，愿景决定成就' – translating to 'Pressure brings motivation; vision determines achievement.'

Urban messages included stress surrounding topics of (1) *career*: ' 听说要绩效工资了，上班的心情更加沉重了' – translating to 'I heard that performance pay is coming, so the feeling of going to work is heavier.'; (2) *finance*:'为了减轻生活压力，年轻人该如何理财' – translating to 'In order to relieve the pressure of life, how should young people manage their finances?'; and (3) *family*:'延迟退休是必然选择，有助于缓解抚养压力' – translating to 'Postponement of retirement is an inevitable choice, which will help ease the pressure of supporting elderly dependency.'

LIWC: Table 1 shows correlated LIWC categories for rural stress, and Table 2 shows correlated LIWC categories for urban stress. LIWC category groups that contain categories significantly associated with rural stress are affective processes, biological processes, cognitive processes, function words, perceptual processes, personal concerns, and singular pronouns.

Words associated with higher negative emotions and cognitive processes such as sadness, discrepancies, tentativeness, auxiliary verbs, and negations were predominant in stress mentions in rural areas. Many words in these categories suggest dissatisfaction, such as 'give up', 'want', and

118

Group	Category	Effect Size	Top Words
Affective Processes	Pos. emotion	-.071	it is good, love, happy
	Neg. emotion	-.070	pressure, burden, exhausted
	Sadness	-.043	give up, tired, pain
	Anxiety	-.030	pressure, upset, anxiety
	Inhibition	-.022	wait, depressed, maintain
Biological Processes	Health	-.041	exhausted, life, health
	Sexual	-.032	love, woman, desire
	Body	-.028	feel, life, body
Cognitive Processes	Discrepancy	-.064	will, want, can
	Tentative	-.027	can, think, if
	Causation	-.025	let, because, so
	Certainty	-.019	everyone, really, every day
Function Words	Aux. verbs	-.050	want, can, may
	Multifunction	-.048	of, yes, have
	Negations	-.037	no, none, don't
	Conjunctions	-.032	then, and, also
	Quantifiers	-.013	many, big, more
Perceptual Processes	See	-.034	video, beautiful, note
	Feel	-.024	heavy, feel, skin
Personal Concerns	Achievement	-.033	jobs, need, work hard
	Home	-.032	family, sleep, go to bed
	Leisure	-.028	easy, health, drink
Pronouns	You	-.055	you
	Impersonal	-.042	people, it, that
	Personal	-.039	you, I, oneself
	She/he	-.027	he, she, people
	I	-.018	I, oneself, people

Table 1: LIWC categories with translated top words associated with stress in rural areas; significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected.

Group	Category	Effect Size	Top Words
Cognitive Processes	Insight	.031	⌐, think, know
Function Words	Specifying articles	.032	on, under, before
Informal Language	Assent	.043	⌐, yes, correct
	Nonfluencies	.031	just is, what, that
Personal Concerns	Work	.039	jobs, need, work hard
	Money	.023	yuan, economy, buy
Pronouns	They	.405	they
	We	.026	everyone, we
Relativity	Time	.026	when, after, month
	Motion	.016	to, on, go to
Social Processes	Family	.018	child, parents, family
Tenses	Progressive tense	.048	⌐, already
	Future tense	.037	will, look, after

Table 2: LIWC categories with translated top words associated with stress in urban areas; significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected. The word ⌐ was untranslated as there is no corresponding English word due to its use as a perfective aspect.

'need'. Categories around stress in rural areas also consisted of personal pronouns and other personal constructs such as health, sexual references, achievement, home, and leisure. On the other hand, urban stress mentions were predominated by time/relativity, work/money, plural personal pronouns, and informal language. Further, while words indicative of future and progressive tenses were significantly associated with urban stress posts, words associated with past tenses were not.

Gallup Polls on Variation of Stress with Urbanization

Data

To corroborate the results above with the broader societal context behind urban-rural differences in stress, we compared the insights from social media data in the context of self-reported comorbidities of stress with survey results from Gallup World Polls, which consist of survey responses from representative individuals (Deaton 2008). Annual national panels representative of the demographics in China across provinces were contacted daily by phone or face-to-

face interviews by Gallup. We aggregated responses across 29 provinces and across 4 different well-being categories – physical, social, and financial wellness along with questions measuring general stress and worry. We used Gallup Polls data from 2006-2011 (the entire duration for which the responses from China were available in the dataset). 17,712 responses were used in this study, averaging 2,952 responses per year across 6 years between 2006-2011. We then regressed the weighted averages of Gallup outcomes per province against the percent urbanization of the province, obtained from the 2015 China statistical yearbook (Statistics 2015). We specifically used, per question, the percentage of all responses that was the most affirmative response (i.e. responded "Yes" or "Satisfied") as the outcome variable for regression. Table 3 shows the questions and the available responses, as well as regression coefficients for the percent urbanization of a province. The regression model is defined in Eq. 2

$$Prov._{\%RespondedYes} = \beta_1 * Prov._{\%Urban} + \epsilon \quad (2)$$

Differential Stress and Well-being in Urban-Rural Areas

Rural residents reported having higher levels of personally experienced stress and worry than their urban counterparts (Table 3). In the language analysis, we found higher occurrence of personal stress with more negative emotion words in the rural LIWC categories compared to urban language (Table 1). On the other hand, urban stress topics expanded beyond personal concerns, and the word stress itself was seen to be applied beyond the individual (Figure 4).

In terms of physical wellness, urban areas were found to be more well-rested and experience less pain than their rural

Category	Question	Responses	Effect Size (C.I.)
General Stress	Did you experience the following feelings during a lot of the day yesterday? How about worry?	Yes, No	-0.316 (-0.539,-0.093)
	Did you experience the following feelings during a lot of the day yesterday? How about stress?	Yes, No	-0.308 (-0.531,-0.084)
Physical Wellness	Did you feel well-rested yesterday?	Yes, No	0.177 (0.012,0.342)
	Did you experience the following feelings during a lot of the day yesterday? How about physical pain?	Yes, No	-0.267 (-0.416,-0.118)
Financial Wellness	Are you satisfied or dissatisfied with your standard of living, all the things you can buy and do?	Satisfied, Dissatisfied	0.369 (0.115,0.623)
	Thinking about the job situation in the city or area where you live today, would you say that it is now a good time or a bad time to find a job?	Good Time, Bad Time	0.483 (0.249,0.717)
	Which one of these phrases comes closest to your own feelings about your household's income these days?	[Living Comfortably], [Getting By, Finding it Difficult, Finding it Very Difficult]	0.256 (0.140,0.373)
Social Wellness	If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?	Yes, No	0.273 (0.056,0.490)
	In the city or area where you live, are you satisfied or dissatisfied with the opportunities to meet people and make friends?	Satisfied, Dissatisfied	0.586 (0.161,1.011)
Civic Engagement	Have you done any of the following in the past month? How about donated money to a charity?	Yes, No	0.240 (0.052,0.427)
	Have you done any of the following in the past month? How about volunteered your time to an organization?	Yes, No	0.123 (0.055,0.191)

Table 3: Regression coefficients and confidence intervals associated with Gallup well-being variables with percent urbanization. Coefficients are the effect size of the percent urbanization variable on the category response, and all correlations are significant at $p < .05$, two-tailed t-test. For the third question in Financial Wellness, 'Living Comfortably' was taken as one class (set to 1) and remaining three were taken as another for regression (as 0).

counterparts. The language analysis corroborated this finding, for rural users expressed more stress surrounding physical pain, tiredness, and insomnia (Figure 3).

In terms of financial wellness, reports of standard of living, job opportunities, and household income comfort increased with urbanization (Table 3). Although both urban and rural users in China expressed work stress on social media, our language analysis showed that rural users express stress surrounding (the lack of) achievement, costs, and desire, all topics that suggest dissatisfaction with one's financial wellness (Figure 3).

Lastly, we found increasing social wellness with urbanization. Urban residents indicated having reliable friends and relatives to count on in times of trouble. Our language analyses showed that both urban and rural users had stress surrounding family and relationships, but notably, topics in rural areas discussed more about personal relationships like marriage (Figure 3) whereas urban topics discussed more about collective responsibilities for elders and youth (Figure 4).

Civic Engagement

We also used Gallup data to examine if urban users in China had more collective and societal concerns compared to rural users as we found in our language analyses that urban users expressed more stress on topics surrounding politics, governance, and society (Figure 4) and used plural personal pronouns like 'they' (Table 2). We observed that provinces with

higher levels of urbanization also had higher levels of civic engagement. The civic engagement questions asked respondents whether they engaged in activities such as donating to charity and volunteering at organizations in their past month. People in urbanized provinces of China reported more donations to charity and more volunteering work than people in rural provinces (Table 3). These findings are corroborated by previous studies that report that higher socioeconomic status in China is associated with higher political participation (Appleton and Song 2008).

Discussion

This study modelled expressions of psychological stressors in urban and rural regions in China by applying natural language processing and hierarchical mixed effects modeling to social media users across counties and provinces in China. Our findings provide a complementary perspective to better understand the urban-rural differences in stress experiences compared to traditional survey assessments (Hawn 2009). We found that rural users tended to be more personal in nature and used more emotions when discussing stress (Table 1) compared to urban users who tended to use relative and temporal terms (Table 2).

In addition, users in the rural areas had more frequent uses of singular personal pronouns, while urban users had more frequent uses of plural personal pronouns when mentioning stress. We also found that rural users appeared more concerned with personal matters, such as health, achievement,

and personal relationships (Figure 3), while urban users appeared more stressed about external and collective issues, such as work, politics, and economics (Figure 4). These findings reflect several results obtained by previous traditional survey-based studies, in which health care, job security, and family disputes increase rural Chinese suicide risks (Zhang and Wang 2012), lifestyle (i.e., reasonable diet, physical activity, etc.) and social support have direct or indirect effects on psychological distress in rural China (Feng, Ji, and Xu 2013), and work is the dominant stress-introducing context in urban China (Lin and Lai 1995). We also used Gallup polls to corroborate insights from the language analysis, supporting the use of social media as a passive sensor of stressors in urban-rural areas in China.

Rural dwellers tended to directly express their symptoms of stress such as frustration, being overwhelmed, and physical pain (Figure 3), while urban residents seemed more concerned with societal events that could eventually trigger stress such as the economic status of the country, the relationship between China and the US, and policies of the banks (Figure 4). As the Internet is more accessible in urban than in rural areas (Song, Wang, and Bergmann 2020), urban citizens are likely more exposed to the most recent news. With higher exposure, they are more likely to talk about and be more concerned about public affairs (Yao et al. 2019). These differences exist beyond controlling for GDP, indicating a fundamentally different lifestyle between rural and urban residents in very specific environments, arguably having different sources of stress (Cui et al. 2012).

Though rural and urban communities both seemed stressed while referring to family, topics in rural communities mentioned mainly about personal relationships like marriage (Figure 3); by contrast, urban communities talked more about responsibilities for elders and youth (Figure 4). This result is consistent with a previous large-scale survey which indicated that first-tier and second-tier developed cities have higher living pressure in housing, education, medical care, and supporting the elderly, while third-tier, fourth-tier, and lower-tier cities have higher levels of family and interpersonal stress (Ying 2016). This can potentially be explained by the Confucian norms of filial piety, which 'requires children to provide care for their elderly parents, putting immense social and moral pressure on adult children to fulfill their parental care responsibilities' (Lee and Xiao 1998; Zimmer and Kwong 2003). The cost of living is higher in big cities, so the pressure to meet these obligations is greater. Moreover, urban individuals expressed more stress surrounding family (e.g., parents, retirement, illness) in comparison to the rural users (Figure 4), potentially exposing more work-family conflicts in urban areas owing to more demanding jobs (Ling and Poweli 2001). While rural residents could provide immediate financial, emotional, and physical care because they are more likely to live in the same household, urban residents (especially migrants) may often be absent and thus cause more stress (Verheij 1996).

Prior work has shown that digital traces can predict Big-Five personality dimensions with ranging from r = .29 to .40 (Settanni, Azucar, and Marengo 2018), which is similar to the highest correlations between psychological traits and observable behaviors in the large psychological literature (r = .3). In this work, we evaluated the influence of users' communities on individuals' language and consequently expected much smaller effect sizes (Hagerty 2000). More importantly, these highly significant correlations demonstrate that Weibo posts encode significant well-being and health information that can be used to study the influence of communities in which individuals are located on their well-being.

Differences Between United States and China

This study complements work which explored similar questions in the context of the United States (Jaidka et al. 2020b). The authors used a dataset of 1.5 billion Twitter posts between January 2009 and December 2015 to identify the topics corresponding to higher stress in rural and urban United States (US).

While there are inevitable differences in how the data in both studies were collected, prepared, and analyzed, it is nevertheless insightful to see whether the comparative differences in rural and urban areas persist across different cultural contexts. One similarity between the two studies is that higher emotional expression was associated with higher stress across rural communities, after controlling for socioeconomic status. The authors also reported that mentions of poor health were higher in rural areas with low socioeconomic status. We, too, found rural users to more likely discuss personal topics such as health on Weibo.

Similar to the US study, we found that discussions of politics and the economy were more likely to be posted by urban rather than rural dwellers. However, one major difference between the two studies is that, US urban communities reporting higher stress were also more likely to discuss relationships on Twitter (Jaidka et al. 2020b). However, in our study, we did not find the same pattern. Our results revealed that urban dwellers discussed not relationships but their family as an indicator of stress (Figure 4), whereas rural areas talked about marriage and romantic relationships (Figure 3). This may be due to the unequal distribution of gender in China's rural areas and the so called 'Caili' culture.

Limitations, Ethics, and Future Work

This study, like several social media based studies, has many limitations. First, our findings may not cover the full picture of stress because of the internet censorship in China (Vuori and Paltemaa 2015) as we only use publicly available posts. We assume a few stress-related words as reported in previous research such as psychosomatic symptoms, substance use, or suicidal ideation may not be present in our dataset because they could harm the establishment of a 'healthy and harmonious Internet environment' (Paltemaa et al. 2020).

China's Bureau of Statistics recommends mapping neighborhood committees to urban-rural regions and we did not have access to such granular location for Weibo users in our dataset. That said, the current county tier classification system that maps counties into urban and rural regions based on a tier system serves as a reasonable proxy as has been seen in prior works (Long 2016). Further, we replicated the findings on language data at the province-level. Discussions around the use of social media based health indices should

include public health experts, computer scientists, lawyers, ethicists, clinicians, policy makers, and individuals from different socioeconomic and cultural backgrounds (Benton, Coppersmith, and Dredze 2017).

There are several potential ways to broaden our findings. First, we analyzed posts that mention several variants of psychological stress which is potentially a subset of all posts that truly indicate a stressed mental state. While language-based estimates of stress have been validated in English using traditional survey instruments (Guntuku et al. 2019a), future work could examine its correlation with other well-being facets in Mandarin. Second, specific sub-populations (e.g. immigrants) who have unique stressors due to diverse reasons (e.g. 'hukou') that separate rural and urban residents into disparate social, economic, and political spheres (Chan 2014) could be examined. Despite living in cities, migrant peasant workers still maintain their rural hukou type and are treated as rural residents, with little to no access to urban social security, which can trigger different degrees and aspects of stress. Further, users are likely to express themselves differently across social media platforms. In this study, we only used data from one major social media platform - Weibo. It would be interesting to examine if the results would differ in other platforms, as has been found in the US (Jaidka, Guntuku, and Ungar 2018).

In summary, our findings suggest that a nuanced analysis of regional social media usage is necessary, in order to situate it in an understanding of the digital divide in internet and social media usage. Differences in social contexts are related to a differential use of social media to cope with daily stressors and act as a buffer for stress and subjective well-being. Despite societal growth and technological advancements, the characteristics of urban and rural life appear to be replicated and reinforced in the online sphere – in context after context, and culture after culture.

Appendix

Stress Keywords along with literal translations: 压力— Pressure, 鸭梨— Pear, 压迫— Oppression, 负担— Burden, 累赘— Burden, 重负— Heavy burden, 重担— Heavy burden, 重压— Heavy pressure, 压抑— Suppress, 沉重— Heavy, 压力大— High pressure, 压力山大— Tremendous stress, 心理压力— Psychological stress, 慢性压力— Chronic stress, 生活压力— Life stress, 工作压力— Work stress, 经济压力— Economic stress, 精神压力— Mental stress, 婚姻压力— Marriage stress, 环境压力— Environmental pressure, 租房压力— Rental pressure, 买房压力— Pressure of buying a house, 职场压力— Career stress, 就业压力— Employment stress, 年龄压力— Age stress, 颜值压力— Appearance stress, 同辈压力— Peer pressure, 学习压力— Study-induced stress, 学业压力— Academic stress and, 考试压力— Exam stress.

Posts containing certain keywords associated with market stress, physical pressure, etc. were dropped to reduce false positives; specifically, 资本压力— Capital stress, 胎压— Tire pressure, 轮胎压力— Tire pressure, 股票— Stock, 大盘— Market index, 指数— Index, 压力位— Resistance level, 压力线— Resistance line, 支撑位— Support level, 支撑线— Support line, 压迫力— Constriction, 承压— Compression, 交通压力— Traffic pressure, 企业压力— Enterprise pressure, 肾脏负担— Kidney pressure and, 皮肤压力— Skin pressure.

Post-Stratification: We also replicated the language analyses for Weibo users at the county level with post-stratified samples on gender so that our sampled data from Weibo can match offline population gender ratios per province (results shown in Supplementary Material 3).

We reweighed each language feature as defined in Eq. 3:

$$PoststratFactor_p = \frac{WeiboMale_p}{WeiboFemale_p} * \frac{PopFemale_p}{PopMale_p}$$

$$Feature_{i_{poststratified}} = \begin{cases} Feature_i * PoststratFactor_p \\ \quad\quad\quad \text{if user is female} \\ Feature_i * \frac{1}{PoststratFactor_p} \\ \quad\quad\quad \text{if user is male} \end{cases}$$
(3)

where p indicates the province that the user is located in.

Code and Supplementary Material at https://github.com/jessecui/WWBP-China-Urban-Rural-Stress

References

Ahmed, S.; Cho, J.; Jaidka, K.; Eichstaedt, J. C.; and Ungar, L. H. 2020. The Internet and Participation Inequality: A Multilevel Examination of 108 Countries. *International Journal of Communication*, 14: 22.

Appleton, S.; and Song, L. 2008. Life satisfaction in urban China: Components and determinants. *World Development*.

Bagroy, S.; Kumaraguru, P.; and De Choudhury, M. 2017. A social media based index of mental well-being in college campuses. In *Proceedings of the 2017 CHI Conference on Human factors in Computing Systems*, 1634–1646.

Beaudoin, C. E.; and Thorson, E. 2004. Social capital in rural and urban communities: Testing differences in media effects and models. *Journalism & Mass Communication Quarterly*, 81(2): 378–399.

Benjamini, Y.; and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300.

Benton, A.; Coppersmith, G.; and Dredze, M. 2017. Ethical research protocols for social media health research. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, 94–102.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research*.

Bouma, G. 2009. Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, 31–40.

Bucher, T.; and Helmond, A. 2017. The affordances of social media platforms. *The SAGE handbook of social media*.

Chan, K. W. 2014. China's urbanization 2020: a new blueprint and direction. *Eurasian Geography and Economics*, 55(1): 1–9.

Chung, J. H.; and Lam, T.-c. 2004. China's "city system" in flux: Explaining post-mao administrative changes. *The China Quarterly*, 180: 945–964.

Cui, X.; Rockett, I. R.; Yang, T.; and Cao, R. 2012. Work stress, life stress, and smoking among rural–urban migrant workers in China. *BMC Public Health*, 12(1): 979.

De Choudhury, M.; Sharma, S. S.; Logar, T.; Eekhout, W.; and Nielsen, R. C. 2017. Gender and cross-cultural differences in social media disclosures of mental illness. In *Proceedings of the 2017 ACM conference on CSCW*, 353–369.

Deaton, A. 2008. Income, health, and well-being around the world: Evidence from the Gallup World Poll. *Journal of Economic perspectives*, 22(2): 53–72.

DiMaggio, P.; and Hargittai, E. 2001. From the 'digital divide'to 'digital inequality': Studying Internet use as penetration increases. *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University*, 4(1): 4–2.

Evans, G. W.; and English, K. 2002. The environment of poverty: Multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. *Child development*.

Feng, D.; Ji, L.; and Xu, L. 2013. The influence of social support, lifestyle and functional disability on psychological distress in rural China: Structural equation modelling. *Australian Journal of Rural Health*, 21(1): 13–19.

Gilbert, E.; Karahalios, K.; and Sandvig, C. 2010. The Network in the Garden: Designing Social Media for Rural Life. *American Behavioral Scientist*, 53(9): 1367–1388.

Giorgi, S.; Guntuku, S. C.; Eichstaedt, J. C.; Pajot, C.; Schwartz, H. A.; and Ungar, L. H. 2021. Well-Being Depends on Social Comparison: Hierarchical Models of Twitter Language Suggest That Richer Neighbors Make You Less Happy. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, 1069–1074.

Guntuku, S. C.; Buffone, A.; Jaidka, K.; Eichstaedt, J. C.; and Ungar, L. H. 2019a. Understanding and measuring psychological stress using social media. In *Proceedings of the ICWSM*, volume 13, 214–225.

Guntuku, S. C.; Klinger, E. V.; McCalpin, H. J.; Ungar, L. H.; Asch, D. A.; and Merchant, R. M. 2021. Social media language of healthcare super-utilizers. *NPJ digital medicine*.

Guntuku, S. C.; Li, M.; Tay, L.; and Ungar, L. H. 2019b. 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, 226–235.

Guntuku, S. C.; Schwartz, H. A.; Kashyap, A.; Gaulton, J. S.; Stokes, D. C.; Asch, D. A.; Ungar, L. H.; and Merchant, R. M. 2020. Variability in language used on social media prior to hospital visits. *Scientific reports*, 10(1): 1–9.

Hagerty, M. R. 2000. Social comparisons of income in one's community: Evidence from national surveys of income and happiness. *Journal of Personality and Social Psychology*.

Hawn, C. 2009. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. *Health affairs*, 28(2): 361–368.

Helliwell, J. F.; and Putnam, R. D. 2004. The social context of well-being. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1449): 1435.

Jackson, L. A.; and Wang, J.-L. 2013. Cultural differences in social networking site use: A comparative study of China and the United States. *Computers in human behavior*.

Jaidka, K.; Giorgi, S.; Schwartz, H. A.; Kern, M. L.; Ungar, L. H.; and Eichstaedt, J. C. 2020a. Estimating geographic subjective well-being from Twitter: a comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*.

Jaidka, K.; Guntuku, S.; and Ungar, L. 2018. Facebook versus Twitter: Differences in self-disclosure and trait prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12.

Jaidka, K.; Guntuku, S. C.; Lee, J. H.; Luo, Z.; Buffone, A.; and Ungar, L. H. 2020b. The rural–urban stress divide: Obtaining geographical insights through Twitter. *Computers in Human Behavior*, 106544.

Jiang, Q.; Zhang, Y.; and Sánchez-Barricarte, J. J. 2015. Marriage expenses in rural China. *The China Review*.

Knight, J.; and Song, L. 1999. The rural-urban divide: Economic disparities and interactions in China. *OUP Catalogue*.

Lederbogen, F.; Kirsch, P.; Haddad, L.; Streit, F.; Tost, H.; Schuch, P.; Wüst, S.; Pruessner, J. C.; Rietschel, M.; Deuschle, M.; et al. 2011. City living and urban upbringing affect neural social stress processing in humans. *Nature*.

Lee, Y.-J.; and Xiao, Z. 1998. Children's support for elderly parents in urban and rural China: Results from a national survey. *Journal of cross-cultural gerontology*, 13(1): 39–62.

Li, L.; Li, A.; Hao, B.; Guan, Z.; and Zhu, T. 2014. Predicting active users' personality based on micro-blogging behaviors. *PloS one*, 9(1): e84997.

Li, M.; Guntuku, S.; Jakhetiya, V.; and Ungar, L. 2019. Exploring (dis-) similarities in emoji-emotion association on twitter and weibo. In *Companion proceedings of the 2019 world wide web conference*, 461–467.

Li, M.; Hickman, L.; Tay, L.; Ungar, L.; and Guntuku, S. C. 2020. Studying Politeness across Cultures Using English Twitter and Mandarin Weibo. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2): 1–15.

Lin, N.; and Lai, G. 1995. Urban stress in China. *Social Science & Medicine*, 41(8): 1131–1145.

Ling, Y.; and Poweli, G. N. 2001. Work-family conflict in contemporary China: Beyond an American-based model. *International Journal of Cross Cultural Management*.

Liu, Z. 2005. Institution and inequality: the hukou system in China. *Journal of comparative economics*, 33(1): 133–157.

Long, Y. 2016. Redefining Chinese city system with emerging new data. *Applied geography*, 75: 36–48.

Lui, M.; and Baldwin, T. 2012. langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*, 25–30.

Paltemaa, L.; Vuori, J. A.; Mattlin, M.; and Katajisto, J. 2020. Meta-information censorship and the creation of the Chinanet Bubble. *Information, Communication & Society*.

Partridge, M. D.; and Rickman, D. S. 2008. Distance from urban agglomeration economies and rural poverty. *Journal of Regional Science*, 48(2): 285–310.

Pennebaker, J. W.; Boyd, R. L.; Jordan, K.; and Blackburn, K. 2015. The development and psychometric properties of LIWC2015. Technical report.

Quercia, D.; Seaghdha, D. O.; and Crowcroft, J. 2012. Talk of the City: Our Tweets, Our Community Happiness. In *Proceedings of the Sixth AAAI ICWSM*, 555–558.

Rentfrow, P. J. 2010. Statewide differences in personality: Toward a psychological geography of the United States. *American Psychologist*, 65(6): 548.

Robinson, L.; Cotten, S. R.; Ono, H.; Quan-Haase, A.; Mesch, G.; Chen, W.; Schulz, J.; Hale, T. M.; and Stern, M. J. 2015. Digital inequalities and why they matter. *Information, communication & society*, 18(5): 569–582.

Saha, K.; Kim, S. C.; Reddy, M. D.; Carter, A. J.; Sharma, E.; Haimson, O. L.; and De Choudhury, M. 2019. The language of LGBTQ+ minority stress experiences on social media. *Proceedings of the ACM on human-computer interaction*, 3(CSCW): 1–22.

Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Dziurzynski, L.; Ramones, S. M.; Agrawal, M.; Shah, A.; Kosinski, M.; Stillwell, D.; Seligman, M. E.; and Ungar, L. H. 2013. Personality, gender, and age in the language of social media: The Open-Vocabulary approach. *PLoS ONE*.

Settanni, M.; Azucar, D.; and Marengo, D. 2018. Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behavior, and Social Networking*, 21(4): 217–228.

Song, Z.; Wang, C.; and Bergmann, L. 2020. China's prefectural digital divide: Spatial analysis and multivariate determinants of ICT diffusion. *International Journal of Information Management*, 102072.

Statistics, N. 2015. China statistical yearbook 2015. http://www.stats.gov.cn/tjsj/ndsj/2015/indexeh.htm. Accessed: 2020-11-10.

Synder, M. 2000. Governmental Control and Cultural Adaptation: A Comparison between Rural and Urban Reactions to China's Fertility Control Policies. https://www.reed.edu/Luce/documents/SnyderLuceReport.pdf. Accessed: 2019-11-10.

Thomala, L. L. 2021. Number of Sina Weibo users in China from 2017 to 2021. https://www.statista.com/statistics/941456/china-number-of-sina-weibo-users/. Accessed: 2021-11-03.

Tian, X.; Batterham, P.; Song, S.; Yao, X.; and Yu, G. 2018. Characterizing depression issues on Sina Weibo. *International journal of environmental research and public health*.

Tian, X.; He, F.; Batterham, P.; Wang, Z.; and Yu, G. 2017. An analysis of anxiety-related postings on Sina Weibo. *International journal of environmental research and public health*, 14(7): 775.

Van Loon, A.; Stewart, S.; Waldon, B.; Lakshmikanth, S. K.; Shah, I.; Guntuku, S. C.; Sherman, G.; Zou, J.; and Eichstaedt, J. 2020. Explaining the 'Trump Gap' in Social Distancing Using COVID Discourse.

Verheij, R. A. 1996. Explaining urban-rural variations in health: a review of interactions between individual and environment. *Social science & medicine*, 42(6): 923–935.

Vuori, J. A.; and Paltemaa, L. 2015. The lexicon of fear: Chinese internet control practice in Sina Weibo microblog censorship. *Surveillance & society*, 13(3/4): 400–421.

Walker, R. E.; Keane, C. R.; and Burke, J. G. 2010. Disparities and access to healthy food in the United States: A review of food deserts literature. *Health & place*, 16(5): 876–884.

Wang, Z.; Yu, G.; Tian, X.; Tang, J.; and Yan, X. 2018. A study of users with suicidal ideation on Sina Weibo. *Telemedicine and e-Health*, 24(9): 702–709.

Ward, N.; and Brown, D. L. 2009. Placing the rural in regional development. *Regional studies*, 43(10): 1237–1244.

White, M. P.; Alcock, I.; Wheeler, B. W.; and Depledge, M. H. 2013. Would you be happier living in a greener urban area? A fixed-effects analysis of panel data. *Psychological science*, 24(6): 920–928.

Woltman, H.; Feldstain, A.; MacKay, J. C.; and Rocchi, M. 2012. An introduction to hierarchical linear modeling. *Tutorials in quantitative methods for psychology*, 8(1): 52–69.

Yao, L.; Li, X.; Li, Q.; and Wang, J. 2019. Temporal and spatial changes in coupling and coordinating degree of new urbanization and ecological-environmental stress in China. *Sustainability*, 11(4): 1171.

Ying, X. 2016. Residents' sense of life pressure: a comparative study of cities. In *Annual report on social mentality of China (2016)*, 69–82. Social Science Literature Press.

Zamani, M.; Schwartz, H. A.; Eichstaedt, J.; Guntuku, S. C.; Ganesan, A. V.; Clouston, S.; and Giorgi, S. 2020. Understanding weekly COVID-19 concerns through dynamic content-specific LDA topic modeling. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2020, 193. NIH Public Access.

Zhang, J.; and Wang, C. 2012. Factors in the neighborhood as risks of suicide in rural China: a multilevel analysis. *Community mental health journal*, 48(5): 627–633.

Zhang, W.; Caines, A.; Alikaniotis, D.; and Buttery, P. 2016. Predicting Author Age from Weibo Microblog Posts. In *LREC*.

Zhao, N.; Jiao, D.; Bai, S.; and Zhu, T. 2016. Evaluating the validity of simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services. *PLoS One*, 11(6): e0157947.

Zhu, S.; Yu, C.; and He, C. 2020. Export structures, income inequality and urban-rural divide in China. *Applied Geography*, 115: 102150.

Zimmer, Z.; and Kwong, J. 2003. Family size and support of older adults in urban and rural China: Current effects and future implications. *Demography*, 40(1): 23–44.