Twitter Analysis: Studying US Weekly Trends in Work Stress and Emotion

Wei Wang*

University of Central Florida, USA

Ivan Hernandez

Northwestern University, USA

Daniel A. Newman

University of Illinois at Urbana-Champaign, USA

Jibo He

Wichita State University, USA

Jiang Bian

University of Florida, USA

We propose the use of Twitter analysis as an alternative source of data to docu-ment weekly trends in emotion and stress, and attempt to use the method to estimate the work recovery effect of weekends. On the basis of 2,102,176,189 Tweets, we apply Pennebaker s linguistic inquiry word count (LIWC) approach to measure daily Tweet content across 18 months, aggregated to the US national level of analysis. We derived a word count dictionary to assess work stress and applied p-technique factor analysis to the daily word count data from 19 substantively different content areas covered by the LIWC diction-aries. Dynamic factor analysis revealed two latent factors in day-level variation of Tweet content. These two factors are: (a) a negative emotion/stress/somatic factor, and (b) a positive emotion/food/friends/home/family/leisure factor, onto which elements of work, money, achievement, and health issues have strong negative loadings. The weekly trend analysis revealed a clear "Friday dip" for work stress and negative emotion expressed on Twitter. In contrast, positive emotion Tweets showed a "mid-week dip" for Tuesday-Wednesday-Thursday and "weekend peak" for Friday through Sunday, whereas work/money/achieve-ment/health problem Tweets showed a small "weekend dip" on Fridays through Sundays. Results partially support the Effort-Recovery theory. Impli-cations and limitations of the method are discussed.

---

* Address for correspondence: Wei Wang, Department of Psychology, University of Central Florida, 4111 Pictor Lane, Psychology Bldg 99 Ste. 320, Orlando, FL 32816, USA. Email: wei.wang@ucf.edu

INTRODUCTION

In the era of popular social media and high technology, researchers can instantly access “Big Data”, a collection of data that is generated from social media, internet, and many other large-scale sources (George, Hass, & Pent-land, 2014). Big data are not only large in size, but also high in velocity and complex in variety. More importantly, big data approaches are typically unob-trusive in data generation and collection. Thus the approach directly addresses some of the well-known methodological problems of small sample and static research (i.e. low power, inability to assess intensive longitudinal trends).

Even more excitingly, big data often contain rich health-related information, coupled with precise location and real-time information, making it especially suitable for public health research. One famous example showed that people s queries on the Google search engine could predict the frequency of flu cases in a given area (Ginsberg, Mohebbi, Patel, Brammer, Smolinski, & Brilliant, 2009; also see Padmanabhan, Wang, Cao, Hwang, Zhang, Gao, Soltani, & Liu, 2014). Services like Google Maps and Waze estimate people s speed via the GPS tracker in their phones when driving on a highway, and from that information can detect traffic jams and road outages. Recently, Facebook pub-lished a study conducted on its own users (approximately 61 million active accounts), where they experimentally manipulated the emotional content of posts people saw when they logged into the site (Kramer, Guillory, & Hancock, 2014). People assigned to see fewer posts with negative content from their friends were more likely to post positive content themselves. Similarly, people assigned to see fewer posts with positive content were more likely to post nega-tive content. Further, they found a contagion effect, where people who were friends with those who saw positive/negative messages were themselves more likely to post positive/negative content. These studies all implicate the power of big data to potentially improve our knowledge of public health, and represent only a small sampling of research possibilities.

Although blooming in many other areas such as public health and manage-ment, big data techniques have not been well exploited in industrial-organisational psychology (see Tonidandel, King & Cortina, 2015), especially in the field of work and organisation health research. The current paper attempts to begin bridging this gap.

BIG DATA AND TWITTER

Big data s popularity among researchers has rapidly increased over the past few years (Lohr, 2013). Although many definitions of Big Data have been offered, varying from general (Microsoft, 2013) to specific (Intel, 2012), one definition that has become ubiquitous formalises the various features of big data into three dimensions: Volume, Velocity, and Variety (Laney, 2001). That is, the increasing size of data, the increasing rate at which it is produced, and the increasing range of formats and people employed represent how the nature of data is changing in the 21st century.

Each day, billions of people create recorded data points on social media, portable devices, financial markets, transportation, and other sources. Not only are these data richer than the sources available a decade ago, but they are also more open and available to interested researchers. Therefore, we are at an important point in time where researchers can ask questions and consider methods not previously possible.

Twitter

Twitter (http://twitter.com)—an online micro-blogging service where users share and receive text messages/"Tweets" of up to 140 characters—provides an ideal source for big data research. Twitter is currently the 8th most pop-ular website in the world. Over 288 million people actively use the site on a monthly basis. If Twitter were a country, it would be the fourth largest by population. This massive user base sends over 500 million Tweets per day, and they arrive at high velocity—about 6,000 Tweets per second. Because of Twitter s short text format, users only need an internet enabled device (80% of active accounts use Twitter on mobile phones) to send a Tweet, which means those messages can serve as a snapshot of people s immediate thoughts. In addition to Twitter s volume and velocity, Twitter also has a great deal of variety in its users—Twitter supports 33 different languages, and 77 per cent of users are outside the US. Of the account holders in the US, 53 per cent are women (similar to the overall population s gender distri-bution). In addition, 60 per cent of users are White, 16 per cent are Black, and 11 per cent are Latino, which slightly over-represents minorities (who have been traditionally under-represented in lab samples). Being a social media website, Twitter may seem to only attract a younger audience. But only 63 per cent of its users are under 35. Thus the size, frequency, and diversity of Twitter s data are distinct advantages.

For big data research, Twitter presents a unique opportunity to analyze lan-guage in a real world setting. Twitter users are not prompted by specific research questions or agendas; rather, they are engaged in casual internet con-versations on a variety of topics of their own choosing. As a result, we can ana-lyze the linguistic content of these Tweets to uncover psychological states of different users, in an unobtrusive, natural setting. In addition, Twitter allows its users to "follow" (receive regular updates from) different people and groups, providing another method to examine a user s interests and attitudes. Using this publicly available information, researchers can better understand people s mental processes underlying a variety of issues.

TWITTER APPROACH TO HEALTH RESEARCH

Indeed, Twitter has been found to be particularly suitable for health research because people openly share health and even personal information (e.g. thoughts, emotions, daily activities) on Twitter. Although intuitively we might expect that people would be reticent to talk about personal health topics on social media, researchers have found that about 10 per cent of Tweets are related to health (Paul & Dredze, 2011a). Moreover, people like to ask health-related questions on Twitter (Paul, Hong, & Chi, 2011).

In recent years, Twitter has drawn tremendous attention from health researchers and practitioners. One of the most prolific research areas in this endeavor is to use Twitter to track and predict diseases like influenza. This approach is similar to Google Flu Tracking, except that Twitter s disease tracking is more powerful due to Tweets typically having more declarative and detailed health-related content than brief search queries in Google. Tweets can indicate both personal conditions and attitudes such as “I got Flu” or “down with swine flu”, or can contain even more details like, “i got fever 102.5 i got flu i got sore eyes my throat hurts taking tylenol”—describing specific symp-toms and treatments (Paul & Dredze, 2011b). Hundreds of millions of Tweets like this, coupled with their precise location and real-time information, can be tremendously informative in tracking and predicting diseases by constructing innovative monitoring systems. Such computational surveillance systems can provide instant results, predicting earlier indication of disease spread than government bureaus reports, which have been relying on doctors manual reports, usually with a 1- to 2-week delay. Indeed, many studies using Twitter have found that flu statistics derived from Twitter analyses are not only extremely highly correlated (r 0.97) with the government statistics from the Centers for Disease Control and Prevention (CDC), but also forecast flu outbreaks about 1 to 2 weeks earlier than the CDC s tracking (Culotta, 2010; Lampos & Cris-tianini, 2010; Paul & Dredze, 2011a).

Influenza only represents one of the many avenues for studying health with Twitter. Researchers have also used Twitter to investigate health topics such as dental pain (Heaivilin, Gerbert, Page, & Gibbs, 2011), sports concussion accidents (Sullivan, Schneiders, Cheang, Kitto, Lee, Redhead, Ward, Ahmed, & McCrory, 2012), and tobacco use (Prier, Smith, Giraud-Carrier, & Hanson, 2011). Twitter can even be used for health interventions such as smoking cessa-tion (Prochaska, Pechmann, Kim, & Leonhardt, 2012) and health promotion, including health program adherence and engagement (Hawn, 2009; Neiger, Thackeray, Burton, Giraud-Carrier, & Fagen, 2013) and promoting health literacy (Dumbrell & Steele, 2013; Park, Rodgers, & Stemmle, 2013).

Recently, researchers have started using Twitter big data for well-being research. This line of research focuses on deriving a computational index of subjective well-being from social media messages as an alternative to self-report well-being questionnaires (Hao, Li, Gao, Li, & Zhu, 2014). By analyz-ing Tweets from 1,300 US counties, Schwartz and his colleagues (Schwartz, Eichstaedt, Kern, Dziurzynski, Lucas, Agrawal et al., 2013) derived Twitter topics and words that predicted life satisfaction over and above traditional predictors. Nevertheless, to our knowledge, there has been no research focusing on work and organisational health using Twitter analysis.

The goal of the current study is two-fold. First, we aim to examine the US weekly trends in work stress and emotions on Twitter, based on the Effort-Recovery model (Meijman & Mulder, 1998). Second, we demonstrate an application of Twitter in work and organisational health research. In particular, we introduce the text analysis technique and demonstrate the development of a "stress" word count dictionary.

WORK STRESS AND EMOTION

Work stress is a widespread phenomenon resulting in many negative consequences for individuals, organisations, and society. In the United States, more than 70 per cent of workers report that their jobs are stressful (Clay, 2011), and one in six Americans reports being extremely stressed (Health and Safety Executive, 2008). The health consequences of this are stunning: More than a third (83 million) of US adults live with one or more types of cardiovascular disease, one of the major causes of which is stress; and the annual total cost of cardio-vascular diseases in the United States was estimated to be $444 billion (Centers for Disease Control, 2011). Work stress also contributes to many other physical and psychological problems, including stroke, headache, insomnia, digestive difficulties, cold and flu, burnout, depression, and anxiety. In addition, work stress can lead to counterproductive behaviors such as absenteeism, lateness, drug abuse, and turnover (Krantz & McCeney, 2002; Spector & Fox, 2005).

Because of its severe negative consequences, psychologists have devoted a great deal of effort to understanding the causes of stress, one of which is work schedules (Totterdell, 2005). Under this framework, abundant research focuses on the 24-hour circadian cycle effects on work stress and job affect. For exam-ple, people working outside of regular hours (i.e. outside of the 8:00 am to 5:00 pm time window) experience more stress and various subsequent health prob-lems (Smith, Folkard, Tucker, & Evans, 2011). Similarly, people with an evening or late-night working schedule experience disrupted circadian physiological rhythms, which cause sleep, digestive, and appetite disorders (Price, 2011).

However, much less research has been done to examine work stress in terms of weekly rhythms. Like the diurnal schedule, the weekly schedule serves as another temporal map that powerfully structures work and life activities (Larsen & Kasimatis, 1990). Previous research suggests that work stress is prevalent on workdays and decreases on weekends (e.g. Almeida & McDo-nald, 1998).

Although previous studies have advanced our understanding of weekly trends in work stress, they suffer some methodological limitations: participant samples were small and homogeneous, and data collection relied heavily on retrospective self-report. For example, the Almeida and McDonald (1998) research—while groundbreaking in its methods—was nonetheless based upon self-report survey data from 48 married couples recruited in the Detroit metropolitan area. These participants were all married, White, with an average of 2.5 children, and worked in relatively prestigious occupations (e.g. health technicians) with aver-age family income above the US median. These participants might not necessar-ily represent all the regions of the US, and the self-report method can be potentially biased due to memory error and experimenter demand effects.

However, we can now address research questions pertaining to weekly health trends using big data, which directly counter some of the methodological limitations of small, local samples and retrospective self-report. More importantly, because social media (e.g. Twitter) precisely record real-time reports, these data are especially suitable for time-related research questions such as stress rhythms. In addition, such data are collected in an unobtrusive (albeit public) manner, which might reduce some of the biases caused by experimenter demand characteristics.

The current study is aimed at understanding US weekly trends in work stress and emotion on Twitter. In particular, this study attempts to examine the Effort-Recovery model (Meijman & Mulder, 1998) by analyzing hundreds of millions of Tweets over 500 consecutive days. According to the Effort-Recovery model, effort expenditure at work leads to stress-related acute load reactions, which in turn lead to after-work pre-stressor levels. To maintain healthy conditions, recovery needs to be completed before a subsequent work-load starts. One of the important opportunities for such recovery is weekends. That is, individuals may experience high work stress and negative emotions in the beginning of a week and then lowered work stress and negative emotions toward the end of the weekdays; then experience the least work stress and neg-ative emotions on the weekends. In contrast, the experience of positive emo-tions is the opposite of negative emotions, in terms of the cycle from the weekdays to weekends. Previous research has empirically tested the weekend s recovery effects on health and job performance (Fritz & Sonnentag, 2005). In the current study, we compare the weekly trends for stress Tweets, negative emotion, positive emotion, and health problems that over 6 million people publicly posted on Twitter. We also attempt to distinguish work-related Tweets from nonwork-related Tweets, and conduct an exploratory analysis to assess whether the weekend recovery effect might be stronger among work-related Tweets. Based on the Effort-Recovery model, we expect to find that (a) stress and negative emotion are higher during the workweek but then dip lower on weekends, (b) positive emotion is lower during the workweek but then increases on weekends, and (c) the expected patterns for stress, negative emotion, and positive emotion are all stronger among work-related Tweets than among nonwork-related Tweets.

METHOD

We retrieved 2,102,176,189 Tweets from 46,908,115 unique Twitter users across a time period of 512 days from 25 May 2009 to 18 October 2010, from Twitter servers. This corpus represents a random sample of 10 per cent of the Tweets available, which was the maximum percentage Twitter allowed researchers to retrieve (now, Twitter only allows researchers to retrieve 1 per cent of Tweets). These data files were over 1,000 gigabytes in size.

First, we filtered these massive data files by only retaining Tweets that were in the United States and written in the English language, which were 222,217,740 (over 222 million) Tweets with 6,163,454 (over 6 million) unique Twitter users. Then we further categorised these Tweets into two groups: Tweets related to work and Tweets not related to work. The categorisation was a challenging and imperfect task, and we adopted the most conservative strat-egy. That is, only the Tweets that used a form of the word "work*" (e.g. work, worked, working, worker, etc.) or "job*" (e.g. job, jobs, jobless, etc.) were iden-tified as work-related Tweets, with the remaining categorised as nonwork-related Tweets.[1] This categorisation method increased the likelihood that most Tweets in the work group were indeed talking about work or job; for instance, "Back to work. Lots to catch up on but all is good. Projects are firing back up and moving ahead now that baseball is done." This categorisation resulted in 8,112,863 work-related Tweets, about 3.65 per cent of all Tweets written in English and posted in the US (and 1,579,506 unique Twitter users, 34.46 per cent of the total English Twitter users in the US).

To conduct daily-level analysis, we considered three categorisations of Tweets (i.e. overall Tweets, work-related Tweets, and nonwork-related Tweets) on a daily basis, and created a text file for each day for each group. This resulted in 512 days 3 3 groups 51,536 files. All the text files were then ana-lyzed as described below.

ANALYSIS

Content Analysis and the LIWC Technique

The key analytic step in big data research with Twitter is to convert the non-numeric data, such as text, into quantitative values. Researchers typically approach the issue of converting text-based data to numeric values with a

[1] We also tried categorising work vs. nonwork Tweets by using the words in the LIWC work dictionary. However, because the LIWC work dictionary was so broadly defined, using the LIWC work dictionary generated a Type I error rate of over 90 per cent.

method called content analysis. This method codes text into a numeric value indicating how much it represents a construct of interest.

The Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007) is one of the most widely used computational methods to convert text to psychological constructs. It is user friendly, supports multiple languages, and has been applied to various fields including literature, personality, political science, etc.

Although the concept of text analysis may appear to be intimidating for beginners, the logic and functionality of LIWC text analysis is surprisingly simple: counting words and calculating word frequencies. Word frequency is calculated against a word count dictionary in terms of percentages, by using the following formula:

$$\text{LIWC word frequency} = \frac{\text{word counts against a dictionary}}{\text{total word count in a text}} \times 100\%$$

where a dictionary refers to a collection of words and word stems (sometimes even phrases) that reflect or measure particular linguistic features or psychological constructs of research interest. A dictionary needs to be defined beforehand. The number of words and word stems in a dictionary varies from several to multiple hundreds. For example, the LIWC "affect" dictionary comprises 935 words and stems. LIWC provides up to 66 built-in dictionaries reflecting various text characteristics, ranging from linguistic processes to spoken catego-ries (for a comprehensive list of the dictionaries, see Table 1 in Pennebaker et al., 2007). Among all the built-in dictionaries in LIWC, the most relevant to organisational health research are 19 dictionaries including social processes (e.g. family, friends, and humans), affective processes (e.g. positive emotion, negative emotion, anxiety, anger, and sadness), biological processes (e.g. body, health, sexual, and ingestion), and personal concerns (e.g. work, achievement, leisure, home, money, religion, and death).

To analyze a text, LIWC calculates the percentage of words in the text that match a dictionary word, out of the total number of words in the text. For example, the positive emotion dictionary includes 408 stems[2] such as "amaz*", "excit*", etc. In a five-word text "I enjoyed my work today", the output by LIWC is 20 (per cent) for the positive emotion dictionary (i.e. one positive emotion word "enjoyed" divided by five total words in the text, and multiply-ing by 100%), 20 (per cent) for the work dictionary (i.e. one work word, "work"), and 0 (per cent) for the negative emotion and leisure dictionaries.

[2] LIWC dictionaries enable fuzzy search with the use of asterisks (*), so that a word stem ending with an * signals all the words beginning with that stem. For example, the word stem "excit*" signals all the words beginning with "excit" such as "excited", "exciting", "excitingly", etc.

Constructing a Stress Word Count Dictionary

Although the built-in dictionaries in LIWC are easy to access and use, the number of dictionaries is limited. Often, a dictionary specific to a researcher s interest may not be available. In such cases, new dictionaries need to be pro-posed and developed. In this study, we are particularly interested in the work-place health-relevant topic of stress. Unfortunately, LIWC does not include a stress dictionary. Therefore we needed to construct our own version of a stress dictionary.

According to Pennebaker et al. (2007), the construction of LIWC diction-aries involves four steps to ensure desired psychometric properties: (1) Word generation; (2) Word rating; (3) Psychometric validation; and (4) Updating and Expansion. We followed their procedures and undertook the first three steps for developing a stress dictionary in the current study.

Word Generation and Rating. The first step is to generate a set of words for the target dictionary from multiple sources, by brainstorming and studying existing literature (e.g. psychological scales). For example, when creating the LIWC affect dictionary and sub-dictionaries, Pennebaker et al. (2007) referred to common emotion measurement scales such as the PANAS (Watson, Clark, & Tellegen, 1988), Roget s Thesaurus, and standard English dictionaries. After generating an initial word list, Pennebaker et al. (2007) instructed three judges to rate whether or not a word on the initial list should be included in or excluded from a dictionary, and they also asked judges to add additional words they felt should be included in a dictionary but which were not on the initial list. A word was included in a dictionary if two out of the three judges agreed, and was removed from the list if two out of the three judges agreed. This step of word generation and selection ensured face validity of each construct dictionary.

Instead of using the bottom-up approach demonstrated in Pennebaker et al. (2007)—which would take a very long time to complete—our current word generation step in creating the stress dictionary took advantage of the premise that stress is one of the negative emotions, and a stress dictionary would there-fore be a subset of LIWC s built-in negative affect dictionary. Specifically, because of stress s negative affective nature, it is fairly reasonable to assume that the stress dictionary is subsumed in the overall negative affect dictionary and is considered a sub-dictionary of the negative emotion dictionary. There-fore, we started with the 511 negative affect words and word stems provided in LIWC, and used this as the initial word collection for the stress dictionary construction.

The second step is word rating. In this step, we used a similar yet more rigor-ous procedure than Pennebaker et al. (2007) did in their original study. Specifi-cally, we employed nine psychology PhD students and instructed them to judge whether each of the 511 negative words should be included or excluded from a stress dictionary, and if any new words should be added. Before the rating, all the raters were trained to understand the definition of a stress word[3] and be familiar with the rating instructions.[4] After word rating, the stress dic-tionary word selection was based on a conservative criterion: a negative affect word remained in the stress dictionary if seven out of the nine raters agreed; and a new negative affect word was added if seven out of nine raters agreed. This rating procedure resulted in 270 negative affect words in the stress dictionary.

Psychometric Validation. After the word count stress dictionary was created, we empirically evaluated its psychometric properties: reliability and validity (American Educational Research Association, American Psycholog-ical Association, & National Council on Measurement in Education, 2014). In psychometrics, reliability is most commonly evaluated by Cronbach s alpha, which assesses internal consistency based on intercorrelations and the number of measured items. In the text analysis scenario, each word in a word count dictionary is considered an item, and reliability is calculated based on each text file s “response” to each word item, which forms an N (number of text files) 3 J (number of words or stems in a dictionary) data matrix. According to Pennebaker et al. (2007), there are two ways to quan-tify such “responses”: using percentage data (uncorrected method), or using “present or not” data (binary method). For the uncorrected method, the data matrix comprises percentage values of each word/stem calculated from each text file. For the binary method, the data matrix quantifies whether or not a word was used in a text file, “1” represents yes and “0” represents no. Once the data matrix is created, it is used to calculate Cronbach s alpha based on its intercorrelation matrix among the word percentages. Pennebaker et al. (2007) claim that the uncorrected method tends to

---

[3] The definition of a stress word that was given to the raters was “A stress word is a word that is related to: (a) demanding situations, or situations that threaten one s well-being, (b) lack of resources or inability to cope with the situation, (c) lack of control, and/or (d) feeling upset, uncertain, pressured, nervous, irritated, frustrated, overwhelmed, or preoccupied with tasks.” This definition was adapted from Lazarus (1966) and Cohen, Kamarck, and Mermelstein
(1983).

[4] Rating instructions were given “When judging whether a word is stress -related, keep in mind that you are not judging whether the word means stress . Rather, you are deciding whether a person who is stressed would be more likely to say the word compared to a person who is not stressed. We expect that at least half of the negative emotion words in this excel spreadsheet can also be classified broadly as stress words. So you should have at least 200–300 words marked as stress words. Please mark 1 in the second column if the word is judged as a stress word, otherwise leave it blank. Also, if you have any words in mind that are not listed here, please feel free to add them in the right column. Thank you!”

underestimate reliability due to low base rate, and the binary method tends to overestimate reliability due to the length of texts.

We assessed reliability based on 510 Tweet files that were randomly selected from 51 states/districts in the US (i.e. 50 states plus Washington, DC). Specifically, we randomly selected 10 Twitter users from each of the 51 states/districts in the US and retrieved their Tweets during the period from May 2009 to October 2010. Hence we obtained 510 independent data files, which further generated a 510 3 270 response matrix after running the stress dictionary for each file. The response matrix yielded reliability of .96 based on the uncor-rected method, and .99 based on the binary method, which are even higher than the reliability estimates found for LIWC dictionaries.

After assessing the reliability of the stress dictionary, we attempted to evaluate its validity. We specifically focused on the two most common forms of construct validity: convergent validity and discriminant validity (Campbell & Fisk, 1959). Convergent validity provides evidence that two measures designed to assess the same construct are indeed related; discriminate validity involves evidence that two measures designed to assess different constructs are not too strongly related. In theory, we expect that the stress dictionary should be positively correlated with other negative emotion constructs to show convergent validity, and not strongly correlated with positive emotion constructs to show discriminant validity. To test these two types of validity, we used the same text files used for the reliability assessment. That is, 510 text files from 510 independent Twitter users, with 10 Twitter users randomly selected from each of the 51 states/districts in the US. The results revealed that the stress dictionary was indeed positively correlated with negative construct dictionaries, including the overall negative emotion dictionary (r 5.66, p <.001), the anxiety diction-ary (r 5.27, p <.001), the anger dictionary (r 5.48, p <.001), the sadness dic-tionary (r 5.27, p <.001), the death dictionary (r 5.15, p <.001), and the health problems dictionary (r 5.27, p <.001). In contrast, the stress dictionary was uncorrelated or negatively correlated with positive construct dictionaries, such as the overall positive emotion dictionary (r 5 2.09, p <.05), the leisure dictionary (r 5 2.08, p <.05), and the money dictionary (r 5 2.13, p <.01). These results provide initial support for the measurement validity for our newly created stress dictionary.[5]

Dynamic Factor Analysis

Before analyzing the weekly trends, we conducted exploratory factory anal-ysis to examine the clustering of dictionaries that represented the same

[5] As recommended by a helpful reviewer, it would have been ideal if we could have also demonstrated high correlations between the Twitter measure and another measure of stress collected in organisations. This was unfortunately not possible in the current design.

latent factor. This analysis provides valuable information on combining similar dictionaries together to avoid presenting redundant information in the trend analyses. The factor analysis began with the time series for the national daily level overall Tweets across 512 days. Note that this analysis is a form of the P-technique factor analysis (Cattell, 1952), which has been advocated for use with single-subject, repeated measures designs involving multiple variables measured on the same person (or same unit) intensively over time. P-technique factor analysis was designed “to identify how groups of variables covary across time for a single individual” (Jones & Nessel-roade, 1990, p. 172).

In order to account for the autocorrelation effect due to measuring consecu-tive days in the time series data, we implemented the dynamic factor analysis routine as described by Lee and Little (2012), using the MARSS function in the R package for Multivariate Auto-Regressive State-Space Modeling (MARSS; Holmes, Ward, & Scheuerell, 2014). The dynamic factor model is different from the classic P-technique factor analysis because it accounts for the time-dependence inherent in the time series data. We ran multiple dynamic factor models, including a one-factor model, two-factor model, and three-factor model, and compared interpretability of factor loadings across models to identify the optimal model. In all cases, we interpreted the models using diagonal and unequal error structures, which is similar to specifying unequal and uncorrelated uniquenesses in confirmatory factor analysis.

Weekly Trends Analysis

After conducting LIWC text analysis for the 1,536 files by using both the built-in LIWC dictionaries and the newly developed stress dictionary, then running the dynamic factor analysis, we next aggregated the LIWC text analysis scores based on a specific day of the week to examine weekly trends. That is, we aggre-gated all the Mondays together, all the Tuesdays together, and so on. We then calculated an affective index for each day from Monday to Sunday by averag-ing the LIWC scores up to the day level.

RESULTS

The dynamic factor analysis of the daily nation-level time trends across the 19 LIWC dictionaries indicated a neat two-factor solution that approached sim-ple structure (see Table 1).[6] The two factors were labeled: negative emotion/

[6] The optimal model was the two-factor solution. For the sake of comparison, the one-factor solution showed loadings very similar to Factor 1 in Table 1, with all the Factor 2 indicators showing near-zero loadings in the one-factor solution; also, the three-factor solution produced an uninterpretable third factor that showed no loadings greater than .29.

TABLE 1
Factor Loadings of Daily Twitter Data using Word Count Dictionaries on US Tweets over 18 Months (Dynamic Factor Analysis, Level of Analysis 5 Days)

Categories	Factor 1: Negative Emotion, Stress, Somatic	Factor 2: Positive Emotion, Food/Friends/Fun, Home/Leisure, VERSUS Work/Money/Achievement/Health
Negative Emotion	.75	.02
Stress	.77	2.01
Body	.63	.11
Sexual	.61	.17
Humans	.51	.05
Sadness	.41	.02
Positive Emotion	.09	.73
Ingestion	2.24	.66
Friend	.01	.68
Leisure	.00	.62
Home	2.23	.62
Family	.04	.54
Religion	.10	.41
Work	2.36	2.87
Money	2.32	2.65
Achievement	2.02	2.60
Health Issues	2.24	2.47
Death	.08	2.05
Anxiety	.03	2.16

Note: N 5 512 days. Dynamic factor model with diagonal and unequal residual matrix.

stress/somatic factor (Factor 1); and positive emotion/food-friends-family-fun versus work/money/achievement factor (Factor 2). The negative emotion, stress, sadness, body, sexual, and humans word counts loaded onto Factor 1; positive emotion, ingest, friend, leisure, home, family, work (negative loading), money (negative loading), achievement (negative loading), health issues (nega-tive loading) loaded onto Factor 2.

Hence we decided to examine the weekly trends of the two factors. It was noteworthy that work, money, achievement, and health issues had negative signs for the loadings on Factor 2, indicating that these three dictionaries reflect the opposite pole of Factor 2, compared to positive emotion, ingest, friend, leisure, home, and family. Thus we also examined the two subfactors in Factor 2 separately, and then also combined the two subfactors by subtracting work, money, achievement, and health issues from positive emotion, ingest, friend, leisure, home, and family.

We also examined four specific dictionaries: stress, negative emotion, posi-tive emotion, and health issues, and plotted the Twitter weekly trends of these

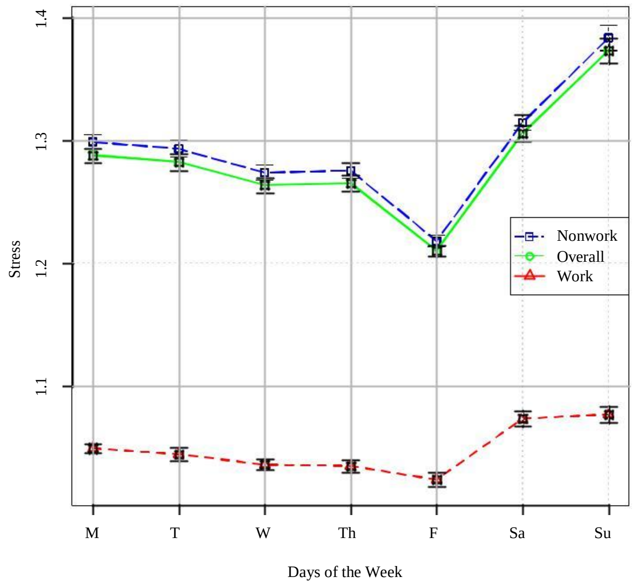

FIGURE 1. Percentage of Tweet Words about Stress, by Day of the Week. Error bars are added for a comparison purposes. Vertical axis represents the composite of word frequencies.

four dictionaries in Figures 1–4. As presented in Figures 1 and 2, the trends of stress and negative emotion on Twitter over the days of the week were very similar and showed a clear Friday dip effect: stress and negative emotion words on Twitter started with higher percentage frequencies on Mondays and slowly decreased from Mondays to Thursdays, and dipped on Fridays, partly sup-porting the expectation from the Effort-Recovery model that stress and nega-tive emotion would be lower on weekends. However, instead of continually decreasing as we expected, the work stress and negative emotions gradually increased on Saturdays and peaked on Sundays. As such, it appears that stress and negative emotion are lowest at the start of the weekend (or perhaps in anticipation of the weekend—i.e. Friday), but actually increase gradually dur-ing the course of the weekend. Recall that these are dips in Tweet percentage, and do not reflect overall dips in Tweet activity.

The trends of positive emotions on Twitter were slightly different (Figure 3). Positive emotions were the lowest in the midweek (i.e. Tuesdays, Wed-nesdays, and Thursdays), partially supporting our Effort-Recovery predic-tions. Then positive emotions gradually increased through the weekend (Fridays through Sundays), which supported our expectations from the Effort-Recovery model.

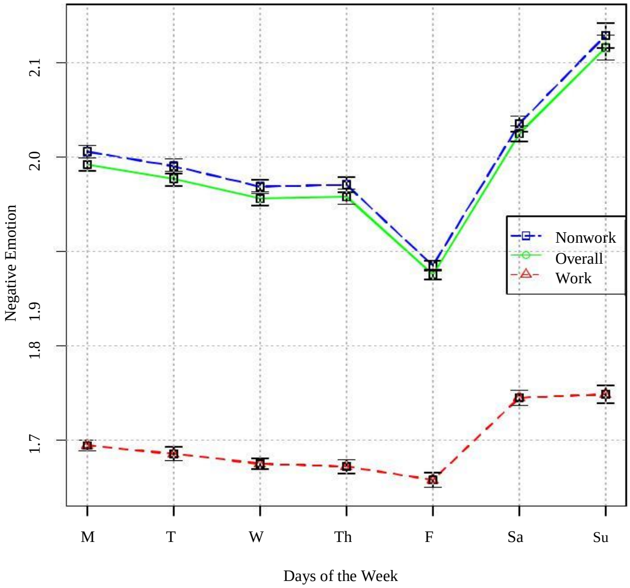

FIGURE 2. Percentage of Tweet Words about Negative Emotion, by Day of the Week. Error bars are added for comparison purposes. Vertical axis represents the composite of word frequencies.

We also separately analyzed the trends for work-related Tweets, nonwork-related Tweets, and overall Tweets (Figures 1 through 3). These trends were similar in each of the stress, negative emotion, and positive emotion dictionary measures, showing that work vs. nonwork Tweet status did not appear to have any effect on the weekly trends. The one key difference was that the mean percentage frequencies of stress, negative emotion, and positive emotion were uniformly lower in the work-related Tweets than in the nonwork-related Tweets. This suggests that work-related Tweets by and large tended to be less focused on emotion and stress words (in terms of the percentage of emotion or stress words in each Tweet), in comparison to nonwork Tweets.

Finally, health issue-related words (e.g. ache*, itch*, pain, sick, drug, etc.) on Twitter present a different trend from the trends for stress, negative emotion, and positive emotion (see Figure 4). First of all, the overall per-centage of health problem words tended to be higher during the workweek and lower on weekends (Friday through Sunday), consistent with the Effort-Recovery model and weekend recovery hypothesis. Second, words for health issues had a much higher frequency in work-related Tweets than in nonwork-related Tweets. This indicates that health problem words were more common in work-related Tweets (people tended to more likely

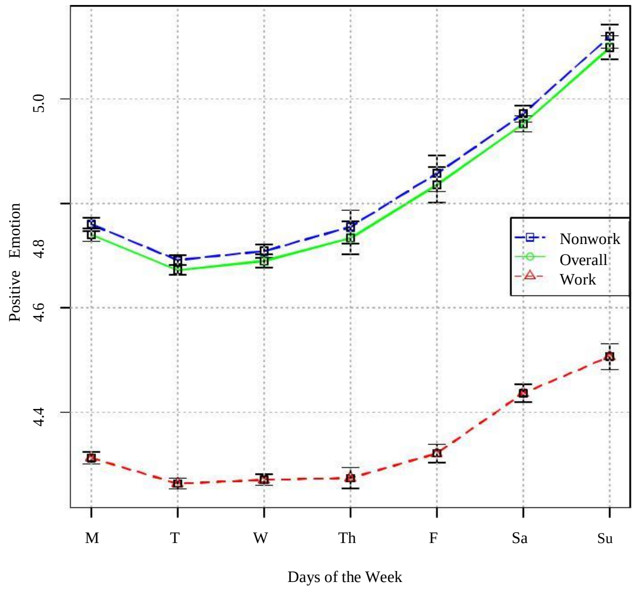

FIGURE 3. Percentage of Tweet Words about Positive Emotion, by Day of the Week. Error bars are added for a comparison purposes. Vertical axis represents the composite of word frequencies.

mention health problem words in work-related Tweets than in nonwork-related Tweets). Further, for work-related Tweets, health problem words unexpectedly showed two peaks: midweek (i.e. Tuesdays through Thurs-days) and weekends (Saturdays and Sundays), with two corresponding dips (on Mondays and Fridays). It is interesting that the frequency of health problem words increased on weekends in work-related Tweets (e.g. some of these Tweets might have involved working on the weekends), but decreased on weekends in nonwork-related Tweets.

The trends at the combined factor level (for the factors revealed by the dynamic factor analysis) are presented in the Appendix. The factor-level trends showed similar patterns to the individual dictionary patterns. For example, the patterns of Factor 1 (negative emotions, stress, somatic) in Figure A were simi-lar to the patterns of stress in Figure 1 and negative emotion in Figure 2 (i.e. showing a “Friday dip”). The patterns of Factor 2A (positive emotion, food, friend, leisure, home, and family) in Figure A were similar to the patterns of positive emotions in Figure 3 (i.e. showing a “Midweek dip” and “Weekend peak”). Finally, the overall pattern for Factor 2B (work, money, achievement, health) was similar to the overall pattern for health issue Tweets shown in Figure 4 (i.e. showing a “Weekend dip”).

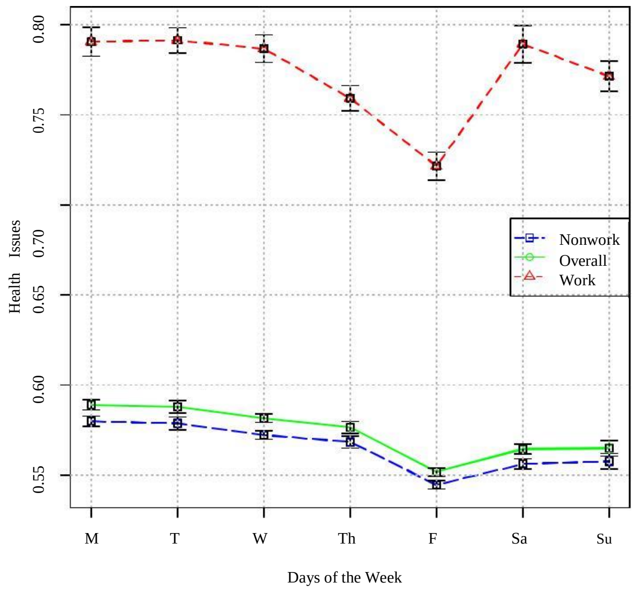

FIGURE 4. Percentage of Tweet Words about Health Problems, by Day of the Week. Error bars are added for a comparison purposes. Vertical axis represents the composite of word frequencies.

GENERAL DISCUSSION

Historically, innovative data collection methods have held promise for advancing psychological research paradigms and meaningfully improving our knowledge of human beings (Miller, 2012). Psychologists have witnessed these phenomena with methods such as self-report survey research, reaction time, electroencephalogram (EEG), functional magnetic resonance imaging (fMRI), and other methodological innovations. Now it is big data. We expect that the application of big data in psychological research will also open doors to new research questions and new answers to conventional questions, increasing our knowledge of phenomena. In this paper, we have described the concept and characteristics of big data/Twitter analysis, as well as its potential for organisational health research. We also developed a stress word count dictionary for text analysis, and attempted to assess work-related stress using Twitter mes-sages. We further demonstrated, by analyzing more than 2.1 billion Twitter messages, how big data might provide new insights into research questions that are difficult for traditional data collection methods to address.

We constructed a stress word count dictionary by following Pennebaker et al. (2007) procedures, with an attempt to derive an alternative stress mea-sure from unobtrusive text analysis that is different from the conventional self-report method. Our stress dictionary measure showed high reliability and desirable convergent validity and discriminant validity. In addition, when fac-tor analyzing the stress dictionary alongside 18 traditional LIWC dictionary measures, the stress dictionary measure strongly loaded onto the negative emo-tion factor, as expected. These results tend to support the psychometric proper-ties of the constructed stress dictionary.

As stress is central to health psychology, work stress is of particular interest to organisational psychologists. In the current study, we segmented the Tweet messages into work-related and nonwork-related Tweets, and then analyzed the two groups separately. The results have shown that the segmentation of work-(vs. nonwork-)related Tweets appears to be meaningful, in some instances. We found that the people expressed a lower rate of emotion Tweets (including posi-tive emotion, negative emotion, and stress Tweets), but described a higher rate of health problem Tweets, in work Tweets than in nonwork Tweets. In other words, Tweets that contained the words “job” or “work” were less likely to talk about emotion or stress, but were more likely to talk about health issues (in terms of word count frequency). As a caveat, we emphasise that these results only describe the aggregation (average) of individual Tweets to the national level; analyzing individual users Tweets might reveal a different pattern.

The weekly trends of work stress and emotion from Twitter clearly display a Friday effect, which is similar to Ingraham s (2014) findings based on Google search queries. For the negative emotions, or “misery” index in Ingraham s (2014) report, Friday is a turning point. Misery decreases as Friday approaches. After Friday, misery starts to increase from Saturday to Sunday. Although the general trends found in Twitter are similar to the results from Google search queries, the results are also slightly different. For example, Goo-gle displays the lowest negative emotions on Saturdays, whereas the dip occurs on Fridays in Twitter. This difference suggests that people may behave differ-ently on different social media platforms (e.g. Twitter, Google, Facebook, etc.).

Most of the findings supported the Effort-Recovery theory (Meijman & Mulder, 1998) and could be interpreted as consistent with weekends recov-ery effects as found by Fritz and Sonnentag (2005)—at least to the extent that stress dips down on Fridays (at the start of the weekend) for both work and nonwork Tweets, health issues dip during the weekend (Friday, Saturday, and Sunday) in nonwork Tweets, and positive emotion increases over the weekend. However, some results—for example, the increases of stress and negative emotion Tweets during the weekend and the increases of health issues in work-related Tweets during the weekend—were poten-tially inconsistent with the Effort-Recovery theory s prediction. It is possi-ble that the weekend spike in health problem Tweets within the set of work Tweets might reflect health issues caused by working during the weekends, but this possibility requires further research. We also highlight that our focus in the current analyses was on the aggregate (national) level of analysis, in contrast to the Effort-Recovery model, which was mainly pro-posed to describe phenomena at the individual person level. At the aggre-gate level, we found an increase of both positive emotion and negative emotion over the weekends, suggesting that—at the national level—people tend to be more emotionally expressive on weekends/Sundays, in terms of both negative emotion and positive emotion Tweets.

In addition, we believe that our findings of weekly trends in job stress and emotions based on the Tweets from the United States might be generalisable to other countries as well. Indeed, a recent Science paper by Golder and Macy (2011) analyzing the diurnal trends in general negative and positive emotions on Twitter across six countries and regions (the USA, Canada, the UK, Aus-tralia, India, and Africa) found that the emotional trends were generally simi-lar across the six countries.

Limitations of Twitter Analysis

Big data and Twitter analysis open new doors for psychological research. As this new approach is expected to bring new insights to conventional psycholog-ical research questions, we predict that big data research will increasingly be adopted by psychological researchers. Yet Twitter analysis is not a perfect approach, and it comes with limitations. One major limitation is that we have only sampled individuals with Twitter accounts. With regard to demographic characteristics, over 60 per cent of Twitter users are under age 35, which may partly undermine the representativeness of older populations in these results. A second major limitation is that we can only assess content that individuals are choosing to make publicly available, which implicates social desirability bias and selective disclosure of information. A third major limitation of Twitter data is its unstructured nature. Unlike conventional laboratory or survey research, big data are not specifically generated for research purposes and have an unorganised structure, which means that most of the constructs are not directly measured. Instead, researchers must indirectly derive the measures, which can pose challenges for interpretation. As one example, the Tweets, “I feel sexy today” and “no one at work said I look sexy today” would be com-bined in Twitter analysis, because they both reflect the thematic content “sexual”. For this reason, it is important for readers to keep in mind that the P-technique factor analytic results in Table 1 are not at the individual person level of analysis, but rather only suggest—at the US national level of analysis— thematic content that tended to appear on the same days.

Also, big data are often naturalistically collected from the field. When com-paring results from big data to the previous results discovered in the laboratory, we may expect a discrepancy in terms of effect size. Indeed, in a recent meta-analysis that compared 203 lab-field pairs of effect sizes, Vanhove and Harms (2015) revealed that effects for field research are routinely smaller than for laboratory research on the same topics. Such a discrepancy in effect size may not necessarily falsify the existing theory. Similarly, we maintain that big data research might reveal similar patterns to laboratory and field research, although perhaps with a relatively smaller effect size than found in lab research, but this would nonetheless provide converging evidence in support of the exist-ing theory (but on the basis of a much, much larger database).

Future Research Directions

Future research applying big data to the study of psychology may focus its analysis at various levels. At the individual level, researchers may address the relationship between individual differences (e.g. personality, education, etc.) and various attitudes and behaviors (e.g. thoughts, emotions, life activities, etc.) that are expressed on social media. At the collective level, one may aggre-gate information to the city, county, state, or even national levels. For example, Hernandez, Newman, and Jeon (2015) analyzed job satisfaction from Twitter at the city level. Golder and Macy (2011) investigated hourly changes of posi-tive and negative affect on Twitter for 84 countries. Because social media data often contain location and real-time information, we would also expect loca-tion and time-based research—e.g. on time trends and regional variation in responses to national events—to become more popular in the future.

In the current study, we used Twitter to explore daily patterns of work stress and emotion. However, researchers can examine many other topics in the work and organisational health domain using this approach. For example, research-ers may study work safety and work accidents by tracking Tweets, and develop a monitoring system similar to Google Flu Tracking or Twitter influenza surveillance. Combining with precise time and location information, such as workplace safety and accident surveillance, might be useful in many ways, including tracking physical health and subjective well-being. In addition, Twitter might serve as a useful tool for studying workplace aggression, as victims might choose to be silent at the workplace but post such information on Twitter.

As discussed above, detecting and identifying work vs. nonwork Tweets or messages can better inform work and organisational health research. Future research is needed to develop computer algorithms to more accurately desig-nate work vs. nonwork Tweets. We also need further work to continue validat-ing the stress dictionary. Although we have examined its convergent validity and discriminant validity, more work is needed to investigate its criterion valid-ity and predictive validity.

In summary, big data approaches like Twitter analysis show exciting poten-tial for health research, and their application in workplace and organisational research is promising. It is our hope that this paper will bring attention to this methodology, and that we will witness a growing research interest in applying big data techniques to the study of work and organisational health.

REFERENCES

Almeida, D.M., & McDonald, D. (1998). Weekly rhythms of parents work stress, home stress, and parent–adolescent tension. New Directions for Child and Adolescent Development, 82, 53–68. doi: 10.1002/cd.23219988205

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). Standards for educational and psychological testing. Washington, DC: AERA.

Campbell, D.T., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. Psychological Bulletin, 56, 81–105.

Cattell, R.B. (1952). The three basic factor analytic research designs: Their interrela-tions and derivatives. Psychological Bulletin, 49, 499–520.

Centers for Disease Control and Prevention (2011). Heart disease and stroke preven-tion: Addressing the nation s leading killers. Retrieved 30 March 2015 from: http://www.cdc.gov/chronicdisease/resources/publications/aag/pdf/2011/heart-disease-and-stroke-aag-2011.pdf

Clay, R.A. (2011). Is stress getting to you? Monitor on Psychology, 42(1), 58–63. Cohen,

S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. Journal of Health and Social Behavior, 24, 385–396.

Culotta, A. (2010). Detecting influenza epidemics by analyzing twitter messages. arXiv:1007.4748v1 [cs.IR]

Dumbrell, D., & Steele, R. (2013). Twitter and health in the Australian context: What type of information are health-related organizations tweeting? Proceedings of the 46th Hawaii International Conference on System Sciences.

Fritz, C., & Sonnentag, S. (2005). Recovery, health, and job performance: Effects of weekend experiences. Journal of Occupational Health Psychology, 10, 187–199.

George, G., Hass, M.R., & Pentland, A. (2014). Big data and management. Academy of Management Journal, 57, 321–326.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457, 1012–1014.

Golder, S.A., & Macy, M.W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science, 333, 1878–1881.

Hao, B., Li, L., Gao, R., Li, A., & Zhu, T. (2014). Sensing subjective well-being from social media. In D. Slezak, G. Schaefer, S.T. Vuong, & Y.-S. Kim (Eds.), Active media technology (pp. 324–335). Heidelberg: Springer. arXiv preprint arXiv:1403.3807.

Hawn, C. (2009). Take two Aspirin and Tweet me in the morning: How Twitter, Face-book, and other social media are reshaping health care. Health Affairs, 28, 361–368.

Heaivilin, N., Gerbert, B., Page, J.E., & Gibbs, J.L. (2011). Public health surveillance of dental pain via Twitter. Journal of Dental Research, 90(9), 1047–1051. doi: 10.1177/0022034511415273

Health and Safety Executive (2008). Work related stress. London: Author.

Hernandez, I., Newman, D.A., & Jeon, G. (2015). Twitter analysis: Methods for data management and a word count dictionary to measure city-level job satisfaction. In S. Tonidandel, E. King, & J. Cortina (Eds.), Big data at work: The data science revolution and organizational psychology (ch. 4). New York: Routledge.

Holmes, E.E., Ward, E.J., & Scheuerell, M.D. (2014). Analysis of multivariate time-series using the MARSS package. User guide: ftp://expo.lcs.mit.edu/pub/CRAN/web/packages/MARSS/vignettes/UserGuide.pdf

Ingraham, C. (2014, 3 December). The Google misery index: The times of year we re most depressed, anxious and stressed. The Washington Post. Retrieved from: http://www.washingtonpost.com/blogs/wonkblog/wp/2014/12/03/the-google-misery-index-the-times-of-year-were-most-depressed-anxious-and-stressed/

Intel (2012). Intel peer research on big data analysis (No. 327825-001). Retrieved from: www.intel.com/content/dam/www/public/us/en/documents/reports/data-insights-peer-research-report.pdf

Jones, C.J., & Nesselroade, J.R. (1990). Multivariate, replicated, single-subject, repeated measures designs and p-technique factor analysis: A review of intraindi-vidual change studies. Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process, 16, 171–183.

Kramer, A.D., Guillory, J.E., & Hancock, J.T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, USA, 111, 8788–8790.

Krantz, D.S., & McCeney, M.K. (2002). Effects of psychological social factors on organic disease: A critical reassessment of research on coronary heart disease. Annual Review of Psychology, 53, 341–369.

Lampos, V., & Cristianini, N. (2010). Tracking the flu pandemic by monitoring the social web. IAPR 2nd Workshop on Cognitive Information Processing (CIP 2010).

Laney, D. (2001). 3D data management: Controlling data volume, velocity and vari-ety. META Group Research Note, 6.

Larsen, R.J., & Kasimatis, M. (1990). Individual differences in entrainment of mood to the weekly calendar. Journal of Personality and Social Psychology, 58, 164–171.

Lazarus, R.S. (1966). Psychological stress and the coping process. New York: McGraw-Hill.

Lohr, S. (2013, 1 February). The origins of “big data”: An etymological detective story. New York Times. Retrieved from: http://bits.blogs.nytimes.com/2013/02/01/ the-origins-of-big-data-an-etymological-detective-story/

Lee, I., & Little, T. (2012). P-technique factor analysis. In B. Laursen, T. Little, & N. Card (Eds.), Handbook of developmental research methods (pp. 350–363). New York: Guilford Press.

Meijman, T.F., & Mulder, G. (1998). Psychological aspects of workload. In P.J.D. Drenth & H. Thierry (Eds.), Handbook of work and organizational psychology (Vol. 2, pp. 5–33). Hove: Psychology Press.

Microsoft (2013, 11 February). The Big Bang: How the big data explosion is chang-ing the world. Retrieved from: http://news.microsoft.com/2013/02/11/the-big-bang-how-the-big-data-explosion-is-changing-the-world/

Miller, G. (2012). The smartphone psychology manifesto. Perspectives on Psychological Science, 7, 221–237.

Neiger, B.L., Thackeray, R., Burton, S.H., Giraud-Carrier, C.G., & Fagen, M.C. (2013). Evaluating social media s capacity to develop engaged audiences in health promotion settings: Use of Twitter metrics as a case study. Health Promotion Practice, 14, 157–162. doi: 10.1177/1524839912469378

Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhang, Z., Gao, Y., Soltani, K., & Liu, Y. (2014). FluMapper: A cyberGIS application for interactive analysis of massive location-based social media. Concurrency and Computation: Practice and Experience, 26, 2253–2265.

Park, H., Rodgers, S., & Stemmle, J. (2013). Analyzing health organizations use of twitter for promoting health literacy. Journal of Health Communication: International Perspectives, 18, 410–425. doi: 10.1080/10810730.2012.727956

Paul, M.J., & Dredze, M. (2011a). A model for mining public health topics from twitter. Technical report, Johns Hopkins University.

Paul, M.J., & Dredze, M. (2011b). You are what you tweet: Analyzing Twitter for public health. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 265–272.

Paul, S.A. Hong, L., & Chi, E.H. (2011). Is Twitter a good place for asking questions? A characterization study. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 578–581.

Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., & Booth, R.J. (2007). The development and psychometric properties of LIWC2007. Austin, TX: LIWC.net.

Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). Linguistic inquiry and word count: LIWC 2001. Mahwah, NJ: Lawrence Erlbaum Associates.

Price, M. (2011). The risks of night work. Monitor on Psychology, 42(1), 38–41. Prier,
K.W., Smith, M.S., Giraud-Carrier, C., & Hanson, C.L. (2011). Identifying health related topics in Twitter: An exploration of tobacco related Tweets as a test topic. Proceedings of the 4th International Conference on Social Computing, Behavioral Cultural Modeling and Prediction, 18–25.

Prochaska, J.J., Pechmann, C., Kim, R., & Leonhardt, J.M. (2012). Twitter5quitter? An analysis of Twitter quit smoking social networks. Tobacco Control, 21, 447–449. doi:10.1136/tc.2010.042507

Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Lucas, R.E., Agrawal, M., et al. (2013). Characterizing geographic variation in well-being using Tweets. Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, 583–591.

Smith, C.S., Folkard, S., Tucker, P., & Evans, M.S. (2011). Work schedules, health, and safety. In J.C. Cambell & L.E. Tetrick (Eds.), Occupational health psychology (2nd edn., pp 185–204). Washington, DC: American Psychological Association.

Spector, P.E., & Fox, S. (2005). A model of counterproductive work behavior. In S. Fox & P.E. Spector (Eds.), Counterproductive workplace behavior: Investigations of actors and targets (pp. 151–174). Washington, DC: APA.

Sullivan, S.J., Schneiders, A.G., Cheang, C., Kitto, E., Lee, H., Redhead, J., Ward, S., Ahmed, O.H., & McCrory, P.R. (2012). What s happening? A content analysis of concussion-related traffic on Twitter. British Journal of Sports Medicine, 46, 258–263. doi: 258 10.1136/bjsm.2010.080341

Tonidandel, S., King, E., & Cortina, J. (2015) Big data at work: The data science revolution and organizational psychology. New York: Routledge.

Totterdell, P. (2005). Work schedules. In J. Barling, E.K. Kelloway, & M.R. Frone (Eds.), Handbook of work stress (pp. 53–81). Thousand Oaks, CA: Sage.

Vanhove, A., & Harms, P.D. (2015). Reconciling the two disciplines of organizational science: The comparability of findings from lab and field research. Applied Psychology: An International Review, 64, 637–673.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063–1070.

APPENDIX

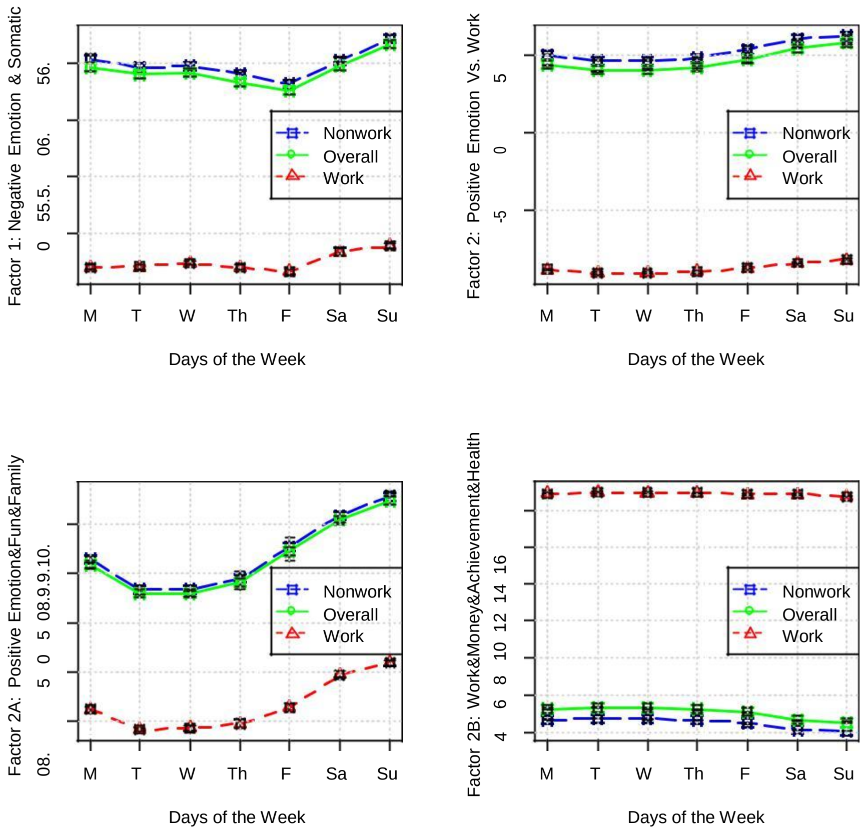

FIGURE A. Percentage of Tweet words about Factor 1 (Negative Emotion/ Stress/Somatic), Factor 2 (Positive Emotion/Food-Friends-Fun/Home/Leisure MINUS Work/Money/Achievement/Health), Factor 2A (Positive Emotion/Food/ Friends/Fun/Home/Leisure), and Factor 2B (Work/Money/Achievement/Health), by day of the week. Error bars are added for comparison purposes.