This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2015.2403839, IEEE Journal of Biomedical and Health Informatics

We Feel: Mapping emotion on Twitter

Mark E. Larsen, Tjeerd W. Boonstra, Philip J. Batterham, Bridianne O'Dea, Cecile Paris, and Helen Christensen

***Abstract*— Research data on predisposition to mental health problems, and the fluctuations and regulation of emotions, thoughts and behaviours are traditionally collected through surveys, which cannot provide a real-time insight into the emotional state of individuals or communities. Large data sets such as World Health Organization (WHO) statistics are collected less than once per year, whereas social network platforms, such as Twitter, offer the opportunity for real-time analysis of expressed mood. Such patterns are valuable to the mental health research community, to help understand the periods and locations of greatest demand and unmet need.**

We describe the "We Feel" system for analysing global and regional variations in emotional expression, and report the results of validation against known patterns of variation in mood. 2.73×10^{9} emotional tweets were collected over a 12-week period, and automatically annotated for emotion, geographic location and gender. Principal component analysis (PCA) of the data illustrated a dominant in-phase pattern across all emotions, modulated by anti-phase patterns for "positive" and "negative" emotions. The first three principal components accounted for over 90% of the variation in the data. PCA was also used to remove the dominant diurnal and weekly variations allowing identification of significant events within the data, with z-scores showing expression of emotions over 80 standard deviations from the mean. We also correlate emotional expression with WHO data at a national level and although no correlations were observed for the burden of depression, the burden of anxiety and suicide rates appeared to correlate with expression of particular emotions.

***Index Terms*—sentiment analysis, mental health, Twitter**

I. INTRODUCTION

EVERY year, a wealth of information is published by the academic community on mental health and wellbeing, including key research questions on predisposition to mental health problems based on various individual and demographic factors (such as age, gender, area of residence and profession) and the fluctuations and regulation of emotions, thoughts and behaviours. Traditionally these data are collected through surveys, which cannot provide a real-time insight into the emotional state of individuals, communities or the global community. Sources of data such as the World Health Organization (WHO; Geneva, Switzerland) Global Health Estimates [1] offer well-defined, high-quality, qualitative measures of the incidence and burden associated with a range of physical and mental health conditions [2], with depression often the leading cause of burden in developed countries [3, 4]. However, with a sampling frequency of less than once per year, there is a great degree of loss of temporal variation.

Online social network platforms, with their large-scale repositories of user-generated content, can provide unique opportunities to gain insights into the emotional "pulse of the nation" [5], and indeed the global community. Twitter (San Francisco, USA) offers the opportunity for the analysis of expressed mood, and previous studies have shown that geographical, diurnal, weekly and seasonal patterns of positive and negative affect can be observed [6, 7]. Studies have also examined the sentiment expressed on Twitter during specific sporting events [8, 9], and to measure levels of happiness of cities within specific countries [10, 11]. Happiness was found to be correlated with demographic features as well as general well-being [11]. The potential of Twitter to detect depression has also been explored [12]. Further detection of emotional patterns relating specifically to mental health variables are particularly valuable to the mental health research community, to help understand the periods and locations of greatest demand (and unmet need) and to allow effective allocation of resources to meet these needs [13].

"We Feel" [14] is a system for the real-time analysis of global and regional variations in emotional expression through Twitter. The system is currently open to the research community, with data accessible through an API for non-commercial use. We describe the system architecture and data processing algorithms and, because of its potential to provide new and evolving information about the mental health of the community, our aim is to validate the data against known patterns of variation in mood. We attempt to do this in three ways. First, we present information on diurnal variations in mood; secondly we link data with prevalence data from countries globally to determine if patterns are consistent with these rates; and finally, we investigate if "events of significance" are associated with changes in expressed emotion. Applications for this system as a tool to the wider

Manuscript received September 26, 2014. This work was supported in part by the NSW Mental Health Commissioner, Mr John Feneley, by CSIRO, and by the Black Dog Institute. M.E. Larsen is supported by the NHMRC Centre of Research Excellence in Suicide Prevention 1042580, T.W. Boonstra is supported by NWO Fellowship 45110030, P.J. Batterham is supported by NHMRC Fellowship 1035262, and H. Christensen is supported by NHMRC Fellowship 1056964.

M.E. Larsen (mark.larsen@blackdog.org.au), T.W. Boonstra (t.boonstra@unsw.edu.au), B. O'Dea (b.odea@blackdog.org.au) and H. Christensen (h.christensen@blackdog.org.au) are with the Black Dog Institute, Hospital Road, Randwick, NSW 2031, Australia, and the University of New South Wales. T.W. Boonstra is also with the MOVE Research Institute, VU University, Amsterdam, The Netherlands.

P.J. Batterham (philip.batterham@anu.edu.au) is with the National Institute for Mental Health Research, Building 63, The Australian National University, Canberra, ACT 0200, Australia.

C. Paris (cecile.paris@csiro.au) is with CSIRO, Cnr Vimiera and Pembroke Roads, Marsfield, NSW 2122, Australia.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2015.2403839, IEEE Journal of Biomedical and Health Informatics

research and mental health community are then discussed.

II. System Design

A. Data Collection

The We Feel system is built as a tool for analysis, in real-time, of the emotional sentiment expressed on Twitter, an online microblogging service that enables users to post and read short 140-character text messages (tweets). Approximately 19% [15] of online adults use this service. The large majority of Twitter accounts are publicly available, allowing any registered or unregistered user to view posted tweets. This contrasts with other social network platforms which have different privacy settings.

Twitter offers a public API which enables programmatic consumption of published tweets as they occur, filtered by specific criteria such as a list of keywords. The standard Twitter feed provides a 1% random sample of all posted tweets. For development and demonstration purposes, a 10% feed was made available for a period of one month by GNIP (Boulder, USA). The default data source for the We Feel system is a custom stream (the "emohose") created using the Twitter query interface to return all English language tweets containing a custom vocabulary of emotional words. The emohose provides a stream of up to 45,000 tweets per minute.

Due to the volume of data collected, Amazon Web Service (Sydney, Australia) infrastructure is used to process the tweet data. Multiple Elastic Compute Cloud (EC2) instances are used to receive the emohose stream and perform classification based on location, gender and emotional content (see Section II.B) – this allows the processing infrastructure to easily rescale. The summary data are then stored in DynamoDB tables for access from the web interface and API.

B. Data Classification

To identify potential emotional tweets, a large vocabulary of emotion terms was compiled from multiple sources, including the Affective Norms for English Words (ANEW) [16] and the Linguistic Inquiry and Word Count (LIWC) [17]. ANEW provides a set of normative emotional ratings for a set of 1034 English words, and LIWC is text analysis software that calculates the degree to which people use different categories of words across a wide array of texts. Validation studies reveal that LIWC successfully measures positive and negative emotions [18]. A crowdsourcing task was conducted to organise the combined emotional vocabulary against Parrott's hierarchy of emotions [19] and Shaver *et al.'s* emotion tree [20]. Each emotional word was classified into six "primary emotion" categories of love, joy, surprise, anger, sadness and fear, with 25 subgroups of "secondary emotions" – see Table I for a sample of the hierarchical structure.

In addition to the emotional categorisation, the We Feel system attempts to classify tweets based on the location and gender of the user. Twitter allows users to enter a location within their profile, however determining a physical location from this free-text entry is challenging. Also, Twitter allows a user to geotag their tweets with a precise location, however this is not commonly used. To maximise the availability of location data, the time zone specified in a user profile was selected as a proxy marker for their physical location. The spatial resolution of this method varies considerably with some countries spanning multiple time zones and, similarly, some time zones spanning multiple countries.

The gender of a tweet poster was estimated by comparing the username of an individual's profile with a large list of male and female names obtained from the BabyNameMap project [21]. However, as it was only possible to categorise approximately one quarter of names, further analysis based on gender was not pursued.

C. Data Visualisation

In addition to the We Feel data API, the web interface provides a range of data visualisation options. An interactive map allows an area of interest to be selected from the global, continental or country/time zone level. Time series and circumplex models of emotional tweet counts, refined by geography, time interval, and emotion of interest, can be plotted, as shown in Fig. 1. A circumplex model projects emotional words onto orthogonal axes of valence (pleasantness of a stimulus) and arousal (the strength of the response to the stimulus). A third dimension, dominance (the degree of control exerted by the stimulus) is not included in the We Feel visualisations. Co-ordinates for the vocabulary of emotional words were obtained from a dataset of affective norms [22].

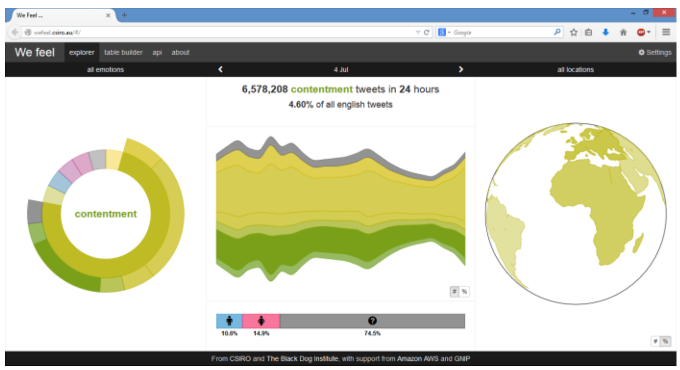

Fig. 1. Screenshot of the We Feel web interface and data visualisations. The circumplex model and emotion filter is on the left hand side, with the time series in the centre, and the geographical area of interest on the right hand side.

D. Ethical Considerations

The data collected by the We Feel system is publically available through the Twitter API. Potentially identifiable information, specifically the tweet content and profile username, are automatically coded for emotional content and gender respectively upon acquisition. Data are analysed in anonymised, aggregated form only. Individual consent from users was therefore not required. This project was approved by the CSIRO Social Science Human Research Ethics Committee.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2015.2403839, IEEE Journal of Biomedical and Health Informatics

III. Results

Data were collected via the We Feel API using the emohose for the period 12 May to 31 August 2014 at 5-minute intervals for all time zones (n=143), primary emotions (n=7) and secondary emotions (n=31). Data for individual tertiary emotions (n=676) were not collected. Data were split into 16 week-long sections, and divided into a training group and testing group for comparison. Due to missing samples at the start of the 1st week (n=11) and at the end of the 16th week (n=1), these periods were allocated to the test groups. Significant events were known to have occurred in the 1st and 14th weeks (see Section III.C), so the 14th and 15th weeks were also assigned to the test group. This resulted in a continuous period of 12 weeks of training data.

In total, 2.73×10^9 tweets were collected and automatically coded over the 12-week training period. The geographic distribution of the source of tweets is skewed with the majority originating in North America (60%), followed by Europe (16%), Asia (11%), South America (7.6%), Africa (2.9%) and Oceania (1.7%). The distribution of the emotional content of tweets is presented in Table I, and shows that the majority were classified as joy (53%), followed by sadness (17%). The other primary emotions were each identified in less than 6% of the tweets.

TABLE I

Categorisation of emotional words into primary, secondary and tertiary categories (all three levels shown for the primary emotion "love" as an illustration, tertiary emotions are not shown for the other primary emotions), and distribution of emotions as a percentage of total tweets captured.

Primary emotion	**Secondary emotion**	**Tertiary emotion**
love (5.8%)	affection (4.0%)	passionate, generous, loved, loving, treasured, romantic, beloved, attracted, faithful, affectionate, compassionate, liked, sweet, accepted, devoted, lovely, blessed, loyal, supporting, kind, considerate, supportive, gentle, gracious, caring, sympathetic, tender, delicate, fond
	lust (1.8%)	naughty, horny, flirty, romantic, attracted, hot, slutty
	longing (<0.1%)	nostalgic
Primary emotion	**Secondary emotions**	
joy (53.0%)	cheerfulness (18.5%), zest (4.9%), contentment (13.3%), pride (6.0%), optimism (3.3%), enthrallment (<0.1%), relief (2.9%)	
surprise (3.6%)	surprise (3.6%)	
anger (4.0%)	irritability (0.7%), exasperation (0.6%), rage (1.6%), disgust (0.5%), envy (0.4%), torment (<0.1%)	
sadness (17.1%)	suffering (0.7%), sadness (3.8%), disappointment (2.2%), shame (4.2%), neglect (1.2%), sympathy (1.3%)	
fear (2.9%)	horror (0.7%), nervousness (2.3%)	
other (13.6%)		

A. Diurnal and Weekly Variations

We investigated the pattern of emotional tweets from the Eastern Standard Time (EST) time zone in more detail. This region comprises of the east coast of the USA and Canada, and generates the largest number of tweets of all time zones with 17% of total number of tweets recorded. Fig. 2(a) shows a breakdown of the secondary emotions within this region, showing the most common emotions classified are cheerfulness followed by contentment. The diurnal pattern was assessed by averaging the number of tweets across the 84-day period at the 5-minute sampling period (Fig. 2(b)). The profile shows a minimum at approximately 4am, rising to a constant level during office hours, followed by an evening peak at approximately 10pm, and then descending to the following morning's minimum. This agrees with previous reports [6, Supporting Information]. Aggregating the number of emotional tweets on a daily basis (Fig. 2(c)) shows a reduction in the number of tweets on Saturday and Sunday compared with the working week.

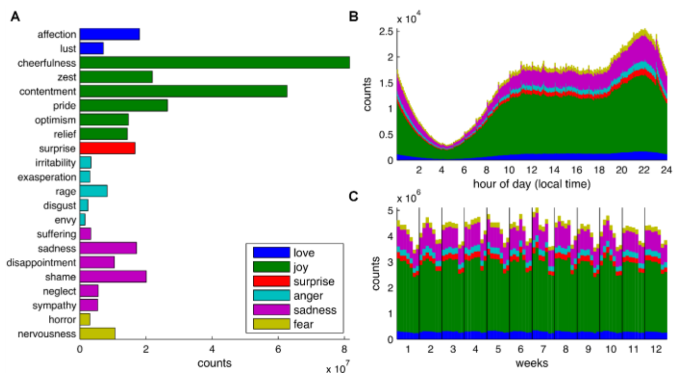

Fig. 2. Number of emotional tweets from the Eastern Standard Time zone (USA and Canada). Panel (a) shows the total number of secondary emotion tweets during the 12-week training period, (b) shows the fluctuation in primary emotions over an average day, and (c) shows the day-to-day fluctuation in the number of primary emotion tweets over the 12-week period.

We used principal component analysis (PCA) as a data-driven approach to examine correlations between the temporal fluctuations of different emotions. PCA is a multivariate statistical procedure that decomposes a set of observations into a set of linearly uncorrelated variables, or principal components (PCs) [23]. The principal components are ordered based on the degree of variation in the data which they account for, and hence PCA is often used as an unbiased means for reducing high-dimensional data sets to a small number of components. Applying PCA to the training data, the first components will capture common repeating diurnal patterns in the data, while the higher components will capture uncorrelated fluctuations in emotions.

We performed principal component analysis on the time series count data to characterise the diurnal weekly pattern in the fluctuation of emotional content. As an illustrative example, we examine three geographic areas: the east coast of the USA and Canada (EST), the United Kingdom, and the east coast of Australia (combining the Brisbane, Melbourne and Sydney time zones). The time series across regions are concatenated, and decomposed into principal components and the corresponding weightings.

Fig. 3 shows the first three principal components, which together explain over 90% of the variance (λ) of the original

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2015.2403839, IEEE Journal of Biomedical and Health Informatics

time series data. The first PC explains λ=87% of the variance and captures the temporal pattern common across all secondary emotions, as shown by the consistent positive weightings. The total number of tweets is at a minimum at approximately 4 am and peaks at approximately 10 pm, closely mirroring the average pattern in Fig. 2. The second PC (λ=2.5%) reflects a modulation of the first PC with a clear separation in the loading of "positive" (love, joy and surprise) and "negative" (anger, sadness and fear) emotions. The PC reveals diurnal fluctuations that are similar across the week showing a peak during the morning between 8-10 am (more positive and fewer negative emotions) and declines steadily during the day to a minimum in the evening around 11 pm (more negative and fewer positive emotions). This pattern is very similar to fluctuations observed in the average happiness level obtained from the 'Hedonometer' [7]. The third PC (λ=1.5%) reveals a distinction between secondary emotions that does not follow the positive/negative primary emotion pattern of the second PC. Positive weighting is observed for lust, disgust, horror, envy and nervousness which peak towards the end of the evening and overnight. We use the results of this PCA in Section III.C to detect events within the time series data.

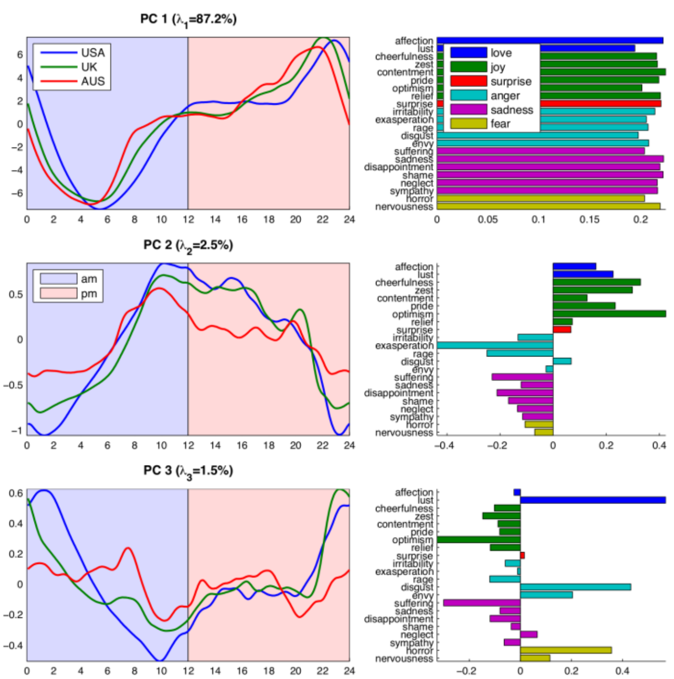

Fig. 3. Principal component analysis of secondary emotion time series of 12 week training data from three geographic regions (east coast USA and Canada, the UK, and east coast Australia). The first three principal components (rows) explain 87%, 2.5% and 1.5% of the total variance (λ) respectively. The PCs are averaged across the 84 days (left column) and the weightings (right column) show the contribution of each secondary emotion to this PC.

B. *Indices of Mental Health*

The previous section illustrated temporal variations in emotional tweets which are consistent with external reports. To investigate whether emotional tweets correlate with indices of mental health in the population, we examined the correlation between rates of secondary emotion tweets with country-level data on the burden of depression and anxiety, and national suicide incidence rates. Mental health data were obtained from WHO Global Health Estimates 2014 Summary Tables [1], which consist of disability adjusted life year estimates for 172 member states. From these data, we extracted the percentage of disease burden attributable to unipolar depressive disorders and to anxiety disorders. Suicide rates were also based on WHO statistics [24].

Due to the irregular mapping between time zones and countries, only 93 countries could be directly identified from the time zone data, of which 83 also had WHO data available. Tweets and mental health data from the USA and Canada were combined as the time zones for these nations overlapped completely. Correlations between proportions of secondary emotion tweets with the mental health indices within the 83 countries are shown in Table II. None of the secondary emotions were significantly associated with depression burden. Countries with a larger proportion of tweets expressing envy, sadness or shame had significantly higher burdens of anxiety. Countries with lower proportions of affection, rage or neglect tweets, or higher proportions of cheerfulness tweets had significantly higher suicide rates.

TABLE II
CORRELATIONS (PEARSON'S CORRELATION COEFFICIENT, R) BETWEEN WHO MENTAL HEALTH INDICES AND SECONDARY EMOTION TWEETS ACROSS 83 COUNTRIES. * DENOTES P<0.05.

	Depression burden	Anxiety burden	Suicide rate
Affection	0.049	0.057	**-0.235***
Lust	-0.131	-0.112	0.024
Longing	-0.082	-0.028	0.043
Cheerfulness	0.034	0.032	**0.267***
Zest	-0.066	-0.002	0.005
Contentment	0.120	0.111	0.084
Pride	0.031	0.009	0.013
Optimism	-0.145	-0.218	0.021
Enthrallment	-0.151	-0.125	0.005
Relief	0.131	0.147	-0.210
Surprise	0.056	0.109	-0.045
Irritability	0.009	0.057	-0.146
Exasperation	-0.121	-0.087	-0.030
Rage	0.015	0.114	**-0.337***
Disgust	-0.114	-0.096	0.041
Envy	0.190	**0.231***	-0.158
Torment	0.057	0.097	-0.154
Suffering	0.008	0.073	-0.166
Sadness	0.160	**0.249***	-0.038
Disappointment	0.191	0.208	-0.183
Shame	0.166	**0.259***	-0.119
Neglect	0.145	0.157	**-0.252***
Sympathy	0.075	0.077	-0.011
Horror	0.151	0.204	-0.069
Nervousness	0.073	0.127	-0.110

To correct for multiple comparisons, we adjusted the significance threshold using a false discovery (FDR) approach that accounts for the number of comparisons and includes a correction for the correlation between variables [25]. The corrected threshold is $p<0.0047$ and after applying this correction only the correlation between rage and suicide rate remained significant ($p=0.0016$). Although the correlation coefficients are relatively low, they show the potential for identifying burden or incidence of mental health problems from emotion expressed on Twitter at a regional or national level.

C. *Event Detection*

Having examined the usual diurnal and weekly patterns, and possible correlations with mental health indices, we examined the feasibility of characterising changes in expressed content in response to significant events. The approximately three-week period of testing data contained two known *a priori* significant events, which would be expected to cause an observable effect: on 13 May 2014 the Australian Federal budget was read, and on 11 August 2014 the death of actor Robin Williams was announced.

To examine the emotional response to these events, the PCA from Section III.A was performed using only the primary emotion data, and the typical diurnal and weekly fluctuations were recalculated. The variation captured by the first three principal components is illustrated in Fig. 4(a), showing that the PCA reconstruction provides a good representation of the training data. The first three PCs captured the repeating diurnal variations, while responses to individual events were captured by the higher components. To characterise these responses, we calculated the residuals after removing these three PCs and computed the mean and standard deviations to normalise the residuals to z-scores (Fig. 4(b)). The normalised residuals for the weeks with significant events were calculated across the EST, UK and Australian areas of interest (Fig. 4(c) and (d)). An increase in the expression of all emotions is observed within Australia at the reading of the budget, which is dominated by large deviations from the normal patterns with z-scores for anger exceeding 20 standard deviations from the mean. A global response is observed following the death of Mr Williams, with z-scores for sadness exceeding 80. Although not illustrated, the remaining two weeks of test data demonstrated isolated spikes in fear, peaking at a z-score of approximately 10.

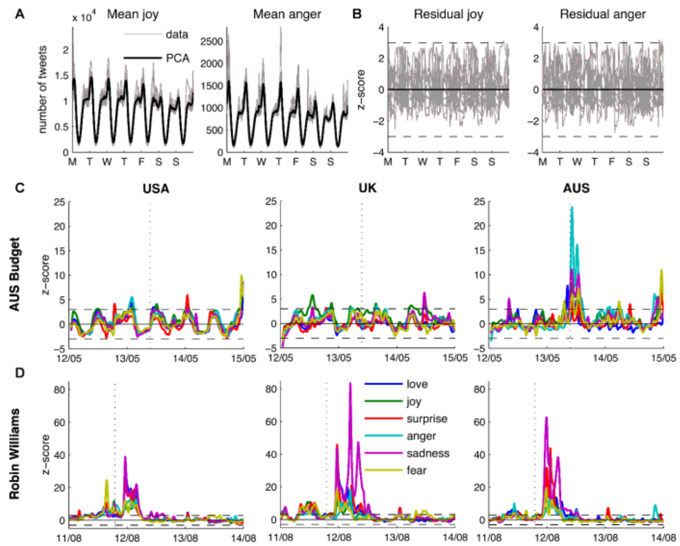

Fig. 4. Effect of significant events on the emotional response on Twitter. Panel (a) shows the PCA reconstruction for joy and anger within the EST time zone, superimposed over the 12 weeks of training data. Panel (b) shows the normalised residual between the training data and the PCA reconstruction within the training data. Panel (c) shows the normalised residual for the week of the reading of the Australian Federal Budget, and (d) for the week of Mr Williams' death. The announcements of these events are shown as vertical dashed lines, with horizontal dashed lines showing a z-score of 3 standard deviations from the mean.

IV. Discussion

We analysed 12 weeks of emotional tweets to investigate whether information captured by the We Feel system can be used to reflect global emotional expression on Twitter. We Feel is a data processing system for analysing global variations in emotional expression through Twitter and provides information in near real-time that is available to the research community through a web interface or API for non-commercial purposes. Over the 12-week training period, the system analysed an average of 22,000 tweets per minute, generating a large-scale, global dataset available to researchers.

The data collected to date has a number of limitations which may impact on the analyses. Firstly, the period of data collection has been relatively short, meaning it is not possible to identify seasonal patterns and changes in baseline. With a longer period of data collection, patterns associated with annual events on fixed dates (such as July 4th in the USA) could be incorporated. Classification errors are an inevitable challenge in sentiment analysis, and it has not been possible here to quantify the accuracy of the coding algorithms. The reliability of individual emotional tweets is therefore unknown, but the aggregate of very large numbers of these data has the potential to give reliable results.

With these considerations in mind, we first examined the emotional content of the tweets and performed PCA to characterise the diurnal and weekly fluctuations in secondary emotions in three geographic regions of the eastern USA and Canada, the UK, and eastern Australia. The principal components show a dominant pattern across all emotions, with a second PC with anti-phase components for "positive" and "negative" emotions, further modulated by individual weightings for specific emotional terms. The patterns, obtained using a fully data-driven approach, are consistent with the results reported elsewhere.

Identification of these typical variations also allowed retrospective identification of extreme patterns associated with significant events, where values over 80 standard deviations from the mean were observed. Although these results appear highly significant, large z-scores were repeatedly observed in time periods without such major events, therefore additional analysis is required to ensure statistical robustness, such as accounting for non-independence of sequential samples from the time series. Supervised machine learning algorithms for novelty detection are likely to provide better identification, in real-time, of significant events, supported by identification and labelling of events within the training data.

Relationships between emotional tweets and indices of mental health were also investigated, and although no secondary emotions were significantly correlated with depression burden, significant positive correlations between envy, sadness and shame tweets and the burden of anxiety were found – although it is not clear that these are directly associated with the nature of the emotions, thoughts and behaviours associated with anxiety disorders. Suicide rates showed a positive correlation with cheerfulness and negative correlations with affection, rage and neglect. The association with affection is perhaps not surprising, and the association with rage may be influenced by coping strategies, however the relationships with cheerfulness and neglect are unexpected. Interpretation of these results is limited by a number of factors. Firstly, there is a risk of spurious correlations being identified and with 75 tests at the 5% level we would expect to find four potentially significant results by chance alone, compared with the seven identified in Table II. After correcting for multiple comparisons, only the correlation between rage and suicide rate remained significant. The indices of burden due to mental health conditions also suffer from a low degree of variance, and suicide incidence statistics can be difficult to compare across countries, both of which may reduce the ability to detect correlations within the data. Also, the data from the burden of disease study was not captured in the same time period as the tweets, and disease may have changed. The inclusion of only English language tweets in an international comparison may also reduce this power. Future implementations of the system could also incorporate data feeds from other social media services which are prevalent in different settings. It should also be noted that only a small proportion of the population in each country are active Twitter users, at only approximately 0.5% of the population of the USA or UK [26]. This is a small fraction, but represents an appreciable number of individuals within these countries. The usage also varies by country, and thus could be included as a variable in correlation analysis. Nevertheless, the sentiment analysis may capture aspects of emotions connected to suicide that have not previously been examined in longitudinal data sets, and thus indicate potential areas of research.

Despite these possible limitations, there is potential for the We Feel system to detect changes in emotional expression on Twitter. Given the difficulty in associating these transient events with prevalence and burden of mental health problems, further data on localised prevalence, anti-depressant prescriptions, and health service utilisation could be used to provide a better reflection of the dynamics of a community's mental health. Improved geographic localisation of the source of tweets may also improve the analysis. This may be particularly relevant for identification of tweets associated with suicide, to identify or intervene in potential suicide clusters.

Our analysis has so far been restricted to examining primary or secondary emotional terms individually. It is possible that specific tertiary emotional terms, or combinations of terms, may be more informative. Other factors, such as the expression of particular emotions at particular times of the day, may also be relevant, so subgroup analysis of clusters within the data may be beneficial.

V. Conclusions

We have presented an overview of the We Feel system for collecting and classifying emotional tweets on a global scale in real-time. Over a 12-week period, 2.73×10^9 tweets were analysed. A series of analyses were performed to demonstrate potential uses of the data, identifying the typical diurnal and weekly variations in emotional expression and using these to detect significant events. Correlations between emotional tweets and indices of anxiety and suicide were also observed, indicating the potential for the development of social media-based measurements for population mental well-being to complement existing data sets. Improvements to the system and possible directions of future analysis have been suggested with the aim of better understanding the emotional pulse of the global community and how this relates to communities' mental health and resources.

Acknowledgment

We thank Neil Jackson for providing us with the opportunity to work with Amazon Web Services in this project, and GNIP for their support in giving us access to their data feed for a limited period to demonstrate the system. We acknowledge the contributions of Dr Alison Calear, Brian Jin and Dr Stephen Wan, and especially Dr David Milne, as the main designer and developer of the We Feel system.

References

[1] World Health Organization. (2014). *Global Health Estimates 2014 Summary Tables*. Available: www.who.int/healthinfo/global_burden_disease/en/

[2] R. C. Kessler and T. B. Ustun, "The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI)," *International Journal of Methods in Psychiatric Research*, vol. 13, pp. 93-121, 2004.

[3] A. D. Lopez, C. D. Mathers, M. Ezzati, D. T. Jamison, and C. J. Murray, "Global and regional burden of disease and risk factors, 2001:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2015.2403839, IEEE Journal of Biomedical and Health Informatics

systematic analysis of population health data," *Lancet*, vol. 367, pp. 1747-1757, May 27 2006.
[4] A. J. Ferrari, F. J. Charlson, R. E. Norman, S. B. Patten, G. Freedman, C. J. Murray, *et al.*, "Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010," *PLoS Medicine*, vol. 10, p. e1001547, Nov 2013.
[5] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnella, and J. N. Rosenquist. (2010). *Pulse of the Nation: U.S. Mood Throughout the Day inferred from Twitter*. Available: http://www.ccs.neu.edu/home/amislove/twittermood/
[6] S. A. Golder and M. W. Macy, "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, vol. 333, pp. 1878-1881, Sep 30 2011.
[7] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth, "Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter," *PLoS One*, vol. 6, p. e26752, 2011.
[8] A. A. A. Esmin, R. L. de Oliveira Jr., and S. Matwin, "Hierarchical Classification Approach to Emotion Recognition in Twitter", *2012 11th International Conference on Machine Learning and Applications (ICMLA)*, pp. 381-385, Dec 2012.
[9] M. Brooks, J. J. Robinson, M. K. Torkildson, S. Hong, and C. R. Aragon, "Collaborative Visual Analysis of Sentiment in Twitter Events," *Cooperative Design, Visualization, and Engineering: Lecture Notes in Computer Science*, vol. 8683, pp. 1-8, 2014.
[10] L. Allisio, V. Mussa, C. Bosco, V. Patti, and G. Ruffo, "Felicitta: Visualizing and Estimating Happiness in Italian Cities from Geotagged Tweets", *Proceedings of the 1st International Workshop on Emotion and Sentiment in Social and Expressive Media*, pp. 95-106, Dec 2013.
[11] L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth, "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place", *PLoS One*, vol. 8, p. e64417, May 2013.
[12] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media", *International AAAI Conference on Weblogs and Social Media*, June 2013.
[13] M. J. Paul and M. Dredze, "Discovering health topics in social media using topic models," *PLoS One*, vol. 9, p. e103408, 2014.
[14] D. Milne, C. Paris, H. Christensen, P. Batterham, and B. O'Dea. (2014). *The We Feel emotion explorer*. Available: http://wefeel.csiro.au
[15] Pew Internet Research. (2014). *Social Networking Fact Sheet*. Available: http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/
[16] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): Instruction manual and affective ratings," The Center for Research in Psychophysiology, University of Florida Technical Report C-1, 1999.
[17] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: LIWC 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, p. 2001, 2001.
[18] J. H. Kahn, R. M. Tobin, A. E. Massey, and J. A. Anderson, "Measuring emotional expression with the Linguistic Inquiry and Word Count," *American Journal of Psychology*, vol. 120, pp. 263-286, Summer 2007.
[19] W. G. Parrott, Emotions in social psychology: Essential readings: Psychology Press, 2001.
[20] P. Shaver, J. Schwartz, D. Kirson, and C. O'Connor, "Emotion knowledge: further exploration of a prototype approach," *Journal of Personality and Social Psychology*, vol. 52, pp. 1061-1086, Jun 1987.
[21] G. Davis. BabyNameMap [Online]. Available: https://github.com/guydavis/babynamemap
[22] A. B. Warriner, V. Kuperman, and M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behavior Research Methods*, vol. 45, pp. 1191-1207, Dec 2013.
[23] I. Jolliffe, *Principal component analysis*: Wiley Online Library, 2005.
[24] World Health Organization. (2004). *Suicide rates by country, year, and sex*. Available: http://www.who.int/mental_health/prevention/suicide/en/Figures_web0604_table.pdf
[25] A. J. Sankoh, M. F. Huque, and S. D. Dubey, "Some comments on frequently used multiple endpoint adjustment methods in clinical trials," *Statistics in Medicine*, vol. 16, pp. 2529-2542, Nov 1997.
[26] D. Mocanu, A. Baronchelli, N. Perra, B. Gonçalves, Q. Zhang, and A. Vespignani, "The Twitter of Babel: Mapping World Languages through Microblogging Platforms," *PLoS One*, vol. 8, p. e61981, Apr 2013.