

Seasonal Fluctuations in Collective Mood Revealed by Wikipedia Searches and Twitter Posts

Fabon Dzogang, Thomas Lansdall-Welfare and Nello Cristianini
Intelligent Systems Laboratory
University of Bristol
Bristol
United Kingdom
Email: firstname.lastname@bristol.ac.uk

Abstract—Understanding changes in the mood and mental health of large populations is a challenge, with the need for large numbers of samples to uncover any regular patterns within the data. The use of data generated by online activities of healthy individuals offers the opportunity to perform such observations on the large scales and for the long periods that are required. Various studies have previously examined circadian fluctuations of mood in this way. In this study, we investigate seasonal fluctuations in mood and mental health by analyzing the access logs of Wikipedia pages and the content of Twitter in the UK over a period of four years. By using standard methods of Natural Language Processing, we extract daily indicators of negative affect, anxiety, anger and sadness from Twitter and compare this with the overall daily traffic to Wikipedia pages about mental health disorders. We show that both negative affect on Twitter and access to mental health pages on Wikipedia follow an annual cycle, both peaking during the winter months. Breaking this down into specific moods and pages, we find that peak access to the Wikipedia page for Seasonal Affective Disorder coincides with the peak period for the sadness indicator in Twitter content, with both most over-expressed in November and December. A period of heightened anger and anxiety on Twitter partly overlaps with increased information seeking about stress, panic and eating disorders on Wikipedia in the late winter and early spring. Finally, we compare Twitter mood indicators with various weather time series, finding that negative affect and anger can be partially explained in terms of the climatic temperature and photoperiod, sadness can be partially explained by the photoperiod and the perceived change in the photoperiod, while anxiety is partially explained by the level of precipitation. Using these multiple sources of data allows us to have access to inexpensive, although indirect, information about collective variations in mood over long periods of time, in turn helping us to begin to separate out the various possible causes of these fluctuations.

Index Terms—Collective Mood, Mental Health, Social Media, Wikipedia, Discrete Fourier Transform, Multiple Tests, Seasonal Cycles, Information Fusion, Big Data.

I. INTRODUCTION

Periodic variations in mood and mental health have been known for a long time, both at the circadian level (24 hour cycles) and the seasonal level (365 day cycles). The reasons behind them are still the object of investigation, and are likely the result of a multitude of factors, including endocrinological, environmental and social components.

While circadian changes in our physiology are well documented (e.g. blood pressure [1], heart rate [2], body tem-

perature [3], melatonin levels [4]), many factors also show evidence of seasonality in their behaviour, which are reviewed in [5], including blood pressure [6], calorie intake [7], cholesterol levels [8], conception [9], death [10] and even violent suicide [11].

Other studies on seasonal patterns within humans have shown stress and immune function fluctuate on a yearly cycle [12], with recent work also demonstrating seasonal changes in cognitive performance [5].

In the realm of mental health, numerous theories behind seasonal depression have been studied, including exposure to light [13], phase-shift in internal clocks [14], and weather [15], with seasonal affective disorder being a well known phenomenon at high latitudes [16], inducing various levels of depression in the winter months.

However, the difficulty in conducting these studies is often in the collection of data from large populations and over long periods of time, in a way that allows for longitudinal analysis. Large-scale studies based on questionnaires often rely on self-reporting and recollection, and can be unreliable and expensive for a multitude of reasons [17]–[19].

Various studies have attempted to address these problems and collect information that relates to mental health from online activities [20]–[24]. While this may relate to a specific sub-population, it has the advantage of allowing constant monitoring of a very large number of subjects and of bypassing the aforementioned issues of self-reporting.

Of these studies on mental health and mood driven by online activity data, many of the studies only consider circadian patterns in mood [20]–[22], or the effect of world events on mood [25], [26]. We have not found systematic studies of seasonality of mood via Twitter, with only one study addressing the topic of mental health using the incidence of Google searches as an indicator for information-seeking about such topics [24].

In this paper, we study the seasonality of changes in mood indicators both in Twitter and in views of pages on the English-language Wikipedia relating to mental health disorders. We find significant seasonality in both negative mood indicators in Twitter and mental health disorders on Wikipedia, peaking in the winter months.

We furthermore decompose these results down to more

specific moods and individual disorders, discovering that there is meaningful overlap between the periods when negative mood indicators are over-expressed and when views of mental health disorder pages peak.

Possible factors which could explain the seasonal variation in mood are investigated using weather data, finding significant correlation with climatic temperature, photoperiod and the rate of change of the photoperiod with some of the mood indicators on Twitter.

II. METHODS

A. Data collection

1) *Twitter posts*: We gathered social media data from Twitter, an online platform that allows users to publish brief textual communications (tweets) of up to 140 characters, which are publicly visible and available via their application programming interface (API). Using the Twitter API, we collected tweets in the period from January 2010 to November 2014, querying for tweets geo-located to within 10km of any of the 54 largest urban centres in the United Kingdom, without specifying any keywords or hashtags. Due to technical issues, 2012 was excluded from the dataset as it did not follow the same collection procedure. For each tweet, we collected the anonymised textual content, a collection date and time, and information about the location from where the tweet was collected (one of the 54 urban centres). Tweets were preprocessed into their constituent tokens using a tokenizer designed specifically for Twitter text [27]. Tokens representing hyperlinks, mentions and hashtags were discarded, along with tokens containing only special characters (e.g. emoticons).

2) *Wikipedia searches*: We gathered the page view logs for pages on the English-language Wikipedia project covering a four year period between January 2012 and December 2015. Each page view records an anonymised search on Wikipedia, but does not capture other information such as the location of the user, how long is spent by the user on the page, or any other interaction by the user. The data collected is aggregated internally by the Wikimedia Foundation, providing a representative sample of traffic to each page on Wikipedia every hour [28].

As the location of users accessing Wikipedia is not provided in the page view logs, it was not possible to determine whether the users are in the northern or southern hemisphere. However, since the northern hemisphere is believed to account for approximately 73.1% of the traffic to Wikipedia globally [29], we analyse the page view logs for cycles on a 365 day period, as the majority of users experience the same seasonal timing. Cycles on shorter periods are not studied here for Wikipedia due to the interference caused by the two different hemispheres.

3) *Weather data*: We gathered Temperature and Precipitation data in the United Kingdom from the Met Office Hadley Observation Center [30], [31] for the years we collected Twitter data. The observations on weather are performed in a triangular area of the United Kingdom enclosed by Lancashire, London, and Bristol.

Photoperiod (day-length) was computed [32] for the UK using London as a centre and covering the same time interval used for the weather observations.

B. Detection of collective Mood and Mental disorders

We wish to study the relative attention devoted in a given time of the year to the expression of negative mood in Twitter and compare it to search patterns about mental health on Wikipedia. We compiled a separate index for each of the four negative mood indicators in the Linguistic Inquirer and Word Count (LIWC) [33]. We also compiled an index of the searches about *mental health disorders* on the English-language Wikipedia [34].

Specifically, we extracted from the Twitter data every occurrence of the tokens associated with the LIWC *negative affect* indicator, and we collected the search volume from Wikipedia for every page on *mental health disorders*. From each we computed the standardized index series' described in Section II-C, which we analysed for significant seasonal variations and for periods in the year where both are over-expressed. To detail the seasonal components found in both indices, we compiled and compared together separate indices for expressions of *anger*, *anxiety*, and *sadness* in Twitter, and for the search volume of refined queries in Wikipedia.

C. Computing the Standardized Index series

The generation of the time series of an index follows the same procedure for Twitter and for Wikipedia. In the following we describe the generation of a Twitter index from individual LIWC tokens, and we close on comments specific to Wikipedia searches.

1) *Generation of the RF series*: The number of occurrences of a token is first computed every hour in the four years interval. Any missing values within the resulting count series are linearly interpolated, with 29th February removed from leap years to keep each year aligned. This count series is smoothed using a three-hour centred moving average to improve the quality of the statistical estimation for rare tokens. The relative frequency (RF) time series is then obtained by normalizing the counts by the total number of occurrences within each hour. The series of total occurrences is itself interpolated and smoothed.

2) *Aggregating the index*: Using the list of tokens in an LIWC category, we compute the index as follows. a) The relative frequency series of each token is detrended by subtracting its centred two years moving window series. b) It is then standardized, giving each token equal weight. c) The index is calculated as the mean relative frequency series over all tokens. The time series of an index is converted to daily samples to enable the comparison with Wikipedia and weather by averaging over 24 hours intervals in the four years interval, providing a series of length 1,460 (365 days \times 4 years).

3) *Analysis and visualization*: Before any analysis, either Fourier transformation or Pearson correlation, the index is clipped to within three standard deviations to reduce the effect of a small number of particularly emotive days within the

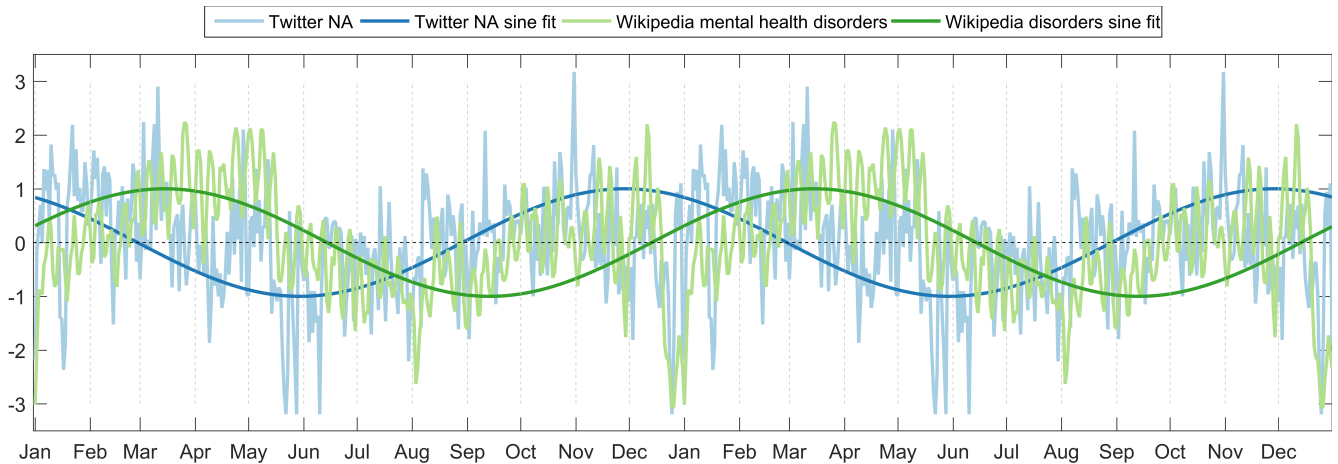


Fig. 1. Series of medians by day of the year for the standardized index of Negative Affect in Twitter UK (blue) and series of medians by day of the year for the standardized index of mental health disorders in English-language Wikipedia (green). The medians are computed across each four year interval, where the series are further standardized and repeated for visual convenience. Overlaid are the sine waves extracted from the standardized index of Negative Affect (light blue) and the standardized index of mental health disorders (light green), each significantly expressed at period 365 days ($p < 0.01$).

series that describe reactions to individual events rather than the collective mood over time. Finally, an index series is standardized for convenience.

When significantly expressed, the 365-day seasonal component of an index is visualized through the series of medians by day of the year over the four years. In Figure 1 the median series is clipped to within three standard deviations and standardized, while in Figure 2 an intermediate smoothing step using a 91-day centred window is applied after clipping and before standardizing. It is performed circularly to provide each day of the year with equal smoothing.

4) *Wikipedia and individual disorders*: Following the same steps we created the Wikipedia standardized index for mental health by computing the number of times each mental health disorder page was accessed each day over the four years we have data, and by smoothing each individual count series in a three-day centred window. Additionally, given the generic description a Wikipedia article provides on each disorder, we also considered them as individual series themselves (e.g. series for *Seasonal Affective Disorder*, *Mood Disorder*, etc.).

D. Significant seasonal components by Fourier analysis

The Fourier Transform on a 1,460 length index series provides a decomposition of the series in terms of individual sine waves that oscillate at each of the harmonic frequencies respectively: each sine component is summarized by the variance explained from the signal and by the phase that indicates the times of over-expression in the corresponding period.

To assess the seasonality of a given index, we tested the frequency response from the Fourier transform at a period of 365 days. We performed a statistical significance test where our null hypothesis, that “the index is not periodic with a period of 365 days” is rejected at a significance level of 1% if the variance explained (the test statistic) from a 365 day sinusoidal movement is not larger than that of the original index series in more than 1% of 100,000 random permutations

of the index. We additionally tested the Twitter indices for significant periodicity at a bi-annual period, repeating the statistical significance test used here¹.

To account for the multiple tests that we perform, we employ the Bonferroni correction to maintain the probability of at least one false rejection at the significance level of 1%.

E. Weather comparisons

Correlation between different attributes of the weather and affect within Twitter are computed using the Pearson correlation coefficient [35]. Significance of the correlation between a mood index on Twitter and a weather attribute w is assessed under the null hypothesis that “the correlation coefficient between the mood index and w is equal to zero”, rejecting the null hypothesis using a significance level of 1% as calculated from the Student’s t-test [36] for the correlation coefficient, again corrected using the Bonferroni correction to account for multiple tests.

III. RESULTS

A. Seasonal components to mood and mental health

We found that negative affect in Twitter is significantly periodic at the one year period, with the phase indicating that it is most over-expressed in late November each year. Similarly, the Wikipedia index of mental health disorders was also significantly periodic at the one year period, with a phase corresponding to over-expression being strongest in mid-March. This can be seen in Figure 1, where the negative affect in Twitter and the Wikipedia index of mental health disorders are visualised as their median yearly pattern, repeated over a two year interval. The idealised sine wave found by the Fourier transform describing their seasonal pattern is additionally shown.

¹Wikipedia was not tested for bi-annual periodicities due to the data being sourced from both hemispheres.

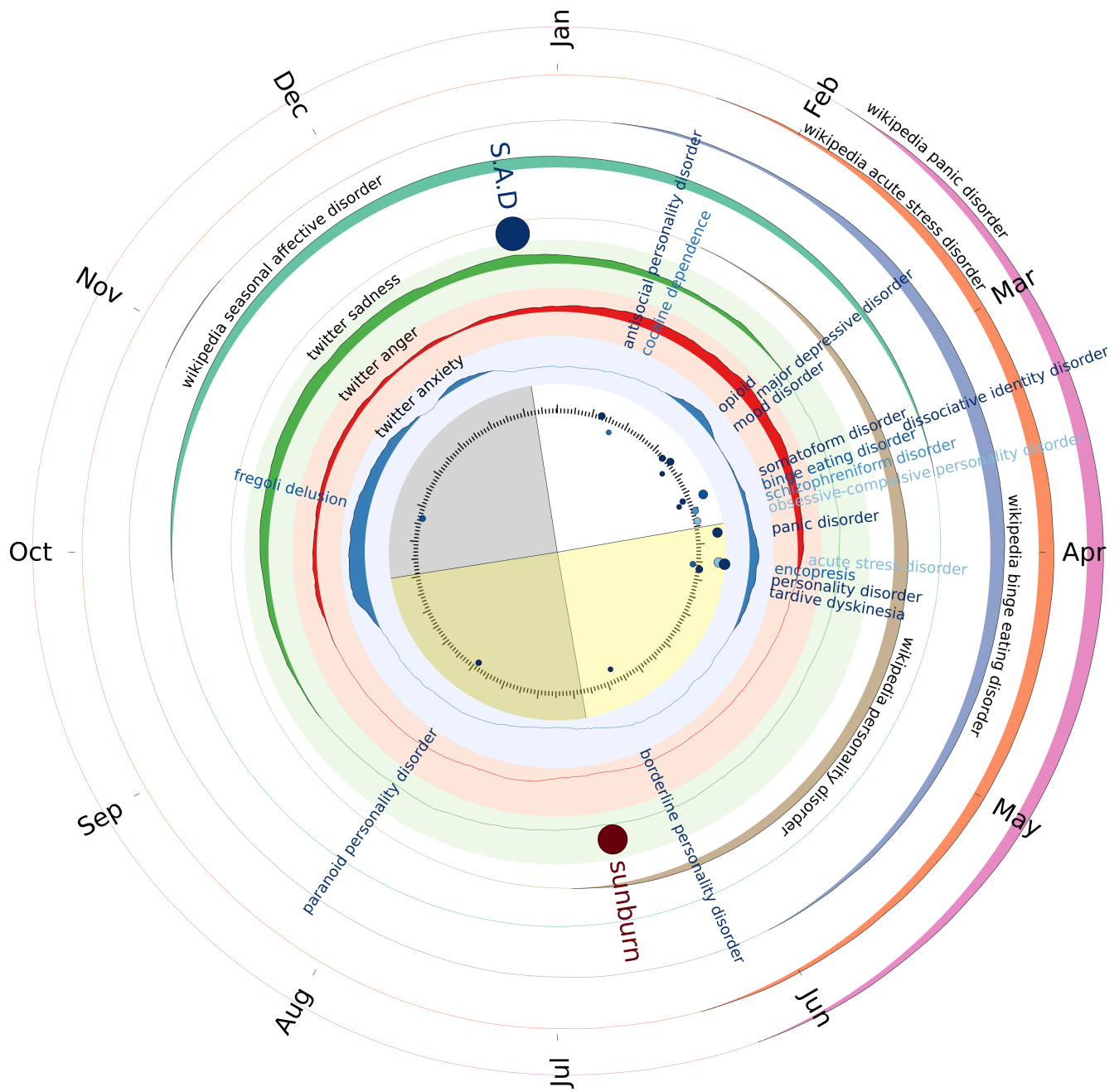


Fig. 2. Seasonal expressions of the negative mood indicators in the United Kingdom and of the mental health disorders on English-language Wikipedia. Semi-arcs illustrate periods of over-expression along the year, in Twitter: Anxiety (blue), Anger (red), and Sadness (green); and in the five most seasonal Wikipedia disorders: Personality disorder (brown), Seasonal Affective Disorder (cyan), Binge Eating Disorder (blue), Acute Stress Disorder (orange), and Panic Disorder (pink). Remaining mental health disorder pages with above 10% variance explained from a sine wave of period 365 days are indicated with a circle, where the radius represents the explained variance and the angle the sine peak date. The horizontal line indicates the Spring and Autumn Equinoxes, the vertical line indicates the Winter and Summer Solstices. Times are computed based on the day-length series in the United Kingdom.

We see that there is a peak in mental health disorders on Wikipedia in March, occurring around the time of the Spring equinox. We do not however see a corresponding peak in mental health disorders in September when the southern hemisphere Spring equinox takes place. A second peak in the mental health disorders instead occurs in November, a period when the northern hemisphere experiences the shortest days

of the year, along with the greatest over-expression of negative affect on Twitter.

In the following section, we further investigate the overlap between the negative affect on Twitter and the second seasonal component of the Wikipedia index by decomposing the negative affect indicator into refined emotions, and detailing the most seasonal disorders on Wikipedia.

B. Decomposing the indices

Breaking the negative affect in Twitter down into different indices relating to anger, anxiety and sadness, we found that while anger and sadness peak in the winter months and are significantly periodic at the one year period, anxiety is better explained by a six month period, peaking in the autumn and spring. In more detail, anger is over-expressed between September and April, anxiety is under-expressed between April and September, and that sadness grows above its mean value in September, declining again in February with its peak found in November.

Focusing on individual mental health disorder pages within Wikipedia, we found that the five pages with the most periodic views are Seasonal Affective Disorder, Personality Disorder, Acute Stress Disorder, Panic Disorder, and Binge Eating Disorder. In each of these cases we found that the periodicity of the indices is significant, and that a large part of their variance can be explained by the seasonal component, ranging from 65.6% for Seasonal Affective Disorder to 17.7% for Binge Eating Disorder. Indeed, a further 13 mental health disorders have more than 10% of their variance explained by a seasonal component.

Figure 2 shows the times when the different mood indices from Twitter along with the five most periodic disorders from Wikipedia are over-expressed throughout the year. It also provides a scatter in polar coordinates of the mental health disorders, with over 10% variance explained from a sinusoidal seasonal component, where the radius represents the explained variance and the angle the sine peak date. The peaking time for the index of the Wikipedia page on sunburn (58.3% of variance explained with a seasonal pattern) is additionally to ground the results. We see that Seasonal Affective Disorder peaks in December at the darkest time of the year, during the Winter solstice. We also note a concentration of mental health disorders around the Spring equinox.

Visually, we immediately see there is strong overlap between the periods when these Wikipedia mental health disorder pages peak and those in which the negative mood indicators are over-expressed on Twitter. The two sources of data combine to show an intriguing picture of collective fluctuations of mental health and mood following a seasonal cycle each year.

C. Association with Weather

Finally, we attempted to tease apart the various possible factors behind the seasonality we found in negative mood indicators on Twitter, where many have been found to be associated with mental health in the literature, including photoperiod duration [37], temperature [38], precipitation [39], and the relative phase-shift with bright light exposure. [40], [41].

As displayed in Table I, we computed the correlation between the negative mood indicators on Twitter with attributes of the seasons in the United Kingdom, such as the temperature, photoperiod, rate of change of the photoperiod and precipitation, detailing the variance explained by each.

TABLE I
VARIANCE EXPLAINED AND CORRELATION COEFFICIENTS BETWEEN WEATHER IN CENTRAL ENGLAND AND THE TWITTER NEGATIVE MOOD INDICATORS IN THE UNITED KINGDOM. BONFERRONI CORRECTION IS APPLIED FOR 16 TESTS, SIGNIFICANCE AT 1% MARKED IN BOLD.

Temperature				
	Variance Explained	Correlation (r)	P-value (p)	
NegativeAffect	8.73e-03	-9.34e-02	7.44e-03	
Anger	1.60e-02	-1.27e-01	3.22e-05	
Anxiety	1.39e-03	3.73e-02	2.61e+00	
Sadness	4.29e-03	-6.55e-02	2.29e-01	
Photoperiod				
	Variance Explained	Correlation (r)	P-value (p)	
NegativeAffect	1.39e-02	-1.18e-01	1.52e-04	
Anger	1.89e-02	-1.38e-01	3.68e-06	
Anxiety	6.49e-04	2.55e-02	5.46e+00	
Sadness	1.82e-02	-1.35e-01	6.31e-06	
Rate of Change of Photoperiod				
	Variance Explained	Correlation (r)	P-value (p)	
NegativeAffect	1.92e-03	-4.38e-02	1.62e+00	
Anger	4.62e-05	6.80e-03	1.28e+01	
Anxiety	8.07e-03	-8.98e-02	1.23e-02	
Sadness	1.24e-02	-1.12e-01	4.57e-04	
Precipitation				
	Variance Explained	Correlation (r)	P-value (p)	
NegativeAffect	7.68e-03	8.76e-02	1.65e-02	
Anger	4.16e-03	6.45e-02	2.52e-01	
Anxiety	8.44e-03	9.19e-02	9.24e-03	
Sadness	7.01e-03	8.38e-02	2.74e-02	

After correcting for multiple tests we found that, in the United Kingdom, longer days are associated with reduced feelings of Sadness ($r = -1.12e-01$, $p < 0.01$) and expressions of Anger ($r = -1.38e-01$, $p < 0.01$), and that the general negative sentiment is also decreased ($r = -1.18e-01$, $p < 0.01$).

In addition, we also find a significant negative correlation between sadness indicators and the rate of change in day-length ($r = -1.12e-01$, $p < 0.01$). This finding lends further support to the phase-shift hypothesis, which postulates that internal body clocks are regulated by not just the amount of bright light exposure, but also on the phase-shift of the photoperiod, a phenomenon exploited to treat patients suffering from seasonal affective disorders [40], [41].

We also found that indicators for anger ($r = -1.38e-01$, $p < 0.01$) and negative affect ($r = -1.18e-01$, $p < 0.01$) are reduced on warmer days. These results point to the influence of the temperature on the regulation of our emotions amongst other factors, while precipitation was found to be significantly correlated with the anxiety indicator ($r = 9.19e-02$, $p < 0.01$), suggesting that days with higher precipitation may elicit feelings of anxiety within the United Kingdom.

IV. CONCLUSION

We have shown a significant seasonal component to different and independent indicators of mood and mental health in large populations as expressed through social media messages on Twitter and access to information on mental health disorders on English-language Wikipedia. The methods demon-

strated in this study can be augmented with even further data, allowing us to have - indirect - access to the mental state of healthy individuals over long periods and for large populations. Furthermore, we go some way in beginning to explain the variance in the mood and mental health disorder indicators by combining the data with weather information, finding significant correlations between the climatic environment and mood indicators on Twitter.

The application of such techniques allows for the potentially real-time monitoring of mood and mental health of a population, signalling broad trends over time without the need to rely on expensive and slow sentinels of mental health as used currently within the profession [42]. Use of social media and information-seeking trends on the web can additionally help overcome issues in collecting data directly from respondents including the stigma of mental health issues [43], interviewer [19] and question bias [17] and other well documented caveats to self-reporting [18].

Further work will aim to identify a more complete picture of the influences of mood on Twitter, complementing weather patterns and information-seeking behaviour shown in this study with the content of the media and real-world events. The analysis of variance explained by various factors can uncover increasingly subtle factors affecting collective mood.

ACKNOWLEDGMENTS

The authors were funded by the ERC Advanced Grant ThinkBIG.

REFERENCES

- [1] M. Millar-Craig, C. Bishop, and E. Raftery, "Circadian variation of blood-pressure," *The Lancet*, vol. 311, no. 8068, pp. 795–797, 1978.
- [2] H. V. Huikuri, K. M. Kessler, E. Terracall, A. Castellanos, M. K. Linnaluoto, and R. J. Myerburg, "Reproducibility and circadian rhythm of heart rate variability in healthy subjects," *The American journal of cardiology*, vol. 65, no. 5, pp. 391–393, 1990.
- [3] R. Refinetti and M. Menaker, "The circadian rhythm of body temperature," *Physiology & behavior*, vol. 51, no. 3, pp. 613–637, 1992.
- [4] A. Cagnacci, J. Elliott, and S. Yen, "Melatonin: a major regulator of the circadian rhythm of core temperature in humans," *The Journal of Clinical Endocrinology & Metabolism*, vol. 75, no. 2, pp. 447–452, 1992.
- [5] C. Meyer, V. Muto, M. Jaspas, C. Kussé, E. Lambot, S. L. Chellappa, C. Degueldre, E. Balteau, A. Luxen, B. Middleton *et al.*, "Seasonality in human cognitive brain responses," *Proceedings of the National Academy of Sciences*, vol. 113, no. 11, pp. 3066–3071, 2016.
- [6] P. R. Woodhouse, K.-T. Khaw, and M. Plummer, "Seasonal variation of blood pressure and its relationship to ambient temperature in an elderly population," *Journal of hypertension*, vol. 11, no. 11, pp. 1267–1274, 1993.
- [7] Y. Ma, B. C. Olendzki, W. Li, A. R. Hafner, D. Chiriboga, J. R. Hebert, M. Campbell, M. Sarnie, and I. S. Ockene, "Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population," *European journal of clinical nutrition*, vol. 60, no. 4, pp. 519–528, 2006.
- [8] E. Kristal-Boneh, G. Harari, and M. S. Green, "Circannual variations in blood cholesterol levels," *Chronobiology international*, vol. 10, no. 1, pp. 37–42, 1993.
- [9] E. L. Rizzi and G. Dalla-Zuanna, "The seasonality of conception," *Demography*, vol. 44, no. 4, pp. 705–728, 2007.
- [10] A. Douglas, T. Allan, and J. Rawles, "Composition of seasonality of disease," *Scottish medical journal*, vol. 36, no. 3, pp. 76–82, 1991.
- [11] M. Maes, P. Cosyns, H. Y. Meltzer, F. De Meyer, and D. Peeters, "Seasonality in violent suicide but not in nonviolent suicide or homicide," *American Journal of Psychiatry*, vol. 150, pp. 1380–1380, 1993.
- [12] R. J. Nelson, G. E. Demas, S. L. Klein, and L. J. Kriegsfeld, *Seasonal patterns of stress, immune function, and disease*. Cambridge University Press, 2002.
- [13] R. N. Golden, B. N. Gaynes, R. D. Ekstrom, R. M. Hamer, F. M. Jacobsen, T. Suppes, K. L. Wisner, and C. B. Nemeroff, "The efficacy of light therapy in the treatment of mood disorders: a review and meta-analysis of the evidence," *American Journal of Psychiatry*, vol. 162, no. 4, pp. 656–662, 2005.
- [14] A. Wirz-Justice, "Biological rhythm disturbances in mood disorders," *International Clinical Psychopharmacology*, vol. 21, pp. S11–S15, 2006.
- [15] M. C. Keller, B. L. Fredrickson, O. Ybarra, S. Côté, K. Johnson, J. Mikels, A. Conway, and T. Wager, "A warm heart and a clear head the contingent effects of weather on mood and cognition," *Psychological Science*, vol. 16, no. 9, pp. 724–731, 2005.
- [16] J. Haggarty, Z. Cernovsky, M. Husni, K. Minor, P. Kermeen, and H. Merskey, "Seasonal affective disorder in an arctic community," *Acta Psychiatrica Scandinavica*, vol. 105, no. 5, pp. 378–384, 2002.
- [17] S. L. B. Payne, *The Art of Asking Questions: Studies in Public Opinion*, 3. Princeton University Press, 2014, vol. 3.
- [18] N. Schwarz and G. L. Clore, "Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states," *Journal of personality and social psychology*, vol. 45, no. 3, p. 513, 1983.
- [19] I. C. Choi and G. W. Comstock, "Interviewer effect on responses to a questionnaire relating to mood," *American journal of epidemiology*, vol. 101, no. 1, pp. 84–92, 1975.
- [20] S. A. Golder and M. W. Macy, "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, vol. 333, no. 6051, pp. 1878–1881, 2011.
- [21] 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, no. 12, p. e26752, 2011.
- [22] V. Lampos, T. Lansdall-Welfare, R. Araya, and N. Cristianini, "Analysing mood patterns in the united kingdom through twitter content," *arXiv preprint arXiv:1304.5507*, 2013.
- [23] J. Li, X. Wang, and E. Hovy, "What a nasty day: Exploring mood-weather relationship from twitter," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2014, pp. 1309–1318.
- [24] J. W. Ayers, B. M. Althouse, J.-P. Allem, J. N. Rosenquist, and D. E. Ford, "Seasonality in seeking mental health information on google," *American journal of preventive medicine*, vol. 44, no. 5, pp. 520–525, 2013.
- [25] V. D. Nguyen, B. Varghese, and A. Barker, "The royal birth of 2013: Analysing and visualising public sentiment in the uk using twitter," in *Big Data, 2013 IEEE International Conference on*. IEEE, 2013, pp. 46–54.
- [26] T. Lansdall-Welfare, V. Lampos, and N. Cristianini, "Effects of the recession on public mood in the uk," in *Proceedings of the 21st International Conference on World Wide Web*. ACM, 2012, pp. 1221–1226.
- [27] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith, "Part-of-speech tagging for twitter: Annotation, features, and experiments," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 2011, pp. 42–47.
- [28] Wikitech, "Analytics/data/pagecounts-raw — wikitech," 2015, [Online; accessed 2-August-2016]. [Online]. Available: <https://wikitech.wikimedia.org/w/index.php?title=Analytics/Data/Pagecounts-raw&oldid=243874>
- [29] W. Foundation, "Wikimedia traffic analysis report - wikipedia page views per country - overview," 2016, [Online; accessed 2-August-2016]. [Online]. Available: <https://stats.wikimedia.org/wikimedia/squids/SquidReportPageViewsPerCountryOverview.htm>
- [30] D. E. Parker, T. P. Legg, and C. K. Folland, "A new daily central england temperature series, 1772–1991," *International Journal of Climatology*, vol. 12, no. 4, pp. 317–342, 1992.
- [31] L. Alexander and P. Jones, "Updated precipitation series for the uk and discussion of recent extremes," *Atmospheric science letters*, vol. 1, no. 2, pp. 142–150, 2000.

- [32] W. C. Forsythe, E. J. Rykiel, R. S. Stahl, H.-i. Wu, and R. M. Schoolfield, "A model comparison for daylength as a function of latitude and day of year," *Ecological Modelling*, vol. 80, no. 1, pp. 87–95, 1995.
- [33] 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.
- [34] "List of mental disorders - wikipedia, the free encyclopedia," 2016, [Online]. [Online]. Available: https://en.wikipedia.org/wiki/List_of_mental_disorders
- [35] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.
- [36] M. G. Kendall, A. Stuart, and J. Ord, *The advanced theory of statistics*. London, 1968, vol. 3.
- [37] R. D. Levitan, "The chronobiology and neurobiology of winter seasonal affective disorder," *Dialogues in clinical neuroscience*, vol. 9, no. 3, p. 315, 2007.
- [38] A. Shapira, R. Shiloh, O. Potchter, H. Hermesh, M. Popper, and A. Weizman, "Admission rates of bipolar depressed patients increase during spring/summer and correlate with maximal environmental temperature," *Bipolar disorders*, vol. 6, no. 1, pp. 90–93, 2004.
- [39] A. Preti, "The influence of seasonal change on suicidal behaviour in italy," *Journal of affective disorders*, vol. 44, no. 2, pp. 123–130, 1997.
- [40] C. I. Eastman, M. A. Young, L. F. Fogg, L. Liu, and P. M. Meaden, "Bright light treatment of winter depression: a placebo-controlled trial," *Archives of General Psychiatry*, vol. 55, no. 10, pp. 883–889, 1998.
- [41] A. J. Lewy, R. L. Sack, C. M. Singer, D. M. White, and T. M. Hoban, "Winter depression and the phase-shift hypothesis for bright light's therapeutic effects: history, theory, and experimental evidence." *Journal of Biological Rhythms*, 1988.
- [42] T. Strine, S. Dhingra, L. McGuire, L. Elam-Evans *et al.*, "Public health surveillance for mental health," 2010.
- [43] W. C. Reeves, T. W. Strine, L. A. Pratt, W. Thompson, I. Ahluwalia, S. S. Dhingra, L. R. McKnight-Eily, L. Harrison, D. V. D'Angelo, L. Williams *et al.*, "Mental illness surveillance among adults in the united states," *MMWR Surveill Summ*, vol. 60, no. Suppl 3, pp. 1–29, 2011.