ACCSCIENCE PUBLISHING

Global Translational Medicine

PERSPECTIVE ARTICLE

YeeZzzy does it: Using Kanye West's tweets to identify sleep and emotional disturbances through Digital Rest-activity Rhythm (dRARs) analysis

Supplementary File

1. Methods

We accessed and downloaded the user-generated text content and date/time data from 1868 tweets using the Twitter.com application programming interface (API) and a custom programming script written in Python. The API provided a programmatic interface to interact with the social media platform and retrieve user-generated text content. Hypertext Markup Language (HTML) queries were used to parse text and time data into JSON format and then imported into R Studio for data cleaning and analyses. Tweets were extracted from the Twitter page between April 14, 2018, and November 4, 2020, as these dates spanned the range of the earliest and latest tweets available. Before linguistic analysis, text data were cleaned to remove special characters which may influence algorithmic outputs. "Re-tweets" and tweets without any linguistic content (such as images/"gifs"/"memes") were included as time-indicative data-points but were omitted from linguistic analysis, due to the presence of linguistic content from other users. In other words, time data from re-tweets images, etc. were used to inform when tweeting activity occurred but were not included as observations of outcome variables in the statistical models. Character symbols conveying emoticons were re-coded into lexical equivalents, according to industry-standard dictionary (Emoji List, v15.0 [unicode.org]) containing the accepted nomenclature assigned to respected UTF-8 encoded characters (*e.g.*, ":)" = "slightly smiling face").

1.1. Temporal data analysis

Timestamps were first converted from native format (UTC) to local time (e.g., pacific daylight time, pacific standard time, eastern standard time) before any temporal analyses. To estimate the period of peak quiescence purported to represent the user's sleep window we adopted a common and well-validated approach used in assessments of human circadian rhythms, by calculating the least active 5 h of each day (the L5 index). To achieve this, we pooled the data and averaged it into one 24-h period. A sliding window function with a step size of 1 h was then used to calculate the total number of tweets which occurred within that 5-h window. We permitted "wrapping" of this window to prevent an assumed to be stationary 24-h period from identifying false L5 estimates. For example, the following L5 window: 20:00, 21:00, 22:00, 23:00. 0:00 is stationary as each of the 1-h bins falls within the bounds of one "finite" 24-h cycle. However, the L5 window 23:00 | 00:00 | 01:00 | 02:00 | 03:00 falls across the boundaries of two. An "unwrapped" sliding window would identify this single period as two independent activity troughs (23:00 – 00:00 and 00:00 – 04:00) biasing the estimated L5. Instead, we permitted "wrapped" L5 windows which spanned between two 24-h periods. We calculated L5 in our data as the 5-h window (in this case, 23:00 – 04:00) which contained the lowest total number of tweets, when averaging across all days. Next, we assigned a binary classifier to each tweet according to whether the tweet was sent within or without the L5 window. For the purposes of interpretation and reporting, we defined tweets within the L5 window as "nocturnal" and "diurnal," and it should be noted that these labels do not necessarily indicate the position on the dark/light cycle. Nevertheless, it is worth noting that the primary geographical latitude of the user (mainland USA, primarily the west coast) resulted in this window always occurring during hours of darkness. These binary labels (0, 1) were then summed to create the total number of nocturnal/L5 tweets. Although many users log daily activity on social media platforms, use is sporadic and non-contiguous in nature, and daily use can neither be assumed nor expected from a given user. Therefore, we also calculated a metric representing the total number of tweets over the preceding seven days which occurred during the L5 period. This allowed us to maximize statistical power by increasing datapoints, and also to test the clinically relevant yet hypothetical influence of accumulated hypothetical sleep/circadian disruption on mood.

2. Data visualization

To facilitate interpretation of the data generated by our methods to the most commonly used benchmark technique in sleep and circadian science, we produced circadian "double plots" to visualize the bins of activity over the entire period in 48-h epochs (rows), displayed across time (columns). Whilst Roenneberg used a dedicated circadian software, we opted for an approach that would facilitate more open-source development of these methods. To do so, we adopted an approach akin to those used to visualize other time-series sleep/circadian data generated by actigraphy and electroencephalography. Such approaches use either binary (activity/no activity) or continuous (activity count/epoch) time series assigned to each predefined time bin/epoch. We opted for a 20-min epoch, as the results of our initial sensitivity analyses resulted in this epoch lengths of 20 min generating the highest visual and temporal resolution, whilst still permitting clear visual delineation of temporal fluctuation rhythms. Twenty minutes are also frequently used as meaningful epochs in studies employing activity methods, which facilitate meaningful comparison. Whilst a standardized consensus epoch-length may ultimately be reached following further investigations, we recommend an iterative approach to tailoring epoch length to each specific population/participant based on the nuances of the observed data, and factors such as tweet resolution (*e.g.*, the number per hour), the length of the total time period and the overall tweet frequency and total numbers of tweets. The local-time timestamps of the tweets were first down-sampled using the MATLAB Signal Processing Toolbox to assign replace the timestamp with the time-epoch in which it fell. Epochs repeated as follows between 00:00 and 11:59 for each day in which tweets were present: hh:00 – hh:19 | hh:20 – hh:39 | hh:40 – hh:59. To then epoch the tweet times, we then simulated a time-series of 20-min epochs across the entire period (total 705,600 bins). The time series was transposed into vertical format, and the rows of tweet data were joined using a left join function so that the timestamps of the tweet were matched to their appropriate bin, and intermediate periods were left empty. We then visualized the data as a 48-h double plot using R, using a heatmap function, where each cell in the heatmap matrix represented a 20-min time bin and was filled if tweet activity occurred during that bin. A color bar was used to represent a continuous value representing the ratio between positive and negative tweets. Dark blue indicated a tweet ratio of 1:0 negative to positive, where as bright red indicated 0:1 negative to positive. Although our analyses conceptualized these two dimensions as orthogonal concepts, we opted for this approach during data visualization as a compromise to using two different double plots to present negative and positive emotion independently, whilst also indicating overall tweet (non-emotional) activity. Data that were timestamped but not used in analyses (e.g., retweets) were indicated by gray shading in the matrix. In addition to this double plot, we also visualized trajectories of positive and negative tweets by plotting the positive emotion (posemo) and negative emotion (negemo) values averaged across the entire period.

2.1. Linguistic enquiry and word count (LIWC)

We used the LIWC algorithm to classify the positive and negative content of tweets. We used the "positive emotion" and "negative emotion" output variables from LIWC as our outcome, which give an overall positivity and negativity score for each tweet based on a validated linguistic dictionary of 6400 words. Due to the large number of possible variables produced by LIWC, we opted to limit our model to the indices of positive and negative affect. Although our *a priori* analyses were limited to the posemo and negemo values generated by the LIWC analysis, the outputs of the LIWC algorithm provide a much more granular picture of emotionality across the spectrum of both positive and negative emotions. To represent these relationships visually, without generating further hypotheses or statistical tests, we conducted network analyses to visualize the relationship among nodes. We selected the 10 emotional constructs with the highest weightings assigned by the LIWC algorithm (Anger, Sadness, Anxiety, Family, Power, Religion, Sexual, Money, Risk, and Death) and established a network matrix combining these values with positive and negative emotion. The relationship (signified by the thickness of the cord) between nodes was plotted using cord plots, and the relative weighting of the emotion categories was indicated by the relative proportion of the circumference band.

2.2. Statistical analyses and diagnostics

We tested the hypothesized relationship between nocturnal tweeting and daytime emotion in relation to the previous night's nocturnal tweeting and average number of nocturnal tweets for the seven previous days. Analyses of positive and negative emotional words specified the inclusion of time of day, month, and the total number of words in the tweet *a priori*, to prevent confounding their confounding effects. Nevertheless, we first tested the relationship of these confounders with positive and negative emotion using zero-inflated generalized linear modeling to test for main and interaction effects of hour, month, and affect (positive vs. negative) both to serve as a validation check for their pre-specified inclusion in models and to test the sensitivity of the model to detect known effects observed at the population level in a single user. Tweets for which there was no data available the previous day or the previous seven days were excluded from the respective analyses. We hypothesized that increased tweeting during the nocturnal period would be associated with decreased positive and increased negative daytime affect (see detailed hypotheses below).

We first probed the central tendency and distributions of our independent and dependent variables using histograms and statistical diagnostic tests. The Shapiro–Wilk test indicated that the proposed predictors (number of nocturnal tweets: last night and past week) and outcomes deviated significantly from a normal distribution (both $P < 0.001$). Similarly, the Anderson–Darling test confirmed the departure from normality for each variable (both $P < 0.001$). Furthermore, positive emotion, negative emotion, and frequency of nocturnal tweets (both daily and weekly), exhibited substantial skewness (>+2 or ≤2) and kurtosis (>+3 or ≤2) values, suggesting outcome and predictor variables followed a log-normal distribution with right-skewing in the presence of zero values. As a result, we employed a zero-inflated Poisson model to account for the excess zeros in the data and appropriately handle both normally distributed predictors and non-normally distributed outcomes. These methods best handle the logarithmic distribution and are robust to deviations from the assumption requirements of linear models. The zero-inflated Poisson model was implemented using the Rigby and Stasinopoulos method facilitated by the use of the Generalized Additive Models for Location, Scale, and Shape (GAMLSS) framework in the R programming language. We controlled the fitting process using 500 cycles to improve the convergence and stability of the model estimation. We evaluated model performance by examining the global deviance, reflecting overall goodness-of-fit, as well as the Akaike Information Criterion, and the Schwarz Bayesian Criterion values, reflecting model parsimony, with lower values indicative of better trade-off between model complexity and fit. Finally, we examined overdispersion, which we calculated as the ratio of the residual deviance to the residual degrees of freedom.

Based on our prior work, we hypothesized that positive and negative emotion values would demonstrate relatively weak levels of correlation, and therefore proposed independent univariate models *a priori*. However, during the data cleaning phase, we verified (during pre-analysis) whether univariate or multivariate modeling may be most appropriate by calculating the correlation between outcome variables, which demonstrated weak correlation ($r = -0.052$, $P > 0.05$) and therefore weak interdependence. This confirmed our assumption and justified the use of univariate models, permitting the exploration of unique associations between each outcome variable and the predictors without assuming dependency between the two.

Our data relied on samples obtained from a single participant. Therefore, to assess the stability and variability of the coefficients estimated by the GAMLSS model, a bootstrapping procedure was performed. Bootstrap is a resampling method that involves randomly drawing samples with replacement from the original dataset. For each bootstrap sample, a GAMLSS model was fitted with the same set of predictors and response variable. The bootstrapping process was repeated multiple times, resulting in a distribution of coefficient estimates for each predictor, which provides a measure of the uncertainty associated with the coefficient estimates. The results showed that the coefficients for the predictor's month_simple, WC, Local_Time_hhEX, and total_noct_tweets_week were reasonably stable, with narrow confidence intervals. The coefficient estimates were statistically significant and the confidence intervals did not span zero, suggesting that these predictors likely have a "true" significant effect on the outcome variable. All analyses were performed on an Intel(R) Xeon(R) Silver 4208: CPU with a 2.10GHz processor and 128 GB of RAM and 16 cores, accessed through a 64-bit Windows operating system. Therefore, to obtain the most accurate estimations of the model's validity and stability estimates, we opted for a much higher than usual number of bootstrapping iterations (20,000). However, a standard computing system may lack the computing power to run such models; therefore, future investigators may need to consider the balance between model stability and computational power.

3. Hypotheses

Hypothesis 1a: Increased nocturnal tweeting during the previous L5 period would be associated with reduced positive emotional scores the following day.

Hypothesis 1b: Increased nocturnal tweeting during the previous L5 period would be associated with increased negative emotional scores the following day.

Hypothesis 2a: Increased nocturnal tweeting during the previous seven L5 periods would be associated with reduced positive emotional scores the following day.

Hypothesis 2b: Increased nocturnal tweeting during the previous seven L5 periods would be associated with increased negative emotional scores the following day.

Supplementary Table

Table S1. Statistical parameters from zero-inflated poisson regression models

	Negative Emotion								
Model 1	**B**	**SE**	**T value**	***P*-value**	**2.5% CI**	**97.5% CI**	**BIC**	**AIC**	**General deviance**
Intercept	3.21667719	0.09531799	33.7467987	<0.0000001	2.85799822	3.55428501	2232.351	2201.372	2189.372
Month	−0.0105418	0.00776453	−1.3576867	0.1748015	−0.0413169	0.02126749			
WC	−0.0509405	0.00250136	−20.36513	<0.0000001	−0.0613504	−0.040916			
Local time	0.00237402	0.00504781	0.47030605	0.63821619	−0.020035	0.02371284			
Number of nocturnal tweets (past week)	−0.0003129	0.00020064	−1.559363	0.11915671	−0.0067408	0.00024523			
Model 2									
Intercept	3.33746195	0.04852571	68.7771932	<0.0000001	3.09970417	3.56113169	5840.143	5809.164	5797.164
Month	0.00781834	0.00433349	1.80416831	0.07143901	−0.0139083	0.02911045			
WC	−0.0481352	0.00121237	−39.703494	<0.0000001	−0.0544039	−0.0417249			
Local time	−0.0043063	0.00244116	−1.7640188	0.0779664	−0.0164334	0.00790276			
Number of nocturnal tweets (previous L5)	−0.0001468	4.48E−05	−3.2760662	0.00108073	−0.000327	−0.00000464			
Positive emotion									
Model 1									
Intercept	3.20845477	0.09633962	33.3035839	<0.0000001	2.85799822	3.55428501	5747.638	5716.659	5704.659
Month	−0.0109185	0.00778174	−1.4030897	0.16083168	−0.0413169	0.02126749			
WC	−0.0513694	0.00252958	−20.307495	<0.0000001	−0.0613504	−0.040916			
Local time	0.00310606	0.00513888	0.60442262	0.5456694	−0.020035	0.02371284			
Number of nocturnal tweets (past week)	0.01043037	0.01892969	0.55100581	0.58172548	−0.0067408	0.00024523			
Model 2									
Intercept	3.33368056	0.04841737	68.8529827	<0.0000001	3.09630629	3.55571741	5856.038	5825.059	5813.059
Month	0.00876233	0.00432678	2.02514007	0.04305939	−0.0130729	0.03026022			
WC	−0.0476967	0.00121543	−39.242614	<0.0000001	−0.0538891	−0.041547			
Local time	−0.0040634	0.00243563	−1.6683058	0.09549866	−0.016033	0.00813939			
Number of nocturnal tweets (previous L5)	−0.0299266	0.00678945	−4.4078038	<0.0000001	−0.0576052	−0.0081686			

Abbreviations: AIC: Akaike information criterion; BIC: Bayesian information criterion; CI: Confidence interval; SE: Standard error; WC: Word count.