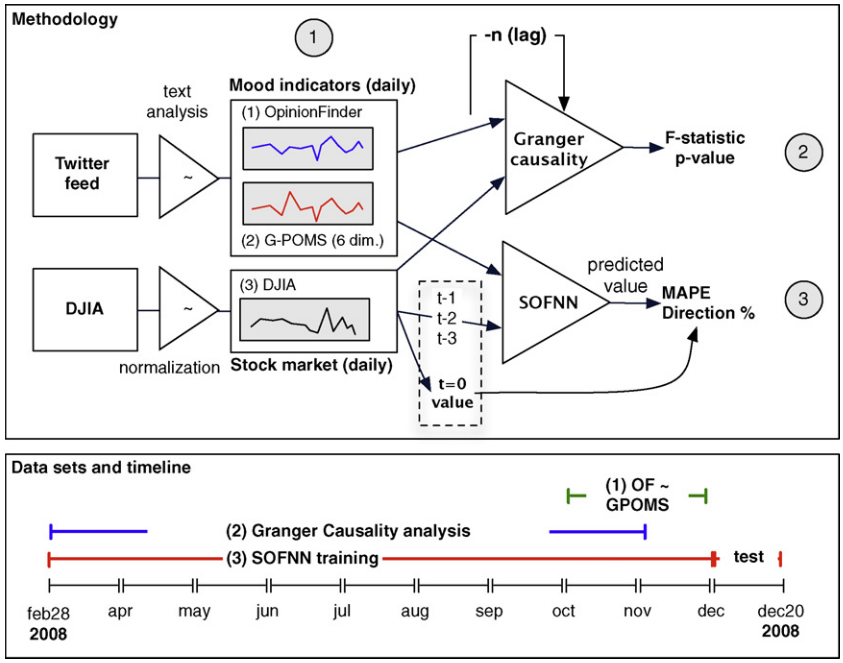


**Fig. 1.** Diagram outlining 3 phases of methodology and corresponding data sets: (1) creation and validation of OpinionFinder and GPOMS public mood time series from October 2008 to December 2008 (Presidential Election and Thanksgiving), (2) use of Granger causality analysis to determine correlation between DJIA, OpinionFinder and GPOMS public mood from August 2008 to December 2008, and (3) training of a Self-Organizing Fuzzy Neural Network to predict DJIA values on the basis of various combinations of past DJIA values and OF and GPOMS public mood data from March 2008 to December 2008.

original POMS terms (in accordance with its co-occurence weight) and via the POMS scoring table to its respective POMS dimension. The score of each POMS mood dimension is thus determined as the weighted sum of the co-occurence weights of each tweet term that matched the GPOMS lexicon. Data sets and methods are available on our project web site.[7]

To enable the comparison of OF and GPOMS time series we normalize them to $z$-scores on the basis of a local mean and standard deviation within a sliding window of $k$ days before and after the particular date. For example, the $z$-score of time series $X_t$, denoted $\mathbb{Z}_{X_t}$, is defined as:

$$\mathbb{Z}_{X_t} = \frac{X_t - \bar{x}(X_{t\pm k})}{\sigma(X_{t\pm k})} \tag{1}$$

where $\bar{x}(X_{t\pm k})$ and $\sigma(D_{t\pm k})$ represent the mean and standard deviation of the time series within the period $[t-k, t+k]$. This normalization causes all time series to fluctuate around a zero mean and be expressed on a scale of 1 standard deviation.

The mentioned $z$-score normalization is intended to provide a common scale for comparisons of the OF and GPOMS time series. However, to avoid so-called "in-sample" bias, we do *not* apply $z$-score normalization to the mood and DJIA time series that are used to test the prediction accuracy of our Self-Organizing Fuzzy Neural Network in Section 2.5. This analysis and our prediction results rest on the raw values for both time series and the DJIA.

### 2.3. Cross-validating OF and GPOMS time series against large socio-cultural events

We first validate the ability of OF and GPOMS to capture various aspects of public mood. To do so we apply them to tweets posted in a 2-month period from October 5, 2008 to December 5, 2008. This period was chosen specifically because it includes several socio-cultural events that may have had a unique, significant and complex effect on public mood namely the U.S presidential election (November 4, 2008) and Thanksgiving (November 27, 2008). The OF and GPOMS measurements can therefore be cross-validated against the expected emotional responses to these events. The resulting mood time series are shown in Fig. 2 and are expressed in $z$-scores as given by in Eq. (1).

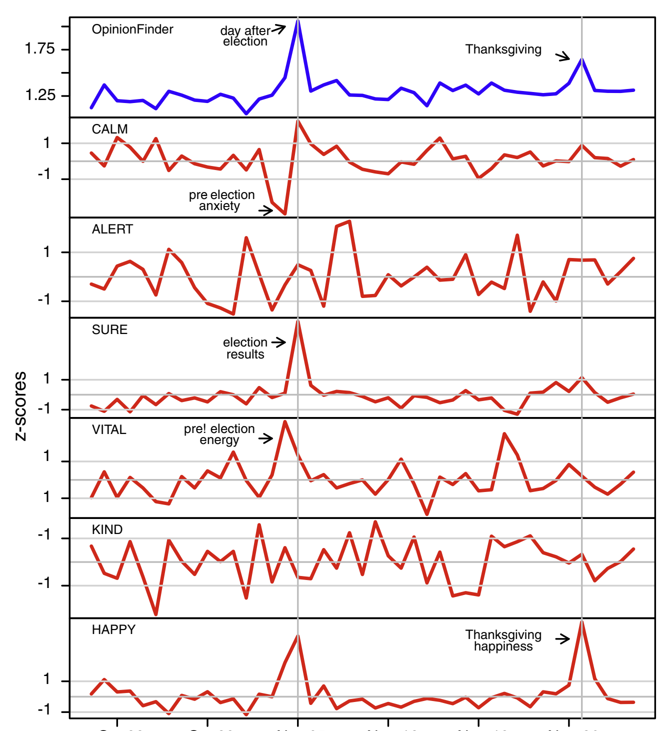


**Fig. 2.** Tracking public mood states from tweets posted between October 2008 to December 2008 shows public responses to presidential election and thanksgiving.

[7] See http://terramood.informatics.indiana.edu/data.

**Table 1**
Multiple regression results for OpinionFinder vs. 6 GPOMS mood dimensions.

| Parameters | Coeff. | Std. Err. | $t$ | $P$ |
|---|---|---|---|---|
| Calm ($X_1$) | 1.731 | 1.348 | 1.284 | 0.205 |
| Alert ($X_2$) | 0.199 | 2.319 | 0.086 | 0.932 |
| Sure ($X_3$) | 3.897 | 0.613 | 6.356 | **4.25e−08***** |
| Vital ($X_4$) | 1.763 | 0.595 | 2.965 | 0.004* |
| Kind ($X_5$) | 1.687 | 1.377 | 1.226 | 0.226 |
| Happy ($X_6$) | 2.770 | 0.578 | 4.790 | **1.30e−05**** |

| Summary | Residual Std. Err. | Adj. $R^2$ | $F_{6,55}$ | $p$ |
|---|---|---|---|---|
| | 0.078 | 0.683 | 22.93 | 2.382e−13 |

* $p<0.1$.
** $p<0.05$.
*** $p<0.001$.

**Table 2**
Statistical significance ($p$-values) of bivariate Granger-causality correlation between moods and DJIA in period February 28, 2008 to November 3, 2008.

| Lag | OF | Calm | Alert | Sure | Vital | Kind | Happy |
|---|---|---|---|---|---|---|---|
| 1 Day | **0.085*** | 0.272 | 0.952 | 0.648 | 0.120 | 0.848 | 0.388 |
| 2 Days | 0.268 | **0.013**** | 0.973 | 0.811 | 0.369 | 0.991 | 0.7061 |
| 3 Days | 0.436 | **0.022**** | 0.981 | 0.349 | 0.418 | 0.991 | 0.723 |
| 4 Days | 0.218 | **0.030**** | 0.998 | 0.415 | 0.475 | 0.989 | 0.750 |
| 5 Days | 0.300 | **0.036**** | 0.989 | 0.544 | 0.553 | 0.996 | 0.173 |
| 6 Days | 0.446 | **0.065*** | 0.996 | 0.691 | 0.682 | 0.994 | **0.081*** |
| 7 Days | 0.620 | 0.157 | 0.999 | 0.381 | 0.713 | 0.999 | 0.150 |

* $p<0.1$.
** $p<0.05$.

Fig. 2 shows that the OF successfully identifies the public's emotional response to the Presidential election on November 4th and Thanksgiving on November 27th. In both cases OF marks a significant, but short-lived uptick in positive sentiment specific to those days.

The GPOMS results reveal a more differentiated public mood response to the events in the 3-day period surrounding the election day (November 4, 2008). November 3, 2008 is characterized by a significant drop in Calm indicating highly elevated levels of public anxiety. Election Day itself is characterized by a reversal of Calm scores indicating a significant reduction in public anxiety, in conjunction with a significant increases of Vital, Happy as well as Kind scores. The latter indicates a public that is energized, happy and friendly on election day. On November 5, these GPOMS dimensions continue to indicate positive mood levels, in particular high levels of Calm, Sure, Vital and Happy. After November 5, all mood dimensions gradually return to the baseline. The public mood response to Thanksgiving on November 27, 2008 provides a counterpart to the differentiated response to the Presidential Election. On Thanksgiving day we find a spike in Happy values, indicating high levels of public happiness. However, no other mood dimensions are elevated on November 27. Furthermore, the spike in Happy values is limited to the 1 day, i.e. we find no significant mood response the day before or after Thanksgiving.

A visual comparison of Fig. 2 suggests that GPOMS' Happy dimension best approximates the mood trend provided by OpinionFinder. To quantitatively determine the relations between GPOMS's mood dimensions and the OF mood trends, we test the correlation between the trend obtained from OF lexicon and the six dimensions of GPOMS using multiple regression. The regression model is shown in Eq. (2).

$$Y_{OF} = \alpha + \sum_{i=1}^{N} \beta_i X_i + \varepsilon_t \tag{2}$$

where $N=6, X_1, X_2, X_3, X_4, X_5$, and $X_6$ represent the mood time series obtained from the 6 GPOMS dimensions, respectively Calm, Alert, Sure, Vital, Kind and Happy.

The multiple linear regression results are provided in Table 1 (coefficient and $p$-values), and indicate that $Y_{OF}$ is significantly correlated with $X_3$ (Sure), $X_4$ (Vital) and $X_6$ (Happy), but not with $X_1$ (Calm), $X_2$ (Alert) and $X_5$ (Kind). We therefore conclude that certain GPOMS mood dimension partially overlap with the mood values provided by OpinionFinder, but not necessarily all mood dimensions that may be important in describing the various components of public mood e.g. the varied mood response to the Presidential Election. The GPOMS thus provides a unique perspective on public mood states not captured by uni-dimensional tools such as OpinionFinder.

### *2.4. Bivariate Granger causality analysis of mood vs. DJIA prices*

After establishing that our mood time series responds to significant socio-cultural events such as the Presidential Election and Thanksgiving, we are concerned with the question whether other variations of the public's mood state correlate with changes in the stock market, in particular DJIA closing values. To answer this question, we apply the econometric technique of Granger causality analysis to the daily time series produced by GPOMS and OpinionFinder vs. the DJIA. Granger causality analysis rests on the assumption that if a variable $X$ causes $Y$ then changes in $X$ will systematically occur before changes in $Y$. We will thus find that the lagged values of $X$ will exhibit a statistically significant correlation with $Y$. Correlation however does not prove causation. We therefore use Granger causality analysis in a similar fashion to [17]; we are not testing actual causation but whether one time series has predictive information about the other or not.[8]

Our DJIA time series, denoted $D_t$, is defined to reflect daily changes in stock market value, i.e. its values are the delta between day $t$ and day $t-1$: $D_t = DJIA_t - DJIA_{t-1}$. To test whether our mood time series predicts changes in stock market values we compare the variance explained by two linear models as shown in Eqs. (3) and (4). The first model ($L_1$) uses only $n$ lagged values of $D_t$, i.e. $(D_{t-1}, \cdots, D_{t-n})$ for prediction, while the second model $L_2$ uses the $n$ lagged values of both $D_1$ and the GPOMS plus the OpinionFinder mood time series denoted $X_{t-1}, \cdots, X_{t-n}$.

We perform the Granger causality analysis according to model $L_1$ and $L_2$ shown in Eqs. (3) and (4) for the period of time between February 28 to November 3, 2008 to exclude the exceptional public mood response to the Presidential Election and Thanksgiving from the comparison. GPOMS and OpinionFinder time series were produced for 342,255 tweets in that period, and the daily Dow Jones Industrial Average (DJIA) was retrieved from Yahoo! Finance for each day.[9]

$$L_1: \quad D_t = \alpha + \sum_{i=1}^{n} \beta_i D_{t-i} + \varepsilon_t \tag{3}$$

$$L_2: \quad D_t = \alpha + \sum_{i=1}^{n} \beta_i D_{t-i} + \sum_{i=1}^{n} \gamma_i X_{t-i} + \varepsilon_t \tag{4}$$

Based on the results of our Granger causality (shown in Table 2), we can reject the null hypothesis that the mood time series do not predict DJIA values, i.e. $\beta_{\{1,2,\cdots,n\}} \neq 0$ with a high level of confidence. However, this result only applies to 1 GPOMS mood

[8] Gilbert and Karahalios [17] uses only one mood index, namely Anxiety, but we investigate the relation between DJIA values and all Twitter mood dimensions measured by GPOMS and OpinionFinder.

[9] Our DJIA time series has no values for weekends and holidays because trading is suspended during those days. We do not linearly extropolate to fill the gaps. This results in a time series of 64 days.