Using Social Media to Monitor Mental Health Discussions—Evidence from Twitter

TECHNICAL APPENDIX

Table A shows the final search terms for identifying tweets related to depression or suicide and terms that were excluded. Figure A shows the raw trend in tweets on Twitter from 2010 to August 18, 2014, based on the search terms for depression or suicide in Table A. The four spikes noted in the paper are evident: World Suicide Prevention Day (WSPD) in 2012, Bell's Let's Talk campaigns in 2013 and early 2014, and Robin Williams' suicide in summer 2014. There are also several smaller spikes, particularly when compared to the immediate time period surrounding each observation. For example, the time around January 2012 seems to have large deviations from the trend when compared to the December 2011 and February/March 2012. Likewise, around September 2013, two large spikes occurred (one of which likely was related to the 2013 WSPD) that could merit attention, but these spikes seem more marginal when compared with the four big spikes that were mentioned previously.

Table A: Search criteria for the depression or suicide monitor using Crimson Hexagon

Search Terms Included[1]	"depression" OR "#depression" OR "depressed" OR "#depressed" OR "mood disorder" OR "#mooddisorder" OR "suicide" OR "#suicide" OR "#suicideprevention" OR "bipolar" OR "mental health" OR "#mentalhealth" OR "#mentalillness" OR "#sad"
Search Terms Excluded	AND -bomb[2] AND -bombs AND -bomber AND -bombers AND -hamas AND -israel AND -israeli AND -palestine AND -palestinian AND -jihad AND -jihadist AND -islam AND -"ISIS" AND -"ISIL" AND -"great depression" AND -recession AND -"economic depression"

[1] The Web site http://hashtagify.me/ was used as the starting point for identifying hashtags (indicated by the symbol "#") and search terms. Search terms were not case sensitive in Crimson Hexagon. As shown in the table, search terms that were included were linked with the logical operator "OR."

[2] "-" = "not." As shown in the table, search terms that were excluded were linked with the logical operator "AND."

[Insert Figure A about here.]

In addition to the exogenous spikes, as the number of tweets grows, the size of the deviations from the trend increases. This increase in volatility can pose problems for analysis as time series modeling and forecasting rely on an assumption of standard variance throughout the time series. Therefore, to standardize the variance of this time series, we applied a logarithmic transformation to the series, as shown in Figure 1 in the main text.

A transformed time series that included data for 2010 revealed an anomaly in the series in the middle of 2010, due to issues surrounding the collection of data from Crimson Hexagon. Therefore, we excluded 2010 from our analysis. However, exclusion of data from 2010 did not substantially alter the analysis or results. Figure 1 clearly shows that after each large deviation, the time series returned to its previous levels and during other periods, seemed to fluctuate around a trend. The trend in the logarithmic series was approximately linear in the first part of the period and leveled off after 2013. This trend is most likely due to the general growth of Twitter instead of any particular growth in individuals with a vested interest in behavioral health joining Twitter or more general interest in tweeting about depression and suicide. The trending behavior of this time series indicates that it violates the stationarity assumption of time series analysis. This indicates an Autoregressive Integrated Moving Average (ARIMA) model, with a first difference, is the appropriate model. The first difference of the series serves to remove the trend and create a stationary series.

The coefficients for the (1,1,2)×(1,1,1)7 ARIMA model are presented in Table B, and accuracy statistics for the model when applied to the 2014 test sample are presented in Table C. Table C presents forecast accuracy for two other methods for comparison: a naïve mean model

Table B: Estimates for the Autoregressive Integrated Moving Average (ARIMA) model for depression or suicide tweets: 2011 to 2013

Variable	Coefficient	Standard error	t-statistic	P-value
AR(1)	0.211	0.0834	2.529	0.011
MA(1)	-0.707	0.0839	8.424	0.000
MA(2)	-0.213	0.0721	2.947	0.003
SAR(1)	0.066	0.0322	2.049	0.040
SMA(1)	-0.967	0.0092	104.8	0.000

AR = autoregressive order; MA = moving average; SAR = seasonal autoregressive order; SMA = seasonal moving average order.

Notes: The model was estimated using Twitter data from January 1, 2011, to December 31, 2013.

Table C: Model forecast performance for depression or suicide using the 2014 test sample

	ARIMA (1,1,2) × (1,1,1)7	Mean forecast	Random walk
Root mean square error	0.188	0.261	0.213
Mean absolute error	0.095	0.144	0.120
Mean percentage error	-0.063	-0.044	-0.019
Mean absolute percent Error	0.781	1.188	0.992

Note: The test sample is based on tweets data from January 1, 2014, to November 28, 2014.

and a random walk model.[1] The statistics measure the average errors of predictions versus the realized values, with smaller differences between the prediction and actual values resulting in a lower score. The mean model takes the average value of the time series as the forecast, whereas the random walk model uses the immediately preceding value as the forecast. The ARIMA model performed very well on the test sample, yielding the smallest errors for all error measures.

Figure B shows the variation in average tweet volume over the period from 2011 to 2013 by day of the week and month. Here, the weekly seasonality is much more apparent, with earlier days of the week having higher average volumes of tweets than later days of the week. To deal with this seasonality, a 7-day seasonal difference also was taken, in addition to the first differencing to create a stationary series. The transformed and differenced series, along with the series' autocorrelogram and partial autocorrelogram are presented in Figure C.

[Insert Figures B and C about here.]

The top panel of Figure C shows that the resulting series after the transformation and differencing is stationary with a regular variance throughout the series. The autocorrelation function decays quickly, with only two lags significantly autocorrelated with the series but no autocorrelation with lags 3 and 4. Even after the seasonal difference, the weekly seasonality can be seen in this panel with significant autocorrelation at lags 7 and 14. The partial autocorrelation function (PACF) graph shows this pattern more clearly, with a regular pattern in the PACF in groups of 7 days, beginning with significant autocorrelations at lags 7, 14, 21, and 28. The significant lags in both the autocorrelation function (ACF) and PACF in both seasonal and nonseasonal periods indicate that an ARIMA model is appropriate with both autoregressive and moving average terms and a multiplicative seasonal component.

In order to validate the model after estimation, it is necessary to have a test data set that is not used in estimation. To obtain the test data, we split the time series in two, with the model estimated with the data from 2011 to 2013 and tested on data from 2014. Basing model selection on the minimum Akaike Information Criterion, an ARIMA model of the order (1,1,2)×(1,1,1)7 was selected. That is, the ARIMA model consisted of one regular autoregressive order, first differenced, and two regular moving average orders, with a multiplicative seasonal component consisting of one seasonal autoregressive order, one seasonal difference, and one seasonal moving average order at the weekly seasonal period.

Figure D presents a plot of the residuals from this model, with the corresponding ACF and PACF. This figure shows that the residuals are well behaved: they are centered on zero and exhibit constant variation. Further, the ACF and PACF graphs show that there is no significant autocorrelation that is left to be explained in the residuals. Note that the two significant results at large lags can be expected due to random chance. The Box-Ljung test offers a more formal test that the residuals are white noise. A well fit ARIMA model will leave no autocorrelation in the residuals that could be explained by additional AR or MA terms; the residuals will be stochastic noise only. With a chi-square value of 32.24, the Box-Ljung test fails to reject the null hypothesis of white noise up to 30 lags, indicating that the estimated model accounts for all autocorrelation in the time series. Additionally, Figure E details a Q-Q plot of the residuals showing that they are approximately normal, particularly through the middle portion of their distribution. Taken together, these diagnostic checks indicate that the ARIMA(1,1,2)×(1,1,1)7 model is well suited for the original series.

[Insert Figures D and E about here.]

APPENDIX REFERENCE

1. Kirchgässner G, Wolters J, Hassler U. Introduction to modern time series analysis. 2nd ed. Berlin, Germany: Springer Science & Business Media; 2012.

FIGURE LEGENDS

Figure A: Daily volume of tweets mentioning depression or suicide: January, 1, 2011, to November 28, 2014

Figure B: Weekly seasonality in tweets mentioning depression or suicide: 2011 to 2013

Figure C: First differenced natural log time series of daily tweet volume for depression or suicide with autocorrelation and partial autocorrelation function plots: 2011 to 2013

ARIMA = Autoregressive Integrated Moving Average.

Notes: Panel 1 plots the difference between the natural log of each day's tweet volume for depression or suicide. Panels 2 and 3 show the autocorrelation coefficient between each day's tweet volume for depression or suicide and its own lagged values. The autocorrelation does not control for other lags while the partial autocorrelation function controls for all shorter lags. The shaded region in the autocorrelation and partial autocorrelation plots indicates the 95% confidence interval. Outside of the confidence interval, the coefficients are statistically different from 0, indicating the presence of autocorrelation.

Figure D: Model residuals of daily tweet volume of behavioral health related tweets for depression or suicide with autocorrelation and partial autocorrelation function plots: 2011 to 2013

ARIMA = Autoregressive Integrated Moving Average.

Notes: Panel 1 plots the residuals of the model, calculated as the actual values minus fitted values. Panels 2 and 3 show the autocorrelation coefficient between each day's tweet volume for depression or suicide and its own lagged values. The autocorrelation does not control for other lags while the partial autocorrelation function controls for all shorter lags. The shaded region in the autocorrelation and partial autocorrelation plots indicates the 95% confidence interval. Outside of the confidence interval, the coefficients are statistically different from 0, indicating the presence of autocorrelation.

Figure E: Normality of residuals for the Autoregressive Integrated Moving Average (ARIMA) model for depression or suicide tweets: 2011 to 2013

Notes: The normal probability plot is a plot of ordered standardized residuals from an ARIMA model against normal scores. Ordered residuals that are approximately the same as the ordered normal scores indicate that the residuals are normally distributed.