Research Proposal: A Tweet for Help? Using Sentiment and Dynamic Network Analysis to Gauge Risk of Suicide with Twitter Data

Bradley Carruthers, Supervisor: Milan Vojnovic

December 8, 2018

Context

Rationale for study

A 2014 report by the World Health Organisation states that more than 800 000 people die by suicide each year worldwide (World Health Organization, 2014) and it is estimated that depression costs the United States $210 billion alone (Greenberg, 2015). These costs of mental illness to society may be surprising to the non-expert, since as Chen *et al.* (2009) point out, there are less obvious negative externalities associated with suicide such as pyschological and financial burdens left on affected family members.

Within the last decade, a vast and growing quantity of heterogenous data has been streaming from social media platforms and their users, providing an exciting opportunity for research and the application of increasingly innovative and original methods such as dynamic network analysis and natural language processing (Gruebner *et al.*, 2017). In my thesis therefore, I plan to use dynamic network analysis to expand on the current groundwork that has been laid in social media sentiment analysis and its application to mental illness (Bollen *et al.*, 2011, Coppersmith *et al.* (2014), De Choudhury *et al.* (2013), Dodds *et al.* (2011), Glass & Colbaugh (2011), Ravi & Ravi (2015), Thelwall *et al.* (2010)).

Research Question

The exploratory research question is as follows: How can dynamic network analysis improve the identification of mental illness in social media data?

Methodology and Timeline

Data

I will scrape large quantities of Twitter data using Twitter's API.

Timeline

Michaelmas Term:

- Step 1: Summarise literature on dynamic network analysis
- Step 2: Summarise literature on sentiment analysis with social media as an application
- Step 3: Set up repository for developing code and trial-run scraping and cleaning a small dataset

Lent Term:

- Step 4: Document research question and model
- Step 5: Refine workflow and model
- Step 6: Deploy pipeline to cloud computing service for analysis on large dataset

Summer Term:

- Step 7: Document analysis, results and visualisations
- Step 8: Refine and finalise results
- Step 9: Conclude writing up of analysis and results

Limitations of Study

The main limitation of the study relates to the quality of the data. Since this would not be an experimental or quasi-experimental setting, there is little ground for answering questions of a causal nature. There is also a risk that dynamic network analysis will add very little to the existing literature and methods for social media sentiment analysis.

Organisation of Study

Chapter 1: Introduction

Chapter 2: Dynamic Network Analysis Literature Review

Chapter 3: Social Media Sentiment Analysis Literature Review

Chapter 4: Model and Pipeline

Chapter 5: Results

Chapter 6: Recommendations for further research and conclusion

References

Bollen, J., Mao, H. & Pepe, A. 2011. Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena. *In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011), 17-21 July 2011, Barcelona, Spain. Presented at the Fifth International AAAI Conference on Weblogs and Social Media.*

Chen, J., Choi, Y.J., Mori, K., Sawada, Y. & Sugano, S. 2009. Those who are left behind: An estimate of the number of family members of suicide victims in Japan. *Social Indicators Research.* 94(3):535–544.

Coppersmith, G., Harman, C. & Dredze, M. 2014. Measuring post traumatic stress disorder in twitter. *In: Eighth International AAAI Conference on Weblogs and Social Media.*

De Choudhury, M., Counts, S. & Horvitz, E. 2013. Social Media as a Measurement Tool of Depression in Populations. *In: Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13.*

Dodds, P.S., Harris, K.D., Kloumann, I.M., Bliss, C.A. & Danforth, C.M. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE.* 6(12).

Glass, K. & Colbaugh, R. 2011. Analyzing social media content for security informatics. *Proceedings - 2013 European Intelligence and Security Informatics Conference, EISIC 2013.*

45–51.

Greenberg, P.E. 2015. [Online], Available: https://blogs.scientificamerican.com/mind-guest-blog/the-growing-economic-burden-of-depression-in-the-u-s/.

Gruebner, O., Sykora, M., Lowe, S.R., Shankardass, K., Galea, S. & Subramanian, S.V. 2017. Big data opportunities for social behavioral and mental health research. *Social Science and Medicine*. 189:167–169.

Ravi, K. & Ravi, V. 2015. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*. 89:14–46.

Thelwall, M., Buckley, K. & Paltoglou, G. 2010. Sentiment in Twitter Events. *Journal of the American Society for Information Science and Technology*. 62:406–418.

World Health Organization. 2014. *Preventing suicide: a Global Imperative*. ed.