VAC Medi+board: Analyzing Vaccine Rumours in News and Social Media

Patty Kostkova, Vino Mano
UCL, London UK
{p.kostkova, vino.mano.13}@ucl.ac.uk

Heidi J. Larson, William S. Schulz
Vaccine Confidence Project, LSHTM, London UK
{heidi.larson, william.schulz}@lshtm.ac.uk

1 INTRODUCTION

Digital health has revolutionised healthcare [1-2], with implications for understanding public reactions to health emergencies and interventions [3], while real-time analysis provides a new opportunity for rapidly detecting changes in public confidence in vaccines [4, 5, 6]. *Medi+board* implements tools for infectious disease surveillance and outbreak management [7], and the novel aim of the *VAC medi+board* is to design an interactive visualisation framework integrating heterogeneous real-time data streams with social network data, to meet information needs as articulated by the LSHTM Vaccine Confidence Project (VCP) investigators.

2 BACKGROUND & RELATED WORK

Social media (SM) allows like-minded people to communicate rapidly from any location – SM can spread accurate information about outbreaks [8-9] or propagate false claims [10] that can undermine vaccination rates [11].

The public health implications of SM have become a popular research topic [12,13, 14]. However, actionable analysis and of evolving concerns requires an interactive system, rather than retrospective case-study analysis. The *VAC medi+board* dashboard fulfils two functions: (1) a flexible dashboard framework provides a visual overview of the news and SM vaccine debate, and (2) a customisable analysis of the social network (SN) actively spreading the debate on Twitter.
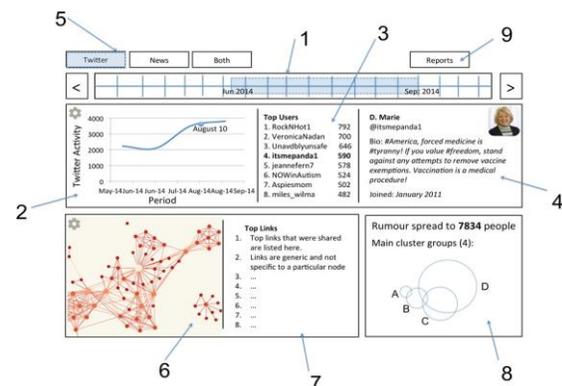
3 VAC Medi+board INTERACTIVE DASHBOARD

Figure 1. A labelled diagram indicating the design of different widgets for *VAC* Medi+Board

The *VAC Medi+board* dashboard is a customizable interactive framework allowing integration of multiple heterogeneous data streams. Figure 1 shows the modular widgets of the dashboard:
1. Global time slider selects the period for analysis
2. Plot showing activity on SM and/or news media over time
3. League table of top users for selected time period or date
4. SM user information for user selected in (3)
5. Data selector to visualise SM activity, news activity, or both
6. SM network graph of top 1000 users/tweets in time period
7. Top links/content of tweets selected in (6)
8. Overlap of community clusters of key influential users (this functionality is in development)
9. Button to generate a printable report for selected time period
10. Search field allows filtering data by keyword

3.1 Data Streams

The main information sources for the vaccination debate are SM and news coverage, though the framework is extensible allowing plugging-in of any other data streams.

The Twitter Streaming API has been set up to fetch all tweets in real-time that contain a set of vaccination debate keywords, pre-defined with domain experts of the VCP. Once we collect this data, we apply an additional filter using high frequency keywords expressing negative sentiment identified by previous studies by the VCP investigators. Currently the main systems for mining news media in real time include *GPHIN*[1] and *MediSys*[2]. For *VAC medi+board* we have used *MediSys*, accessing its regularly updated 'Vaccination' RSS feed every five minutes.

3.2 Implementation and Visualization

The dashboard is implemented as a web application, developed in Javascript and Node.js. Chart.js has been used for line graph visualisations and Sigma.js has been used for the network graphs. MongoDB is on the server-side of our application to store the SM and news data. The use of dynamic programming methodologies, incrementally pre-processing as much of the raw SM data as possible and caching the results, allows real-time processing and visualisation.

In *VAC medi+board* there are two different types of network visualisations - by tweets-as-nodes ("what information is being spread"), and by users-as-nodes ("who spreads information"). When nodes represent tweets, the size of each node reflects its retweet count, an estimate of the magnitude of spreading. When nodes represent users, node size reflects users' retweet count, and retweet relationships are denoted by directed edges drawn from each initial tweeter to each subsequent retweeter. As default n=1000, we visualise the top 1000 nodes with the highest retweet count values within the selected time period.

3.3 Demonstration: HPV 3-Month Study

To demonstrate this system, we present data collected over 3 months (the 12th of October 2015 to the 2nd of January 2016) with special attention given to the recent Human papillomavirus (HPV) vaccine scare in this period [15]. A total of 875,088

---

[1] http://www.who.int/csr/alertresponse/epidemicintelligence/en/

[2] http://medusa.jrc.it/MediSys/homeedition/en/home.html

tweets and 4,020 news articles related to vaccination were collected in this period, of which 150 news articles and 92,954 tweets referred to HPV. These are plotted over time in Figure 2, with tweets shown in red and news articles (with scale multiplied 200-fold for visibility) in green.
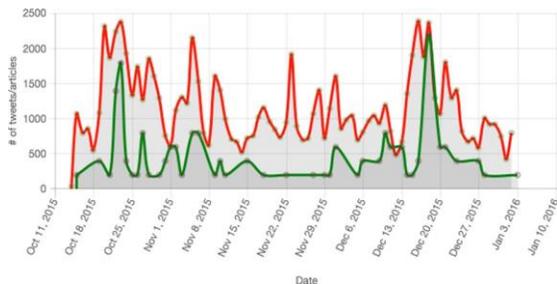

Figure 2. HPV-related tweets (red) and news articles x200 (green) plotted over time.

Two particularly robust peaks can be observed, which both relate to real events – one, in Oct 2015, coincides with widely-discussed incidents in the Colombian town of El Carmen de Bolivar, where hundreds young women experienced fainting and other side effects after HPV vaccination [5]; another, in Dec 2015, coincides with the publication of an article by a VCP investigator (HL) advocating for the HPV vaccine, which attracted both positive and negative reactions on SM. HPV-related tweets accounted for 15% and 21% of all vaccination-related tweets in the first and second peaks, respectively, compared to an average of 11% across the 3-month period.
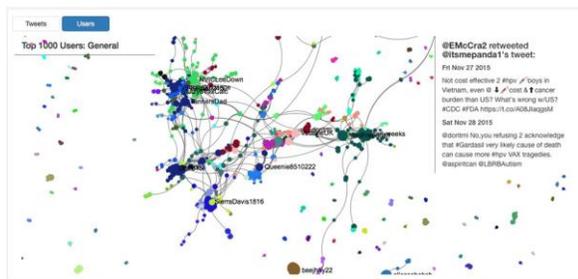

Figure 3. An example of the vaccination debate network graph where the tweets are related to HPV.

Figure 3 shows several HPV debate related clusters within the vaccination network graph, which users can browse to gain an in-depth understanding of the interlinked conversations. This is informative of how information is diffused from one user to another via the twitter network, and VCP experts confirm that the range of topics is similar to what is known about the HPV debate globally through other sources, bolstering the interpretation of the Twitter discourse as a microcosm of real-world public perceptions.

4   CONCLUSIONS

The *VAC medi+board* project is an interactive dashboard integrating SM discourse on Twitter and news coverage from the monitoring tool MediSys, covering public debate related to vaccines. The dashboard visualizes the streams, allows for drilling down to the top news, tweets, and users, and supports visualization of SN to investigate information dissemination. An analysis of HPV-related content confirms that news and twitter data can help public health experts explore vaccination phenomena, and that the depictions derived rapidly from these data are consistent with what is already known about the domain from traditional qualitative research methods.

Work in progress includes automatic identification of new emerging themes and analysis of users influenced by the key influencers, to assess the impact these individuals and organizations have.

5   ACKNOWLEDGMENTS

6   REFERENCES

[1]   P. Kostkova. Grand Challenges in Digital health. Front. Public Health.

[2]   P. Kostkova. *Public Health*. Chapter in Y Mejova, I Weber, MW Macy (Eds): "Twitter: Digital Socioscope", Cambridge University Press, UCA, 2015, pp.111-130

[3]   P Kostkova, D. Fowler, S Wiseman, J Weinberg. 'Major infection events over 5 years – how is media coverage influencing online information needs of healthcare professionals and the public?" JMIR, 2013, 15(7):e107.

[4]   World Economic Forum 2013 Global Risks 2013, Eighth Edition, Geneva, Available at http://reports.weforum.org/global-risks-2013/ [16 Mar 15]

[5]   H Larson et al. The state of vaccine confidence 2015, project report, available: www.vaccineconfidence.org (Jan 10 2016)

[6]   Andree Panisson, Marco Quaggiotto. From cortisone to graphene: 60 years of breakthroughs in PubMed publications. In the proceedings of WebSci 2014

[7]   P Kostkova et al. Integration and visualization public health dashboard: the medi+board pilot project. Prof 23rd WWW conf, pp. 657-662, 2014

[8]   M Szomszor, P Kostkova, C St Louis: *Twitter Informatics: Tracking and Understanding Public Reaction during the 2009 Swine Flu Pandemic*s, IEEE/WIC/ACM Int Conf on Web Intelligence & Intelligent Agent Technology 2011, WI-IAT, Vol.1, pp.320-323

[9]   P. Kostkova, M Szomszor, C St Louis. #swineflu: The Use of Twitter as an Early Warning Tool and for Risk Communication in the 2009 Swine Flu Pandemic ACM TMIS, SI on Informatics for Smart Health and Wellbeing, Vol 5, No 2, Article 8, July 2014, doi 10.1145/2597892

[10]   Centre for Disease Control and Prevention. Ten great public health achievements – US 1900-1999, MMWR Morbidity Mortality Weekly Report 48:241-243, 1999

[11]   C Ash, Summun bonum. Science Translations Medicine 2:61ed8, 2010

[12]   S Aral et al. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proceedings of NAS USA, 106:21544-21549

[13]   M Salathe, QV Duy, K Shashank, DR Hunter. The dynamics of health behaviour sentiments on a large online social network. EPJ Data Science 2013, 2:4

[14]   Larson et al Measuring vaccine confidence: analysis of data obtained by a media surveillance system used to analyze public concerns about vaccines. Lancet Infect Dis 13(7), pp. 606-13 2015

[15]   H. Larson et al.. Tracking the global spread of vaccine sentiments: The global response to Japan's suspension of its HPV vaccine recommendation. Human Vaccines & Immunotherapeutics. Vol 10, Iss 9, 2014, 2543-2550