COSMOS: OPENING UP SOCIAL MEDIA FOR SOCIAL SCIENCE

Twitter users now post an astounding 500 million tweets per day, while Facebook users communicate over 9 million messages per hour. It all adds up to a huge amount of interesting information. Datasets from these and other social media platforms therefore open up a fresh field of study for social scientists by offering another completely new window on community life.

Three aspects of social media records make them particularly valuable to researchers. Firstly, they give academics the opportunity to study society's reaction to events in (almost) real-time. Secondly, they can provide a much richer picture of events than simply using more traditional statistics such as crime, census or election figures alone.

And lastly, the informal, real-time reporting can offer an insight into phenomena that might remain below the threshold of official datasets. For instance, local government may monitor Twitter discussions of crimes in their area to enable them to formulate an early response to law-breaking, before it has been tracked by official statistics.

Clearly, social media offer new, exciting opportunities for scientific research. That is why COSMOS, the ‘Collaborative Online Social Media Observatory’ aims to open up data from these platforms for academic use by social scientists.

WHAT WAS THE PROBLEM?

Prior to COSMOS, a scientist who wants to use data from a social media platform would have encountered a number of problems:

1. how to obtain the information
2. how to deal with the volume of data
3. how to analyse the records

HOW WAS THE PROBLEM SOLVED?
COSMOS has largely solved these problems for datasets from Twitter. Via the Twitter API a random sample of 1% of all tweets is continuously collected by COSMOS and these are loaded into a database. From this sample the user can select the information that is relevant for his or her research topic. Next, the user can apply a number of analysis tools.

This list of tools is impressive:

- A gender detection tool which can show the gender of the tweeter
- A language detection tool to determine the language used in the tweet
- A sentiment analysis tool using natural language programming techniques to identify if the tweet is positive or negative
- A tension detection tool to assess if the tweet is using emotionally or politically charged terms
- A geospatial detection tool that makes it possible to visualize the geographical distribution of the tweets
- A social network visualisation tool that illustrates the connections between tweets, for example showing how they have been retweeted.

Using these tools, the researcher can gain a valuable insight not just into the message content, but the demographic details of those behind them.

**WHAT ARE THE ETHICAL CONSIDERATIONS?**

Ethical concerns about social media for research emerged in 2014 when a publication in the Proceedings of the National Academy of Sciences USA attracted worldwide attention. In this study, researchers conducted a large-scale experiment on the ‘contagiousness’ of emotions among Facebook users. For a number of users, positive messages from others were blocked, while a second group had negative messages suppressed. The researchers observed the effect of ‘emotional contagion’: posts by the Facebook users became more positive or negative in line with the messages they were viewing.

However, the outcry about this experiment was because the Facebook users did not consent to being part of it. Informed consent is generally seen as a condition for involvement of humans in research.

What about ethical considerations in using COSMOS tools themselves? At this moment, COSMOS only uses tweets and all Twitter users know that their tweets are public. However, Professor Rob Procter of COSMOS states that if a tweet is republished in a scientific publication, approval by the tweeter is usually sought in advance.
But he also views ethical considerations surrounding social media as part of a much broader perspective. Procter sees the potential for the development of a ‘public sociology’ in which the public will be involved in social research. Such a ‘public sociology’ would make use of crowdsourcing tools in which the public participates in coding and/or labelling tasks of large datasets. This public involvement could help ensure that researchers are trying to find answers to questions that are important to people. In this way, the use of social media messages might become much more acceptable to the public.

**KEY LESSONS LEARNED**

The COSMOS team numbers 12 active researchers, a combination of computer scientists and social scientists, but Procter emphasises the importance of social scientists taking the lead in this project. Why is this so crucial? To analyse social media records, computational tools are needed that respect social science methodology. Therefore a bridge has to be built between computer scientists who develop the computational tools and social scientists who are looking for interesting patterns.

**FUTURE DEVELOPMENTS**

COSMOS offers many research possibilities although the tool is still in its infancy: the COSMOS desktop tool will be on offer from September 2014 onwards for researchers outside the COSMOS team and will enable them to analyse relatively small social media collections. At the other end of the scale, the COSMOS Cloud will be made available in the second half of 2015 allowing researchers to create and analyse very large datasets from Twitter using the cloud.

Procter sees future developments for COSMOS along two lines. Firstly, there are plans to open up information from other social media platforms. However, negotiations with organisations like Facebook take enormous effort and Procter emphasises that he would like to see governmental involvement to support this. Secondly, a number of new analytics tools are envisaged: the crowdsourcing tool mentioned above; a tool to identify topics in tweets; and machine-learning tools that analyse relationships between social media on the one hand and conventional datasets on the other.

What about the future of COSMOS itself? The scheme came into being through a merger between the Jisc-funded Twitter Workbench and the Jisc-funded Cardiff Online Social Media Observatory projects. So far, the COSMOS project has been funded by research grants from various organisations, of which Jisc deserves a special mention because of their support in the start-up phase of the project.

However, in the near future COSMOS will develop a business model to generate income of its own. Procter foresees different service levels, varying from a free offering of Open Source software to a hosted environment that can be used immediately for a fee. There is some commercial competition, mainly software focused on sentiment analysis, but this is difficult for researchers to use because of the ‘black box’ nature of the software (i.e. the algorithms used in the software are proprietary). Nevertheless, the call for such powerful tools will no doubt increase as public use of social media continues to be ubiquitous: the quest now is to enable academics to make the most of those datasets.
FIND OUT MORE

Find out more about COSMOS on the project website

References

Photo of social media apps shared on Flickr by Jason Howie under a Creative Commons license (CC BY 2.0)

Photo of barcode arm shared on Flickr by Gareth under a Creative Commons license (CC BY NC 2.0)

Big and Broad Social Data and the Sociological Imagination: a Collaborative Response, 2014; Housley et al; Big Data & Society July-September 2014 vol. 1 no. 2; DOI: 10.1177/205395171454513 available online here


Experimental evidence of massive-scale emotional contagion through social networks; Adam D.I. Kramer et al (2014) PNAS June 17, 2014 vol. 111 no. 24; DOI: 10.1073/pnas.1320040111 available online here

Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack; P. Burnap et al; Soc. Netw. Anal. Min. (2014) 4:206 available online here

Based on an interview with Professor Rob Procter from Warwick University by Maurits van der Graaf (July 15, 2014)

August 2014

With thanks to Professor Rob Procter, and to Maurits van der Graaf (interview and text) and Nicola Yeeles (text editor).
This work is made available under a Creative Commons Attribution 3.0 Unported licence.
Knowledge Exchange is a co-operative effort that supports the use and development of Information and Communications Technologies (ICT) infrastructure for higher education and research.

www.knowledge-exchange.info