

Can *Flickr* activity predict elections?

Tobias Preis, *Warwick Business School* (Tobias.Preis@wbs.ac.uk)

Suzy Moat, *Warwick Business School* (Suzy.Moat@wbs.ac.uk)

Mammoth amounts of data are now being generated through society's extensive interactions with technological systems, automatically documenting collective human behaviour in a previously unimaginable fashion (1-3). Analysis of such *big data* opens up new opportunities for a more precise and extensive quantification of real world social phenomena that was difficult to attain using complicated and expensive surveys and laboratory experiments alone.

A particularly fruitful area of research has focused on the analysis of queries to search engines and other online information services such as *Google* and *Wikipedia*. Strong links have been found between changes in the information users are seeking online and events in the real world, ranging from reports of flu infections (4) to stock market movements (5-6). A recent study has shown that Internet users from countries with a higher per capita GDP are significantly more likely to search for information about years in the future than years in the past (7).

The primary goal of this mini-project is to expand on findings identifying links between behaviour online and the real world (4-8) to analyse election outcomes. This includes analysing the relationship between activities of *Flickr* (9) users around elections and election outcomes. A particular focus is expected on analysis of a dataset of votes for the US Congress.

Previous preliminary experience in the following three areas is of advantage to complete this mini-project successfully:

- (1) *basic programming skills and experience with automated data download;*
- (2) *experience with statistical programming packages;*
- (3) *writing reports for an academic audience.*

This mini-project can be the basis for a PhD project in this area. Possible directions include incorporating other data sources such as *Google*, *Wikipedia*, *Twitter*, news archives and others and developing agent-based models of this complex human behaviour.

Please do not hesitate to contact us if you have questions about this mini-project.

References:

1. Moat, H. S., Preis, T., Olivola, C.Y., Liu, C. & Chater, N. Using big data to predict collective behavior in the real world. *Behav. Brain Sci.* (in press).
2. Lazer, D. *et al.* Computational Social Science. *Science* **323**, 721 (2009).
3. Vespignani, A. Predicting the Behavior of Techno-Social Systems. *Science* **325**, 425 (2009).
4. Ginsberg, J. *et al.* Detecting influenza epidemics using search engine query data. *Nature* **457**, 1012 (2009).
5. Moat, H.S., Curme, C., Avakian, A., Kenett, D.Y., Stanley, H.E. & Preis, T. Quantifying Wikipedia usage patterns before stock market moves. *Sci. Rep.* **3**, 1801 (2013).
6. Preis, T., Moat, H.S. & Stanley, H.E. Quantifying trading behavior in financial markets using Google Trends. *Sci. Rep.* **3**, 1684 (2013).
7. Preis, T., Reith, D. & Stanley, H. E. Complex dynamics of our economic life on different scales: insights from search engine query data. *Phil. Trans. R. Soc. A* **368**, 5707 (2010).
8. Preis, T., Moat, H. S., Stanley, H. E. & Bishop, S. R. Quantifying the Advantage of Looking Forward. *Sci. Rep.* **2**, 350 (2012).
9. <http://www.flickr.com/services/api/flickr.photos.search.html>