



AQMeN (Applied Quantitative Methods Network)

Collaborative Small Grants Scheme

END OF AWARD REPORT

Social media and the 2014 Scottish Independence

Referendum: The Twitter dimension

Mark Shephard, Stephen Quinlan, Stephen Tagg

University of Strathclyde

Lindsay Paterson

University of Edinburgh

February 2014

Table of Contents

The project	3
Overview	3
Methodology.....	3
Activities	5
Analysis.....	5
Findings/ Outputs.....	5
A: Project beneficiaries	7
B. Project impact.....	7
C. Collaborative components.....	9
D. Innovative use of advanced quantitative methods.....	9
E. Knowledge exchange value of the report	9
F. Attached documents and resources	10

The project

Overview

This small grant enabled our existing ESRC/AQMeN funded project exploring social media platforms in the Scottish independence referendum to broaden out and focus on other social media channels besides online discussion forums – in this instance Twitter. Online discussion of political matters is today at an unprecedented level with Twitter providing a core forum for political discussion in ‘real time’. There are estimated 500million active Twitter accounts worldwide today, with 30million alone in the United Kingdom, demonstrating the potential reach of this medium.¹ But because of the relative newness of Twitter, investigations exploring its impact on politics, in particular referendum campaigns are few and far between. This project seeks to fill this void and provide new perspectives and knowledge by examining Twitter content relating to the 2014 Scottish Independence referendum campaign. Our central objective is to assess the discourse on Twitter, an unmoderated forum of discussion and compare it to a moderated forum, namely the BBC HYS discussion forums of the Scottish independence referendum (which we were already in the process of monitoring). We aim to explore the content of interactions in terms of what issues are being discussed, and what sentiments are expressed in respect of independence, political parties, national identity, and politicians.

As a result of a decision to extend our data collection period² to ensure more data and greater variability, thereby enhancing the validity of our research, we are only in a position to provide some preliminary results based on a portion of the data at the time of submission.³ But preliminary conclusions would seem to suggest that the two different forums do differ on many aspects including the amount of focus on issues, which politician is mentioned more, and the level of profanity seen. However, there are similarities in our samples on other metrics including the amount of overtly flaming behaviour displayed.

Methodology

Content and sentiment analysis

Our research design was based on a content analysis of a sample of tweets using the hashtag ‘#indyref. This involved devising a coding frame to measure the content of tweets on a range of

¹ Source: http://www.mediabistro.com/alltwitter/twitter-top-countries_b26726 (Accessed: 30 May 2013).

² We had hoped that this small grant would act as a seed corn proposal to engage in more extensive analysis of Twitter into 2014 as part of our wider ESRC funded project exploring social media platforms in the referendum. However, due to funding limitations this has not been deemed possible. As such, this informed our decision to extend our data collection.

³ Data is based on our analysis of tweets on 26 November 2013.

different dimensions (see the analysis section for more). The coding frame was designed to ensure the maximum level of comparativeness with a similar content analysis of the BBC HYS discussion forums that as had already been conducted. The coding was carried out by one member of the team (the graduate student) between November 2013 and January 2014. To ensure reliability and replication, intra-coder and inter-coder tests (for e.g.: Neuendorf, 2002; Hayes and Krippendorff, 2007; Krippendorff, 2013) involving two other investigators were conducted on all relevant variables with the results blind of the original coder and of the other inter-rater tester. Using Krippendorff's κ as our base measure of reliability, preliminary tests suggest acceptable levels of reliability between coders. In the future, we intend to move beyond the content analysis and apply sentiment analysis to the data.

Case selection

Our starting point was to focus on tweets using the most-common referendum related hashtag '#indyref' on a series of key dates in the referendum campaign and where a corresponding BBC HYS online thread would also be discussing Scottish independence. We took the two most recent key dates in the referendum campaign at the time of making our decision for analysis, namely: 1) 18 September 2013 – one year to go until the referendum in 2014, and 2) the day on which the Scottish government launched its White Paper for an independent Scotland (26 November 2013).

Sample

Our next step was to devise a mechanism to gather a sample of tweets. As our interest is to compare the flavour of discussion on twitter vis-à-vis the flavour on the BBC HYS discussion board, we decided to concentrate only on tweets during the period in which the relevant BBC HYS discussion board was 'live'.

For the data gathered on 26 November 2013, we used the R library 'twitterR' as well as other packages including 'ROAuth', 'RCurl', 'bitops', 'digest', and 'rjson' and devised code which allowed us to interact with an API server and provide us with a selection of tweets. This procedure yielded approximately 5,000 tweets for analysis. Tests of the sample suggest that we have sufficient representation of tweets corresponding to the whole time frame we are interested in.

However, it was not possible to utilize this approach to obtain data for 18 September 2013 given this procedure can only select samples from a relatively recent time point. Gathering of tweet data is complicated by the fact that tweet data over time is archived and can only be retrieved through professional organizations, which can in cases involve substantial fees. To gain

a selection of twitter data for the 18 September 2013, we instead approached *topsy.com*. Through a trial license, they provided us with a full sample of tweet data using the '#indyref' at their disposal for that date which yielded in excess of nearly 5,000 tweets. Due to time limitations on our project, we had to select a random sample of these 5,000 tweets for inclusion in our content analysis, which yielded an n of 1,500.

While we do not claim that our analysis provides a complete picture as to the discourse on Twitter more generally relating to the Scottish referendum, we argue that our contribution provides us with a flavour of some Twitter interactions which we then proceed to compare with the interactions we observed on the BBC HYS forums.

Activities

1. Decision on mechanisms to grab Tweets & devising of necessary 'R' code to do so: September-October 2013.
2. Two key dates in the campaign selected followed by data collection: October/November 2013.
3. Coding frame devised: October 2013-November 2013.
4. Data collation and creation of dataset: November 2013-January 2014.
5. Data analysis: January 2014 (preliminary only).

Analysis

Our analysis thus far has been based on a content analysis with each individual tweet as the unit of analysis. We then analysed each tweet on a number of dimensions including:

1. Its sentiments towards independence.
2. Mentions of political parties and politicians and the tone expressed towards them.
3. The tone of the tone (excessive use of punctuation/CAPITALIZATION as indicators of flaming) and use of profanity.
4. Level of policy focus of the tweet.

It is our intention to apply sentiment analysis (we are currently running this on the BBC HYS data for the main grant).

Findings/ Outputs

Findings

Thus far, our conclusions are based on a preliminary analysis of a portion of the data (see footnote 2). Some emerging trends include:

1. *Issue discussion:* As expected contributions to online discussion boards are much more likely to be policy issue driven than contributions on Twitter. Whereas 43% of comments made in our analysis of BBC HYS forums are associated with policy, only 18% of tweets in our sample could be classified as policy related. The economy though in both forums was the issue that was the foci point of discussion.
2. *Tone of discussion:* There would appear to be similar levels of flaming in both forums as measured by the amount of CAPITALIZATION or excessive punctuation used in both, which is low overall. However, Twitter does seem to contain more profanity compared to the BBC discussions, which is to be expected given that Twitter is an unmoderated forum and the BBC is a moderated one.
3. *Politicians:* It would appear that politicians are mentioned to the same extent in both our samples – about 1/5 of tweets and 1/5 of comments in the BBC HYS forums mention a politician. Alex Salmond is the most mentioned in both but we do see evidence that Twitter mention of politicians is more diverse than what we observe in the BBC HYS forums.

Outputs

1. *Knowledge exchange:*

- Forthcoming academic article detailing our results – submission made to American Political Science Association Conference – Washington DC August 2014.
- Forthcoming blog post outlining our preliminary findings – due June 2014.

2. *Data and quantitative training:*

Two datasets of a random selection of tweets associated with the #indyref on two key dates in the referendum campaign thus far, namely the one year anniversary until the referendum (18 September 2013, n=approx. 1,500) and the launch of the White Paper by the Scottish government (26 November 2013, n=approx. 5,000).

This dataset will be made publically available at the completion of all analysis of the research at the end of 2014.

3. *Career development:*

- Mark House (MA graduate) obtained skills in data collection, collation and inputting, data analysis including reliability tests and enhanced their programming experience of SPSS, Microsoft Excel, and their ability to conduct content analysis.

- Stephen Quinlan (Postdoctoral fellow) and Mark Shephard (PI on overall project) enhance their knowledge and skills of Twitter and programming skills in 'R'.

A: Project beneficiaries

A1: Graduate student on the project

MA graduate student obtained skills in data collection and sampling, collation and inputting, data analysis, and enhanced their programming experience of SPSS.

A2: Academic and research staff on the project

Staff gained further experience of Twitter, sampling of tweets and 'R' packages. They also produced an academic article.

A3: Academic community at large.

The creation of a large-scale new dataset encompassing Tweets associated with the '#indyref' will be of benefit to the academic community at large given that it will be made available for secondary data analysis upon completion of our own analysis.

A4: AQMeN

Data emanating from this project will form the basis of the AQMeN workshop on using the 'R' programme in April 2014. We also are scheduled to contribute to the AQMeN spring 2014 newsletter announcing our result's findings.

B. Project impact

As we promised, we did use this pilot as the basis (in large part) for a grant extension application that was rejected by the ESRC.

We expect a number of further impacts through this project:

- 1) We want to use our findings to further inform the educational materials we have been producing for schools and colleges on how to engage with social media.
- 2) April 2014 – we are running an AQMeN workshop on content analysis of social media that will include data and exercises derived from this funding
- 3) Conference – academic dissemination we applied to the American Political Science Association for their Aug/Sept 2014 conference in Washington DC.

- 4) Academic article (contrasting these Twitter findings with comparable time measures we have collected and analysed from the BBC's Have Your Say threads on independence news stories)
- 5) Media (blog + TV/radio derived from 3 and 4 above)
- 6) AQMeN events: seminar and newsletter

C. Collaborative components

Collaboration on this project came in the form of co-operation between the two universities (Strathclyde and Edinburgh) and also the cross-disciplinary nature between colleagues (Shephard, Quinlan, and Paterson – political science; Tagg – Marketing).

D. Innovative use of advanced quantitative methods

Using mixed-methods approaches (NVivo, SPSS and R-tm) to analyse the content and sentiment of tweets we are building upon existing measures designed for the quantitative analysis of social media on the independence referendum. These will be demonstrated in a hands-on AQMeN seminar in April 2014.

E. Knowledge exchange value of the report

As mentioned in our application (see under 'Timetable for completion of the project'), KE activities and subsequent publications are designed to coincide with the working schedule of the main ESRC social media grant which complements and integrates into this project. As such, this project was preceded by the schools and universities experimental projects and the BBC Have Your Say content and sentiment analysis project. Consequently, we are to deliver on this project over the coming months:

- 1) We want to use our findings to further inform the educational materials we have been producing for schools and colleges on how to engage with social media.
- 2) April 2014 – we are running an AQMeN workshop on content analysis of social media that will include data and exercises derived from this funding
- 3) Conference – academic dissemination we applied to the American Political Science Association for their Aug/Sept 2014 conference in Washington DC.
- 4) Academic article (contrasting these Twitter findings with comparable time measures we have collected and analysed from the BBC's Have Your Say threads on independence news stories)
- 5) Media (blog + TV/radio derived from 3 and 4 above)
- 6) AQMeN events: seminar (TBA) and newsletter
- 7) Contact parliaments and youth parliaments and relevant groups to disseminate core findings (TBA)

F. Attached documents and resources

The full comprehensive dataset will be deposited with AQMeN along with all relevant syntax files from any publications arising upon the completion of analysis and confirmation of publication.