Social Media for Election Communication and Monitoring in Nigeria

PREPARED BY DEMOS FOR THE DEPARTMENT FOR INTERNATIONAL DEVELOPMENT

JAMIE BARTLETT, ALEX KRASODOMSKI-JONES, NENGAK DANIEL, ALI FISHER, SASHA JESPERSON

DEMOS

Published by Demos March 2015

Unit 1, Lloyds Wharf
2 – 3 Mill Street
London SE1 2BD

hello@demos.co.uk
www.demos.co.uk

CONTENTS

EXECUTIVE SUMMARY

Nigeria has witnessed an exponential growth in internet and social media use. From a modest 200,000 users in 2000, by 2015 around 30 per cent of the population is online, increasingly on smart phones. The use of social media in elections initially became noticeable in the preparations for the 2011 Nigerian elections, and now receives widespread media attention for its role in informing, engaging and empowering citizens in Nigeria and across Africa.

Social media activity presents a novel way to research and understand attitudes, trends and media consumption. There is a growing number of academic and commercial efforts to make sense of social media data sets for research or (more typically) advertising and marketing purposes.

This project examines the potential of social media for monitoring and communication purposes, using the 2015 Nigerian elections as a case study. The purpose of the research is to develop an understanding of the effectiveness of social media use for communication and monitoring during the 2015 Nigerian election, and draw out lessons and possibilities for the use of social media data in other elections and beyond.

Methodology

Over the period 18 March – 22 April 2015, researchers collected 13.6 million tweets posted by 1.38 million unique users associated with the Nigerian Presidential and State elections held in March –April 2015. Only English language tweets were intentionally collected, using English language search terms. We also collected data – posts and interactions – from 29 election relevant public pages on Facebook. For both data sets we undertook a series of automated and manual types of analysis, in four phases. First we determined the nature and type of data available. Second, we analysed the demographic details of those who posted content. Third, we constructed network maps to illustrate the relationships between users. Finally we compared social media data with CASE2015 (Content Aggregation System for Elections), which is an election monitoring system that gathers reports about the conduct of the election through social media, SMS and apps. We compared our data set to 796 reports received through SMS and apps during the election period.

Findings

Nature and type of data available

We found that Twitter was ten times more active over the election period than at 'normal' times. We generated 12.4 million tweets about the elections over the period; and these tweets tended to be divided into 'reportage' (i.e. people describing events) and 'comment' (i.e. people commenting on events). By using automated classification, it was possible to break the data down into further useful categories. For example there was a significant volume of tweets about violence (408,000). Much of the top content was widely shared news reports or campaigning – which demonstrates the value of more nuanced analysis of less popular content. While there was a significant volume of rumours being spread on Twitter, we found relatively few cases of ethnic or racial slurs being used.

Demographics of users

There were 1.38 million unique Twitter users posting content about the election on Twitter, and (in our data set) 216,000 Facebook users interacting with content on popular public Facebook pages. Within the Twitter data, seven of the 10 most popular accounts (in terms of mentions or retweeted content) were media outlets. When compared against known demographics of social media users, it is possible to draw a number of general conclusions about how representative users were of the broader population. Although internet and social media penetration is growing quickly – especially via mobile phones – it still represents a small proportion of the whole population:

- Awareness and use rates are much lower among older and less educated Nigerians.
- Men are significantly more likely to use social media than women.
- Although breakdowns of social media penetration by region are not available (we have not been able to locate any), the north of the country has less internet availability than the south. This suggests there will be an intrinsic bias in social media data towards southern states.
- There was a significant proportion of 'organisational' accounts in the Twitter data set.
- 2.91 million tweets were identifiable as being from Nigeria (71 per cent of those which included some location data). Of the 4.1 million tweets that had some location data, 1.14 million came from users in Lagos and 454,000 from Abuja. Only around one per cent of all tweets included a precise latitude-longitude geo-tag.

Specifically in relation to Nigerian social media users, although not a perfectly representative sample, it appears there were a number of quite different types of users on social media: citizens, citizen-journalists, media bloggers, official news outlets, NGO / CVO, political campaigners and supporters.

Networks of users

Different networks (for example conversations about violence and those about problems voting) revealed different accounts are influential, which illustrates the value of building theme based networks. On Facebook, users tended to like or comment on only one post. A smaller group of active users who commented on posts tended to be quite loyal, and more likely to engage with a single candidate's page than on several. Twitter accounts supported by DfID clustered closely together, suggesting they are reaching a similar group of other users rather than across the network as a whole.

Comparison with CASE data set

It is very difficult to directly compare CASE 2015 and Twitter data in relation to electoral misconduct. The majority of cases of polling misconduct data found on Twitter was not recorded by CASE, although this partly because CASE collects information about events at polling stations, while Twitter data is far broader. Most of the CASE incidents were found on Twitter, although they were not easy to find. We found that Twitter offers a far wider variety of data about electoral misconduct (and is not limited to polling stations); and in far greater volume. Its value is as a source of citizen-led reporting that can be either investigated further, or used to provide more colour and context to existing methods of reporting.

Implications

The research was able to find that there was a very large volume of potentially useful data available on Twitter that was not available through other sources. Much of the content is popular, viral content – often news related – while the potentially more useful 'on the ground' information is often buried beneath the major news stories. Based on this pilot, we consider there are five broad capabilities that social media and social media monitoring can provide:

- Better understand the network of influencers / journalists / commentators.
- Ability to detect and characterise unexpected events quickly as they occur (for example, violence).
- A source of supplementary data about electoral misconduct.
- Track and respond quickly to rumours or misinformation.
- An evaluation tool to complement existing evaluation systems.

Strengths and weaknesses of social media data

Social media data has certain strengths (relational, real or near real time, reactive, voluminous) and weaknesses (unpredictability, lacking in measurable demographic variables, reliability, uncomprehensive). Generally speaking, we think social media offers a valuable supplementary source of insight and information, alongside traditional (and better established) sources of research such as polling or focus groups. It is imperfect, but still useful if used for the right purposes.

What are the methodologies that prove effectiveness?

Automated data analytics, key word sampling and classifiers create a number of new methodological challenges.

The challenge of this technology is that performance can be highly differential. In some cases, automated 'classifiers' that filter or categorise large data sets of this type perform extremely well in accurately classifying data sets. In other cases, for example with more event-specific data (such as 'electoral violence', 'other violence' or 'accreditation'), it is less accurate.

Research ethics are an increasingly important consideration when undertaking research or 'monitoring' using social media. While such research should not be undertaken without particular caution and consideration, research without informed consent can be justified when no details about an individual are likely to be divulged, and where the risk of harm to research subjects is fully minimised. Even with open source data, however, certain conditions still ought to be met. Any use of this type of technology or monitoring capability should be done with careful consideration of its ethical use; ideally with reference to GSR's Professional Guidance 'Ethical Assurance for Social Research in Government'.[1]

Recommendations

How social media should be used for election monitoring / voter education (for Nigeria and elsewhere)

Based on this pilot study, we recommend that social media would be best used in a number of ways for election monitoring and education. Some of these are not limited to election monitoring, and can be of use outside elections too:

Pre-election / strategic understanding of environment

- Identify potentially influential voices and accounts which might be important to engage with or listen to.
- Given the likely presence of social media in future elections in Nigeria and elsewhere, social media clearly provides an excellent opportunity for citizens to contribute to election monitoring. Campaigns in the lead-up to elections or other flash points to educate people on how to use social media to

monitor / communicate (for example, using certain hashtags, accounts, adding location data) could increase the volume of citizen engagement in the election process.

During-election / tactical insight

- Identify and understand emerging events. 'Event' in this case does not only mean a physical 'real world' event; it can also refer to a purely digital event, such as rumours.
- Supplement existing media monitoring research work, by analysing social media to identify multimedia citizen generated information about events.
- We believe that automated Twitter analysis can, to a high degree of accuracy, identify citizen reports about electoral misconduct. This can be either a) new, undocumented events or b) colour and detail for already reported events.

Post-election / research and evaluation

- Analysing these data sets in the way set out above would allow researchers to identify the most common complaints or concerns about electoral misconduct, or better determine citizens' experience of voting. Twitter in particular offers a way to assess citizen views on a scale hitherto not possible. That can be used, with accompanying demographic data and caveats, as a valuable research tool.
- Better understand the possible reach and activity of certain organisations or movements – particularly those directly supported by DfID.

Practical steps

In the event that DfID considers commissioning a capability (or alternatively a consultancy arrangement, where insight is delivered) in order to use social media research for election monitoring and education, the following steps would be valuable:

- Invest in developing internal expertise in how social media monitoring works, ideally combining a number of disciplines, most principally machine learning, computer programming, and social sciences.
- Review the current range of social media monitoring tools currently available and the current API access rights to the major social media platforms. Determine which of these capabilities can be satisfactorily done using free software in house. Others might be better achieved with academic partners which would allow for great flexibility in application.

- Any tenders written or commissioning decisions taken should include input from individuals with a good knowledge of the strengths and weaknesses of social media monitoring software and methodology. We recommend a healthy level of scepticism about commercial offers – especially those that are not open and transparent about every stage of the research / monitoring process.
- Carefully review all research ethics guidance likely to be relevant for this type of work, and set out broad guidelines to ensure any activity is conducted to existing ethical research standards.

1 BACKGROUND

There has been a change over the last decade in the way people access, consume and produce media: a shift away from mainstream media and toward internet-based content and social media. Thirty per cent of Nigerians use the internet – of which 70 per cent are using social media (Facebook, YouTube and Twitter all count in the top ten most visited sites in Nigeria). This is changing the way people get their news, and learn about issues.[2]

Social media activity presents a novel way to research and understand attitudes, trends and media consumption. There is a growing number of academic and commercial efforts to make sense of social media data sets for research or (more typically) advertising and marketing purposes.[3] From the inception of Ushahidi to collect and map reports of violence during the post-election period in Kenya in 2007, to the reliance on Twitter during Iran's 2009 elections, social media platforms have become important tools to track and map irregularities and violence, but also for communication beyond one way messages from leaders to the people. As yet, very little analysis focuses on the role and effectiveness of social media in election communication and monitoring.

This project examines the potential of social media for monitoring and communication purposes, using the 2015 Nigerian elections as a case study. The purpose of the research is to develop an understanding of the effectiveness of social media use for communication and monitoring around the 2015 Nigerian election, and draw out lessons and possibilities for the use of social media data in other elections, and beyond.

This report contains the following sections. First, we present a rapid review of existing research work on the subject. This covers both elections and social media use in Nigeria, and any research which examines the potential of social media for election monitoring and communication. Second, we set out the methodology employed. The types of analysis undertaken on the data were not always precisely what had been outlined in the original research proposal, but rather were driven by the available data. (Any changes were agreed in advance with DfID.) Third, we set out the results in full, divided into Facebook and Twitter results. These are presented in a way which reflects the original research questions. Fourth, we discuss the implications of these results according to a number of specific research questions asked, and propose a number of recommendations.

Literature review

Introduction

Nigeria has witnessed an exponential growth in internet usage. From a modest 200,000 users in 2000, now an estimated 51 per cent of the population use the internet.[4] There are a total of 186,410,197 active mobile lines in Nigeria as of February 2015 according to the Nigerian Communications Commission,[5] a twofold increase from the 93 million reported in 2011.[6]

Much of this increase is driven by a growth in mobile web access. A Gallup poll from 2012 found that almost 73 per cent of Nigerians owned a mobile phone.[7] That figure is now expected to be over 80 per cent. The Mobile Africa 2015 study, which surveyed 3,500 mobile users in five countries across Africa, reported that 47 per cent of Nigerians used their phone to access the internet.[8]

While social media use has increased, it remains fragmented. According to We Are Social, a London-based social media communications agency, Facebook penetration in Nigeria only stood at six per cent in 2014, around 11 million users.[9] This figure, however, has been growing rapidly.[10] In 2014, the Broadcasting Board of Governors reported that, while two thirds of Nigerians have heard of social networking services, just 29 per cent of Nigerians had used at least one platform in the past week. Awareness and use rates are much lower among older and less educated Nigerians – 51 per cent of those 35 or older have heard of social networking services and just 13 per cent of this age-group said they used such a service in the week the past week.[11] A 2015 study of social media users in Makurdi, the capital of Benué state, found younger people tend to make up the bulk of overall internet users. The study also reported that use of social networking sites within the sample was greater among men (70 per cent) than women (28 per cent), as well as finding some correlation between levels of education and use of the internet/social media.[12]

Social media is used for lots of reasons. On Facebook, the most popular pages in Nigeria include Kaymu (an online marketplace), Naij and Information Nigeria (news agencies), Pastor Enoch Adeboye (a Christian preacher), P-Square and Young Paperboyz (popular Nigerian musicians), Omotola Jolada (a Nollywood actress) and Goodluck Jonathan.[13] The most popular Twitter accounts, by contrast, are almost exclusively Nigerian musicians and music producers. The most popular politician on Twitter (with 455,986 followers at the time of writing) is Babatunde Fashola, the former governor of Lagos state, who was much acclaimed for his work on tackling traffic problems, crime and poor infrastructure.

The elevated position of social media in Nigerian society and public life can also be seen from the changing nature of news websites. The third most visited site in Nigeria, Sahara Reporters, relies heavily on reporting by citizen-journalists for its content and has been at the forefront of publishing multimedia content on social platforms including Twitter, Facebook, Instagram, Tumblr and YouTube. Its popularity as a news platform (with over 1.5 million likes on Facebook) is testament to the influence that social media can have.[14]

Aside from more mainstream social media platforms (Facebook, Twitter, YouTube), Nigerians have a strong presence on platforms such as 2go, a South African social networking site, and Eskimi, a mobile social network and media platform (with a reported nine million members as of 2014).[15]

Social media and elections in Nigeria

Relative to other countries in the region, Nigeria has a long history of social media activism.[16] One example is the protests staged in January 2012 against the government's announcement of the removal of Nigeria's fuel subsidy, which resulted in a 120 per cent increase in the per litre pump price of petrol. The announcement provoked a series of demonstrations across the country and internationally, both on the streets and online using the hashtag '#OccupyNigeria'. The episode may have played a role in the subsequent re-instalment of the subsidy payments by Jonathan's government.[17] A second example of social media's political influence in Nigeria concerns the reporting of the Islamist militant group Boko Haram. The #BringBackOurGirls Twitter campaign, initially started by Nigerian lawyer Ibrahim Abdullahi, gained international attention. The hashtag alone has been used in more than 4.5 million tweets globally since the campaign began. The issue of civilian security and terrorism in northern Nigeria subsequently became a major part of election campaigning for the All Progressives Congress (APC) parliamentary candidate General Buhari.

The use of social media specifically during the elections first became noticeable in the preparations for the 2011 general elections. In a review of these elections, the Policy and Legal Advocacy Centre documented at least three main ways in which Nigerians were using social media. First, to share information relating to the elections. This included the development of novel technologies that allowed people to access data and information in real time. One example was *Revoda,* a mobile application which enabled a parallel vote count, access to polling unit results, transmission of collected results and additional information about the entire electoral process.[18] Second, social media platforms were used by political parties, candidates and governmental organisations for campaigning and raising awareness. The Independent National Electoral Commission of Nigeria (INEC) used the opportunity to develop its communication channels and engage with citizens through Facebook, YouTube and Twitter. INEC's Situation Room was established, enabling people to directly contact the organisation to report misconduct and concerns about the poll. The Commission received about 4,000 tweets in the three days during the presidential election. Finally, Nigerians used social media "to improve the efficiency of election observation".[19] Citizens were able to share information and pictures such as results from their polling units. Although this may not have prevented malpractice and falsification of results, it meant that the public was aware of the trends in different locations and more likely to challenge any falsified results. Civil society organisations were also instrumental in leading campaigns for transparency and accountability, as demonstrated by projects such as *Reclaim Naija,* an election incident reporting system that allowed feedback to be easily aggregated and analysed. This allowed Nigerians to report incidents of violence and electoral malpractices through text messages. Between the National Assembly elections of 9 April 2011 and the presidential election of 16 April 2011, citizen observers submitted 6,000 incident reports to the platform. Another project, The Social Media Tracking Centre, harvested social media reports from the elections before mapping incidents and monitoring the process of the polls over time. [20] At the end of that election, the INEC's chair Attahiru Jega stated that the use of social media enhanced transparency in the electoral process and made the INEC more accountable to the public in the conduct of elections.

By 2015, citizen journalism and observation were often finding their way into the mainstream news as media organisations increasingly invited their subscribers to report on online platforms.[21] One noticeable feature was the expanded use of hashtags as flashpoints for political discussion and advocacy. On the eve of the 2015 elections, between 40 to 50 active hashtags linked to Nigerians actively discussing the elections were identified.[22] 'Hashtagging' in this way also became a way of identifying political affiliation and support for candidates among the electorate.

2015 also saw an increase in the use of social media by political parties. For example, StateCraft, a Lagos-based communications company, was responsible for APC candidate Muhammadu Buhari's digital drive intended to appeal to younger people.[23] President Goodluck Jonathan appointed Obi Asika, the chair of Social Media Week Lagos (an international conference focused on change in social media technologies) as his Senior Special Assistant on Social Media. Political parties have also branched out in to other mediums to engage voters. Both front running parties staged 'Google Hangouts', in which candidates answered questions from young Nigerians. The APC also tried to crowdsource funding using a mobile platform, designed to tap in to the social media networks of its supporters to raise money for campaigns.

The importance of social media extended beyond polling day. Following the presidential inauguration, Nigerians posted tweets that included the hashtag #BuhariFixThis to offer their suggestions for the priorities of Buhari's first term in office. The Centre for Democracy and Development West Africa also developed an app, 'the Buharimeter' designed to track the progress of electoral promises and provide a forum for political discussion. Civic technology organisation BudgIT began a social media campaign #OpenNASS, which calls for transparency and publication of the full details of the expenditure by the national assembly to encourage openness in the new government.

There are examples from elsewhere in Africa of how social media is affecting politics in similar ways. In Kenya, the Ushahidi platform that was established after the 2007 election violence was instrumental in collating and mapping citizen reports of electoral misconduct, receiving 45,000 visits to its website. The success of the initiative resulted in the launch of Uchaguzi in 2013, a programme designed to repeat citizen electoral monitoring for the Kenyan presidential elections. The website recorded over 3,000 incident reports in the days surrounding the elections, which included nearly 400 security reports and issues of voting irregularities, registration problems and polling station difficulties.[24] In 2012, Senevote was developed by the Senegalese election watch coalition (COSCE) and resulted in 74,000 individual observations of activities at polling stations.

2 METHODOLOGY

For this study, we collected data from three main sources: Twitter, Facebook, and YouTube. These sites were used because they are known to be popular in Nigeria, and these platforms allow researchers to collect and analyse data from them in a relatively easy and structured manner.

It is possible to manually collect social media data in a number of ways - copying, screen grabbing, note-taking, and saving web-pages. However, where large volumes of data are involved, the most appropriate method is to collect the data automatically through connection to a platform's 'Application Programming Interface' (API). The API is a portal that acts as a technical gatekeeper of the data held by the platform. It allows an external computer system to communicate with and acquire information from the social media platform. Each API differs in the rules it sets for this access: the type of data it allows researchers to access, the format it produces these data in, and the quantities of data produced. Some APIs can deliver historical data stretching back months or years, while others only deliver very recent content. Some deliver a random selection of social media data taken from the platform, while others deliver data that matches the queries – usually keywords – stipulated by the researcher. In general, all APIs produce data in a consistent, 'structured' format, and in large quantities. Facebook's and Twitter's APIs also produce 'meta-data' – information about the data itself, including information about the user, their followers, and profile. This meta-data can be a rich source of information valuable to social media researchers, often containing information on everything from the sender's device type, to their account creation date, and location.[25]

Data collection and classification

Twitter

Over the period 18 March – 22 April 2015, researchers collected 13.6 million tweets posted by 1.38 million unique users. Data were collected by a handcrafted selection of words associated with the Nigerian Presidential and State elections held in March – April 2015 (see annex for a full list). These data were collected in six phases, to reflect the changing dynamics of events. Only English language tweets were intentionally collected, using English language search terms (although other languages and tweets combining English language and other languages, for example Pidgin English, may have also been included in the tweet).

Nigerians generally engage in social media in English. However, conversations are often flavoured with a mix of local languages, especially Pidgin English, Hausa, Yoruba and Igbo which are the most widely spoken languages. A few social media accounts are dedicated to communicating in the local languages; however, the majority of political conversations are in English. This is likely due to the fact that

Nigerian languages are more spoken than written, and because the exigency of communicating to a broad national audience requires the use of a common language.[26]

Once collected, the data were broken into smaller categories of meaning. This was performed by algorithms called 'classifiers'. To build a classifier, the researcher first decided on the categories into which the data were to be classified. (How to choose the categories into which data are to be classified is an important consideration. In our experience, categories are best chosen based on a preliminary review of the data, and not driven by preconceived notions of what categories are hoped for). In this research, the categories were chosen by analysts following a preliminary, manual review of the data collected. We selected categories which we thought would be most amenable to testing various modes of analysis in order to answer the research questions, but were also reflective of categories found in the data. For example, in the original proposal we planned to divide this into data that was about 'communication' and data that was about 'monitoring'. However, the data we found was not amenable to this division.

The researcher then 'trained' a classifier by manually categorising an initial sample of tweets, and inputted them into the Method52 software. The software thereby learnt which units of language and patterns of language use are associated with the different sorts of tweet. As the analyst trains the classifier, the software reports back on how accurate the classifier becomes – that is, how its own decisions compare to the decisions of the analyst on a privileged 'gold standard' set of tweets. The use of classifiers in this way is a typical approach to dealing with very large sets of 'natural' language. Hence it is often called 'natural language processing' (NLP).

To break the data down into categories suitable for further analysis, classifiers were fed into one another to create a tree like structure shown in table 1, below. For example, identifying those tweets complaining about issues with the card readers required two more classifiers: one to split the 5.5 million reportage tweets up and identify those discussing problems voting, before a final algorithm identifying the 46,000 referring to card readers from within that data set.

To answer the questions relevant to this study, classifiers were built to categorise 'reportage' tweets into tweets about results (1.5 million), those about problems voting (386,000), and those about violence (408,000). How well these classifiers performed at this task is discussed in the Annex, below. This allowed us to create several more manageable data sets, which we could then subject to a variety of analyses.[27]

Facebook

Data collection on Facebook is slightly different to that on Twitter. There are several types of API access to Facebook data, most of which have been designed for app makers, such as a Public Feed API, a Keyword Insights API, a Marketing API and an Atlas API.[28]

For this work we used the Public Feed API which allows researchers to access all data that has been posted on selected public Facebook pages. (Public pages are usually run by an administrator that decides who can upload posts, but usually any users who have 'liked' the page will be able to share, comment or like the posts uploaded.) Access to all Facebook data is predicated on the user's settings and who has agreed to share information with them. Facebook's privacy structures are complex – potentially, any single user can have a distinct privacy setting for every piece of data they share. The Public Feed API will only return data that is public.

We identified nine public pages associated with candidates. These were usually 'fan pages'. This returned 1,137 posts made on those pages. The Public Feed API also allows researchers to collect all interactions with each post. This provided us with 539,000 likes, 218,000 comments and 108,000 shares. Each interaction is associated with one of the posts. We also identified 15 public pages associated with formal news outlets, and collected 28,000 posts made on those pages. This included 3.8 million likes, 1.6 million comments, and 1.2 million shares. Finally, we identified five Facebook pages associated with non-formal news outlets, such as notable bloggers. We collected 5,621 posts, 209,000 likes, 158,000 comments and 87,000 shares.

Facebook's Public Feed API collects data on a volume, not a time, limit. Therefore the data collection period was varied for each page. However, we capped the time collection to only include dates broadly similar to those used for the Twitter data collection.

Other

Twitter and Facebook are often used by users to share links to third platforms or sites. From a random sample of 388,000 tweets, we collected the YouTube links being circulated on Twitter. We collected 695 in total, representing 23 million views and 145,000 likes. However, there was very little analysis we were able to conduct on this data.

Analysis type

Once the data were collected, we conducted four waves of analysis. First, we undertook a 'use-type analysis', which examines the volume and nature of data available. We built a number of classifiers to sort the data into these different use types. We then classified the data into these different types of uses and subjected them to a series of more manual types of analysis to better understand what type of data was being generated that might be relevant.

Second, we conducted a demographic analysis, in which we analysed any relevant demographic information about users. We used Method52 and the data analytics platform Qlik to determine, as far as the data allowed, the location of users, age of users, gender of users, and some other variables. This was to better understand who was posting content.

Third, we conducted a series of network analyses which looked at the way these users connected to and communicated with each other. We used Method52, R, Python and the data analytics platforms Qlik and Gephi to explore the role of relationships in the communication of information about the election to answer, as far as the data allowed, a number of questions. We produced a number of different network maps in order to better understand the nature of these networks on social media, including the likely reach of users.

Finally, we compared our online data against offline data provided by CASE 2015 to examine the extent to which online data was a useful gauge of offline events. In order to compare online and offline data, we accessed 796 CASE reports collected during the election and inputted them into Qlik. We then added around 350,000 tweets related to violence during the election (we judged this would be the most comparable data set).

Caveat

This research has been conducted using proprietary software, developed by the research team in partnership with the University of Sussex. Although the techniques used would be the same – natural language processing and network analysis – other research groups would likely have their own methods and may have reached different conclusions.

3 RESULTS

Use type analysis (type of data available)

Proportion and volume of data relate to the election

On Twitter, we collected 13.6 million tweets over the data collection period, of which 12.4 million were relevant to the election. By comparison, the search terms we selected produced just 260,000 tweets over a more 'average' week collected 19-26 June 2015. (This comparison between election and non-election tweets is based only on the generic sampling terms we used, which were place names, such as 'Nigeria', 'Abuja' and 'Lagos', and not election specific sampling terms such as 'NigeriaVotes'. This makes for a more accurate comparison). Including the enormous spike of tweets on both voting day and election day, we estimate Twitter to have been almost ten times more active during the period of collection than subsequently.

The collection terms we used were mostly tailored to the election - candidate usernames and election-specific hashtags formed the vast majority of collection terms. We also collected data using broader terms, which included words like 'Nigeria' and 'Lagos'. Therefore we built a relevancy classifier (accuracy scores for every classifier are included in full in the Annex) which removed 1.2 million tweets that were not related to the election, such as discussion of Naija music and Nollywood.

We built classifiers to split the data up into more manageable categories. We found the single largest distinction between data was between tweets which were 'reportage' (i.e. people describing events) and 'comment' (i.e. people commenting on events). There were 5.5 million tweets in the reportage data set and 4.8 million tweets in the comments data set.

On Facebook, data were not subdivided into specific categories because of the range and diversity of data. We were able to find the content that was interacted with the most by users of the site. ('Interactions' refer to content that a user has either 'liked', commented on, or shared. It is a useful proxy for reach because each time a user interacts with content it is, potentially, shown on their friends' timeline.)

We identified nine pages which were public fan pages of the main candidates and parties. In total we collected 1,137 posts. These posts received a total of 539,727 likes, 218,906 comments and 108,716 shares. These posts were most frequently photos followed by status updates. Unsurprisingly, the production of posts increased as the election approached and dipped again once the result was known.
We also identified 15 news pages which were Facebook pages of popular official news websites. These produced 28,767 posts and received a total of 3,818,580

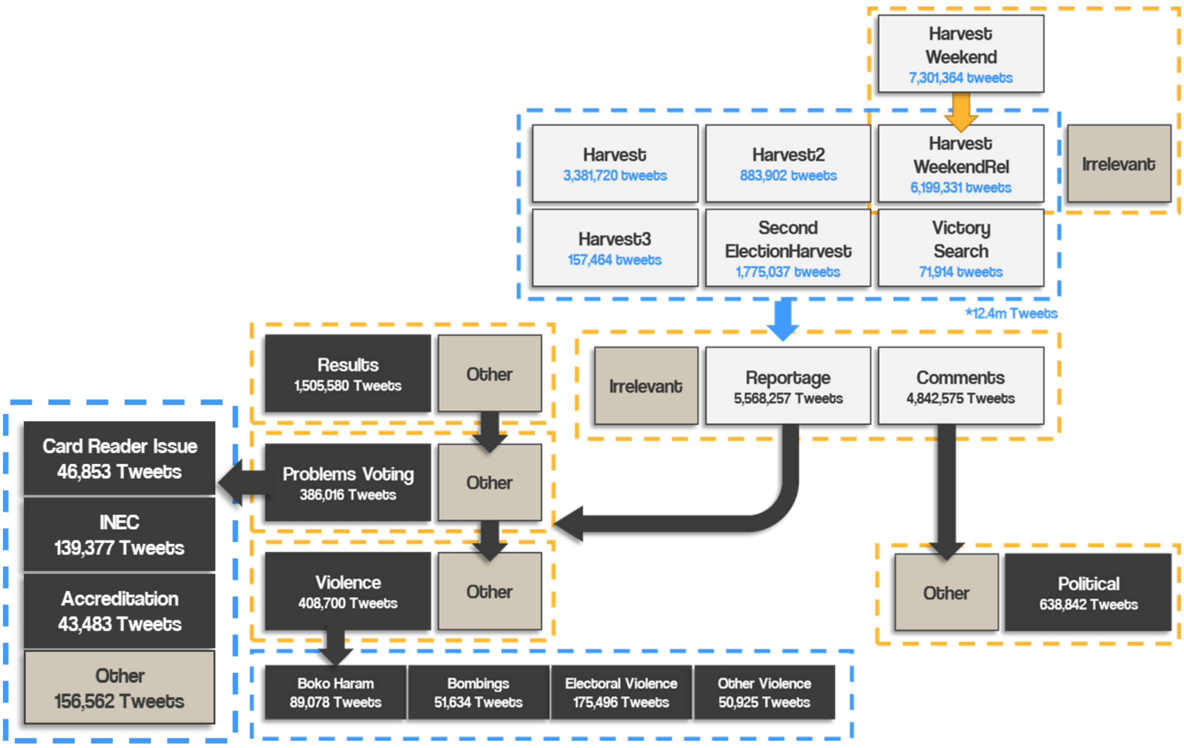
likes, 1,593,938 comments and 1,224,187 shares. These posts were predominantly links as they connect users to content on the news organisation website.[29]

We identified five pages which were Facebook pages of popular bloggers and non-formal news outlets from which we collected a total of 5,621 posts. In total these posts received 209,772 likes, 158,179 comments and 87,061 shares. Post production of bloggers was more sporadic than previous categories, reflecting their purpose of commenting rather than reporting on events. Similar to the posts of formal news organisations, the majority of posts are links, as Facebook is a location from which to drive traffic to their site.

Proportion and volume of subjects being discussed

Based on a review of the data, we built further classifiers to better understand the type of Twitter data available. Reportage was then further divided into tweets about results (1.5 million), those about problems with voting (386,000), and those about violence (408,000). Based on our classifications, we found the following categories across the data sets. This provides a top level view of the sorts of common themes and subjects discussed.

Figure 1 - Types of data available



Public confidence in the electoral process and in the final results

To answer this question, we categorised two types of data from Twitter, drawn from our total data set.

First, we identified 408,383 tweets over the period that were about violence. We built a new classifier to discriminate between different types of tweet within that data set (as per figure 1, above). This allowed us to split the data between tweets that were about Boko Haram (approximately 89,000, from 48,000 unique users); bombing (approximately 51,000; from 22,000 unique users); electoral violence (approximately 175,000; from 53,000 unique users); other violence (approximately 50,000; from 19,000 unique users); and other (i.e. not about violence but inaccurately classified in the preceding step: approximately 34,000).

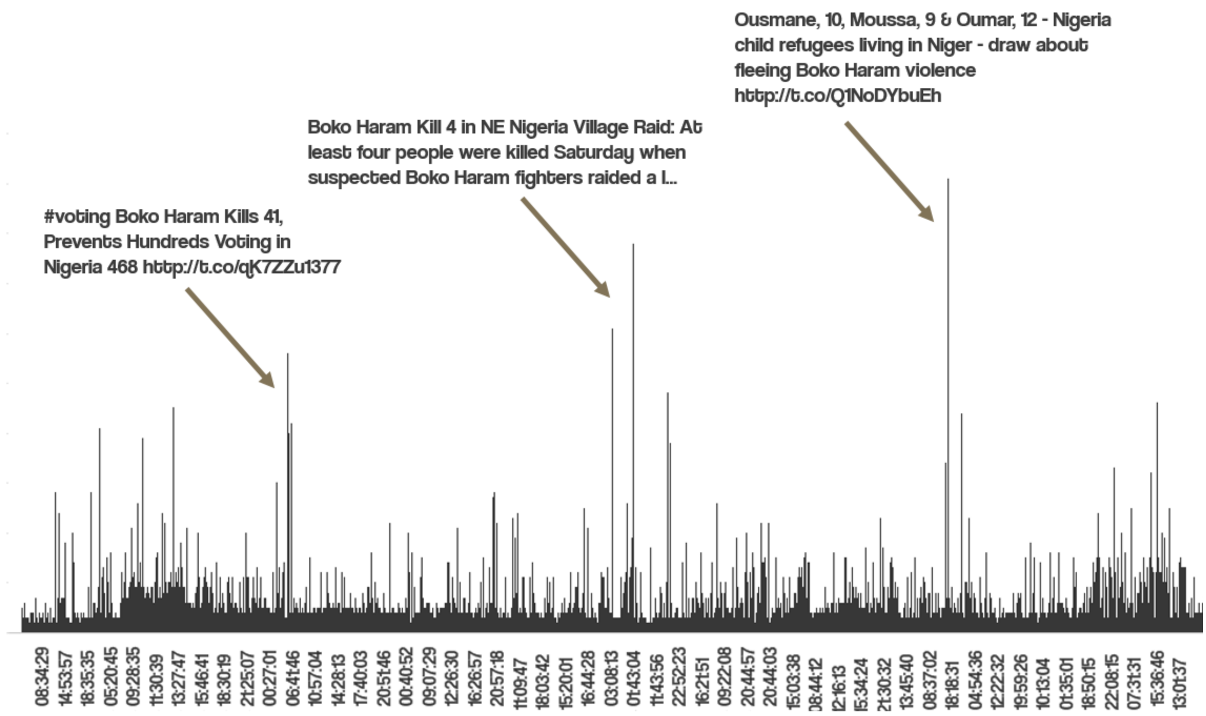
A more granular analysis of tweets over time highlights the individual pieces of content that were most widely circulated. It is a simple task to identify the most popular retweets from within each cluster of the violence data set. These are set out below. The international interest in Boko Haram is clear, and shows how valuable it is to have a classifier that can filter these tweets out (leaving the more useful data about election specific issues). The single most widely shared piece of information was the bombing that targeted the polling unit in Enugu. It is not, however, possible for us to ascertain how many people actually viewed these tweets.

Figure 2 - Top tweets in the violence data set

		Retweets
Boko Haram	One year later: 219 #Nigerian Girls are missing Kidnapped by #BokoHaram http://t.co/1i3qjrMCrT #BringBackOurGirls http://t.co/YSmrjPrOzh	672
Bombing	Breaking News: Bomb Blast at Polling Unit in Enugu State	623
Other Violence	FLASH: Deadly infectious disease hits Irele township in Ondo state. Victims wake up blind and dead a few hours later, at least 25 dead.	433
Electoral Violence	BREAKING: APC Agents are being barred from accessing collation centre in Apapa #SitRoom15 #NigeriaDecides	606

However, it is debatable whether this type of content is of any real interest to electoral observers, since this is typically widely shared news reports. This underlines the need to look beyond simple 'volume' counts and find the valuable content not being picked up and circulated as frequently. Figure 3 below sets out the volume over time of tweets about violence. The larger spikes are the widely shared content. However, the smaller spikes might be valuable, local insight (we discuss this below).

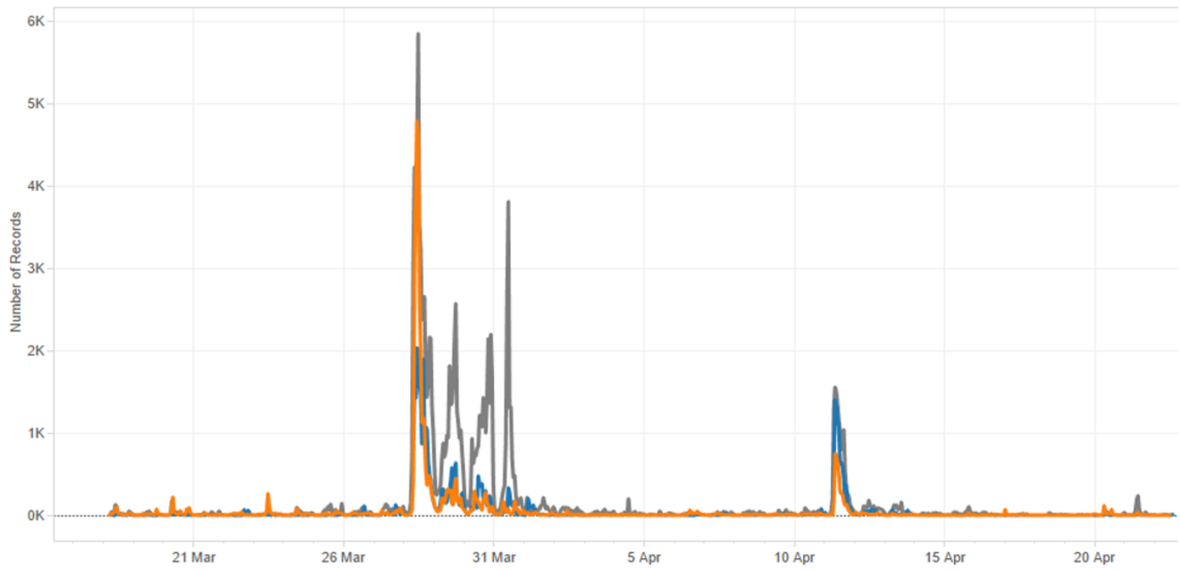
Figure 3 - Violence tweets plotted over time



Second, we categorised 386,000 tweets about problems with voting (henceforth 'problems voting'). We built another classifier to distinguish between four categories that appeared frequently in the data: card reader issues (approximately 48,000 tweets; from 20,000 unique users); INEC (approximately 139,000; from 61,000 unique users); accreditation (approximately 43,000; from 15,000 unique users); and other problems (approximately 156,000; from 66,000 users).

By plotting these data sets as a time series, we can see that accreditation and card reader issues dominate on election day, while INEC is the subject of debate up until the results are announced. These major spikes are represented by the large spikes in figure 4, below; while smaller spikes represent potentially more valuable data, but are harder to identify.

Figure 4 - Problems voting data



Tweets in the problems voting set contained a great deal of useful 'meta-data' (data about the data) that allow for greater insight into the data. This includes: images / photos of real-time activity 'on the ground' that are useful in determining what is taking place (13 per cent of tweets in the entire dataset contained a photo); a time-stamp which allows researchers to determine the exact moment a tweet was sent; and geo-location, which is not common, but can be supplemented with algorithmically-determined location, and would allow DfID to place quite a high proportion of these reports geographically.

Rumours and hateful language

We also examined the extent to which rumours and hateful language circulated on Twitter.

Reports of violence of many types circulated frequently on Twitter. @INECNigeria – the independent national electoral commission – looked to dispel false rumours it received on Twitter, usually tweeting on the hashtag #FalseReports. To understand the lifecycle of a false rumour on Twitter we received a set of false rumours from DfID. (These were tweets dispelling rumours that had been posted by the @INECNigeria account and covered topics from simultaneous accreditation of voters to hijacking of voting material.)

From these instances, identifying the original rumour was difficult; the @INECNigeria account did not 'reply' or 'mention' the user which originally sent the false information (a practice we would suggest adopting). It is difficult to estimate the numbers of false rumours circulating on the platform. With nearly 800,000 reports of either violence or problems voting, we are reliant on

@INECNigeria or DfID's CASE data to confirm whether those reports were accurate or not.

However, by searching for words contained in the tweet, we were able to identify a number of tweets which could have been picked up. For example, one report of simultaneous accreditation was linked to Jos North in Plateau. By looking for reports referencing Jos North, we were able to identify sixteen unique tweets in the 'problems voting' and 'violence' datasets (figure 5, below).

Figure 5 - Tweets about Jos North and accreditation

"@iamSeyiPeters: Underage voting in Jos North LGA, Plateau state? UPD #Nigeriadecides #Nigeria2015 http://t.co/7jQ88a1ZF0"@INECAlert	@inecnigeria Plateau State REC is under pressure to rigg Jos North and Wase LGCs in favour of PDP. Rigging in Jos South LGA #NigeriaDecides
"@SayNoCampaign: Card Reader failures reports at PU 53, Tudun Wada LGA, Jos North LGA, Plateau State #Nigeriadecides"	@yiaga @amplifiedr @osiwa1 at PU 31/05/09/028 Congo Junction Jos North accreditation on course but with delay in machine capturing!
Card Reader failure in PU53, Tudun Wada LGA, Jos North LGA, Plateau State cc @inecnigeria #nigeriadecides #nigeria2015	12:54 RT @MykelSogoye: In many polling units in Jos North, no materials as at 12:30. No INEC official as well. #Nigeriadecides"
"@NigeriaNewsdesk: Some youth in Jos North LGA of Plateau State have set ablaze the home of a PDP chieftain, Adams Alkali."	A boy has been shot by a @PoliceNG officer from Sector 1B at PU38 Ali Kazaure Ward, Jos North, Plateau @inecnigeria #Nigeriadecides
"@pmnewsnigeria: In many polling units in Jos North, no materials as at 12:30. No INEC official as well. #Nigeriadecides #SitRoom15"	Ali kazaure, Jos North. @Tha_Wierdt: Which unit? "@Naseeem_666: Ballot box snatcher shot dead in Jos, by security forces. #Nigeriadecides .
#PollStream #Nigeria2015 @Channelstv INEC supervisor missed up distribution process in Tundun wada/Kabong Area of Jos north, plateau state	Jos North LG: elections have been completed, however there are no inec officials at d collation center to collate d results #Nigeria2015
@APCNigeria #Nigeriadecides PU 31-05-10-055 in Jos north Naraguta B has abt 730 accredited voters but only 200 ballot papers provided"	Voting kicked off @12 & accreditation still on goin inNaraguta B, Jos North LGA #NigeriaDecides #Nigeria2015 http://t.co/lE1dTFR12e
@inecNigeria officials are yet to arrive in a PU here in Jos North, Masallacin Jumma'a, 'Yan Dankali Cc @zarahmusa @APCNigeria @omojuwa.	"@NigeriaNewsdesk: Angry youths attack PDP member in Jos North http://t.co/70DtAtNhL9â€¦ via @todayngr"see big lies waiyooo munshiga UKU a JOS

There was a typically a delay between a rumour being circulated and @INECNigeria publicising its falsehood of between two and four hours (based on rumours we could trace back to a clear source). As a consequence, only very few rumours were actually dispelled, likely due to capacity. In Jos North region alone there were 51 unique reports of violence (a number that swelled to 371 after retweets and replies) and electoral malpractice, and @INECNigeria only responded to one.

We also identified individual users whose insight could have been useful in verifying or corroborating reports, as well as false rumours. One Twitter user was reporting from her local polling station with regular updates on violence, electoral malpractice and policing, including photographic data. Verification of this type of content is potentially a fruitful source of information.[30]

We also identified a small number of racial or ethnic slurs and searched for them within our entire data set. However, we found very few instances of these terms being employed. Therefore we instead searched for examples of words for ethnicities being used ('Igbo', 'Hausa', 'Fulani', 'Yoruba' and 'Efik'). Igbo was by far the most mentioned (nearly 11,000 instances).

The data showed that there was a general spike in words denoting ethnicity around the presidential poll itself. Only a spike in words which are considered politically charged or politically relevant by Twitter users would be expected at election time, so this suggests some prominence of ethnicity in discourse surrounding the election. Given the historic importance of ethnicity when it comes to region and candidate support in Nigeria, this was unsurprising. There was a further spike in mentions of 'Igbo' in early April, following the election. This was largely due to a controversy surrounding the Oba of Lagos, who was reported to have threatened Igbo residents into voting for the All Progressives Party candidate, Akinwunmi Ambode. The backlash that followed on Twitter was noted and catalogued by Nigerian media as ObaGate.[31]

The data were not available on Facebook to answer the question.

Most commonly shared stories about the election over the period and how far was their reach / engagement

In order to better understand the most widely shared content, we created a random sample of the entire data set, and then found the most retweeted tweets and most shared tweets within that set. (We did this because the entire data set of 12.4 million tweets was too large to process for this purpose.) The sample was based on a random selection of three per cent of tweets (372,000) drawn in proportion to the original collections.

We then undertook a qualitative review of the top 25 tweets and links contained in the data set and categorised them into categories chosen by analysts, to provide a sense of what sort of content was popular. The vast majority of the top tweets are reportage of the election results, with a scattering of campaigning, electoral reportage and business promotion. The links shared are more diverse – there is a plurality of news articles, but the sample also contains humorous vines, documentary videos, blogs and links to the websites of specific electoral organisations. This illustrates the fact that much of the content is popular, viral content, while the potentially more useful 'on the ground' information is likely to be buried beneath the major news stories.

Table 1 - Most popular tweets

	Tweet Text	Number of Retweets	Type
1	Muhammadu Buhari wins Nigeria's presidential election as Goodluck Jonathan admits defeat http://t.co/LPbSLD9grK #NigeriaDecides	247	Results reportage (news link)
2	"General Muhammadu Buhari @ThisIsBuhari of @APCNigeria is hereby returned the winner"- Jega http://t.co/WK5YKlg8Xp	210	Results reportage (INEC retweet)
3	Just received a phone call from Gov Mukhtar Ramalan Yero congratulating me for winning the Kaduna Gubernatorial election. I thanked him too.	200	Results Reportage (politican)
4	All eyes on Jega #Nigeriadecides #Jega #Nigeria #Lagos #np #naija #KCA #RT #lt #lrt #Abuja #Nigeria2015 #INEC http://t.co/DGCDdT7GBA	196	Election reportage (musician retweet)
5	Professor Joseph Chikelue Obi : " The 2015 Nigerian Elections are Over. President Buhari must now quickly move into Aso Rock & Start Work ".	172	Results Reportage
6	Kudos to Mr. President for congratulating his winning opponent GMB on his victory. Nigeria is one. #APC	163	Results Reportage
7	I am in Canada, I'm giving results to my cousin that voted PDP in Nigeria. He does not have electricity to watch #NigeriaDecides	159	Self-reportage
8	WE WON #LAGOS!!! THANKS TO EVERY ONE WHO CAME OUT TO VOTE #APC #AMBODE WE REALLY APPRECIATE YOUR SUPPORT. NOW MY LEGACY CONTINUES...	152	Results Reportage (politican)
9	#APC is here for a new Nigeria. every single person RICH OR LESS PRIVILEGE matters alot to us. #VOTE #CHANGE AND WE WILL CONQUER BOKO HARAM	152	Campaigning
10	Nigeria http://t.co/Kcaz1jCpPL #Nigeria #HVAC #Lagos #Ac #usa #KCA #engineering #it #RT #Africa #Abuja #business http://t.co/mCHhIX7oa1	149	Business
11	By the wish and will of Nigerians, who do you see winning this presidential election? Retweet for Buhari Favorite for Jonathan CC @omojuwa	148	Campaigning
12	#BREAKING Opposition candidate General Buhari wins Nigeria's presidential election. #NigeriaDecides #Nigeria2015 http://t.co/UEMY8FFMLZ	141	Results reportage (news link)
13	Professor Joseph Chikelue Obi hereby Publicly Congratulates the Incoming Nigerian President (General Muhammadu Buhari GCFR) on His Big Win.	134	Results reportage
14	Nigeria Website http://t.co/c8nXZDvxYQ #Nigeria #webdesign #Websites #Lagos #Africa #WordPress #webiz #html5 #np #RT http://t.co/JtAgiwZWWo	133	Business
15	'Buhari' is the fastest growing brand in Africa. RT if you agree	132	Business
16	#BREAKING President Jonathan concedes defeat, congratulates General Buhari #Nigeriadecides	131	Results reportage
17	BORNO HAS 1m PVCs & YET TO BE ANNOUNCED, WE ARE LEADING WITH 2MILLION+ VOTES ALREADY. SO WHAT DOES THIS MEAN? #Nigeriadecides #Nigeria2015	125	Results reportage
18	#Ondo AA:386 AD:1237 ACPN:2406 ADC:1227 APA:1139 APC:299,889 CPP:1012 HOPE:184 KOWA:223 NCP:846 PDP:251,368 PPN:734 UDP:184 UPP:221	125	Results
19	#Katsina AA:183 AD:283 ACPN:402 ADC:498 APA:1671 APC:1,345,441 CPP:976 HOPE:47 KOWA:215 NCP:330 PDP:98,937 PPN:254 UDP:117 UPP:72	117	Results
20	#Enugu AA:433 AD:269 ACPN:479 ADC:478 APA:715 APC:14,157 CPP:237 HOPE:110 KOWA:203 NCP:761 PDP:553,003 PPN:407 UDP:1,623 UPP:290	117	Results
21	#Ekiti AA:94ACPN: 534APC:120,331CPP:330HDP:94 KOWA:108 NCP:377 PDP:176,466 PPN:388 UDP:60 UPP:145	113	Results
22	Nigeria Music http://t.co/obSgq3ot6m #KCA #Afrobeat #RT #np #lt #Afrobeats #Nigeria #Lagos #Music #Abuja #Naija #we http://t.co/VkF2Bojc3T	112	Business
23	Congratulations Nigerians. Thank you for lending your voices. Let's join hands in nation building. God bless Nigeria! http://t.co/HBQxKKR0nq	109	Results reportage
24	Buhari closes in on historic win as vote counting draws to a close #NigeriaDecides http://t.co/JzTpwDyMWc http://t.co/R6hpYv71uF	107	Results reportage
25	Congratulations, General Muhammadu Buhari as #NigeriaDecides http://t.co/yV1MCLKJ4G http://t.co/k43RPmdLy2	100	Results reportage
26	#NigeriaDecides: Nigerians are so creative!! #LetNigeriaDecide http://t.co/L1942Mar1R	100	Campaigning
27	#Kaduna AA:218 AD:273 ACPN:424 ADC:546 APA:1611 APC:1,127,760 CPP:824 HOPE:105 KOWA:176 NCP:754 PDP:484,085 PPN:549 UDP:79 UPP78	100	Results
28	I would have endorsed Prof. Jega for President because of the way he calmly handled that situation but we already have one. PRESIDENT BUHARI	99	Results reportage
29	This is my favorite photo! Women in Maiduguri defiant against Boko Haram and have come out to vote! #Nigeriadecides http://t.co/xSfhZEdxAE	98	Election reportage

Table 2 - Top Links shared

	URL	Number of Retweets	Description
1	http://atiku.org/aa/2015/03/31/congrat	122	Congratulatory message by Atiku Abubakar, GCON, former Vice President, on the election
2	http://bbc.in/19FtQlP	258	BBC: Nigeria election: Muhammadu Buhari wins presidency
3	http://www.trendinalia.com/twitter-trending-topics/nigeria/nigeria-	179	Trendinalia: list of trending hashtags in Nigeria
4	http://pollwatchng.com	149	Poll Watch Nigeria
5	http://aje.io/fkwf	111	Al Jazeera: Buhari closes in on historic win
6	http://cnn.it/1OPghkR	111	CNN: Nigeria votes: Forget the candidates, democracy was the real winner
7	http://bbc.in/1CIldCI	107	BBC: Africa live: as it happened
8	http://aje.io/9la6	104	Al Jazeeria: Buhari secures historic election victory in Nigeria
9	http://youtu.be/EdMez0yhNxg	102	Video: The true history of Arabs, Islam and Jihad
10	http://goo.gl/Mg2S2o	97	#NigeriaDecides: Nigeria 2015 Presidential, National Assembly Elections – LIVE
11	http://dlvr.it/8mBMNH	80	Entertainment express: MOPOL Shoots Police Boss in Bauchi
12	http://ec-media.sndcdn.com/u811uDbusZa	80	?
13	https://vine.co/v/OLUT9qQ5IMW	75	"Funny" vine
14	http://bit.ly/NigeriaPortal	74	The Nigeria Wall website (news)
15	http://omojuwa.com/2015/03/jonathan-sambo-david-mark-meet-with-	74	Omojuwa.com: Jonathan, Sambo, David Mark Meet With Service Chiefs and NSA, New Plans
16	http://youtu.be/mXOSQj4xjPY	73	Video: The true history of Arabs, Islam and Jihad
17	http://www.elections.premiumtimesng.com/#!/	72	Times Election Centre: Abia supplementary election results
18	http://www.hasyourlifechanged.com	72	Campaigning App/Game
19	http://bbc.in/1xwvQZd	71	BBC: Africa live: as it happened
20	http://www.shammahmarket.com/2015/04/cedaj-nigeria-enterprise-	70	Shammah Market: CEDAJ Nigeria Enterprise; A Leading Water Storage Tanks Supply,
21	http://nyti.ms/19FANU1	69	NY Times: Beleagured, Nigerians seek to restore a general to power
22	http://thndr.it/1xfUz1p	67	APC website
23	http://bbc.in/1C7PupG	62	BBC: Africa live: as it happened
24	http://bit.ly/1OJ8O6Q	58	Sahaha Reporters: Nigerian Elections 2015
25	http://m.channelstv.com	56	Channels television livestream
26	http://youtube.com/watch?v=fq2ycv	54	?
27	https://vine.co/v/OU5Kq30q1x5	53	"Funny" vine
28	https://vine.co/v/OwKzrgw5lFi	51	"Funny" vine
29	https://vine.co/v/OxmnKvAgguH	51	"Funny" vine

In order to have a better sense of the sort of posts being shared, we took a sample of the 200 most popular Facebook posts (i.e. those with the most likes and shares from candidate pages, news sources and blogs (informal news channels) respectively). Similarly to above, we manually categorised posts to get a better sense of which issues and content from their representatives voters were most interested in.

For candidate pages, we found that a third of popular posts (29 per cent) were encouragements to vote and thanks to political supporters. 4.5 per cent were posts about development legacies of the Jonathan administration (including agriculture and energy gains) while 3.5 per cent promoted women's issues including Women's Day.

Over a quarter of popular posts from news sources (26.5 per cent) dealt with vote counting and election prediction stories, while 9.5 per cent focused on Nigeria's future political developments (including Buhari's potential cabinet and policy decisions). Eight per cent of posts sampled were stories celebrating the peacefulness of the transition, and acceptance of polling results by all sides. For blogs and informal news, the most liked and shared stories tended to be those that directly supported candidates (13.3 per cent) while ten per cent that directly criticised candidates. The sample also showed a much greater number of stories dealing with ethnicity and politics in Nigeria (12.6 per cent).

Demographic data about users

We used Method52 and the data analytics platform Qlik to determine, as far as the data allowed, the location of users, age of users, gender of users, and other variables, for example if they are a 'member of the public' or part of an organisation. Because Facebook and Twitter data sets are very different, we have divided this section in two parts.

Twitter data

Number of users

On Twitter, we identified 1.38 million unique users contributing to the data set. This is calculated by taking the total count of unique usernames in the overall dataset. Given Nigeria is a country of approximately 170 million people, this is a small subsection of the population, particularly as the data include users from the wider international community.

In terms of the most popular accounts mentioned (i.e. containing @name) in any tweets within the data set, seven of the ten were media outlets. This shows how influential mainstream media outlets are in driving online traffic in Nigeria.

Gender

Demographic information isn't automatically available through Twitter. We therefore constructed a classifier to categorise the names and descriptions of those users contributing to our overall dataset and tried to determine whether they were male, female, or if the account belonged to an institution. Over half (52 per cent) of the unique users contributing to our dataset were male, and around a third (33 per cent) were female. The remainder were from organisational accounts.[32]

Account type

We built a second classifier to analyse the influence of news and other organisations in the dataset. This found 15 per cent of tweets came from civil society and other organisations, and just ten per cent of tweets were sent by news sources and journalists. However, it is possible this does not reflect their true contribution, as the news organisations were retweeted heavily. When we look at the percentages by *retweeted* username, the percentage of news organisations is 53 per cent.

Location of tweets

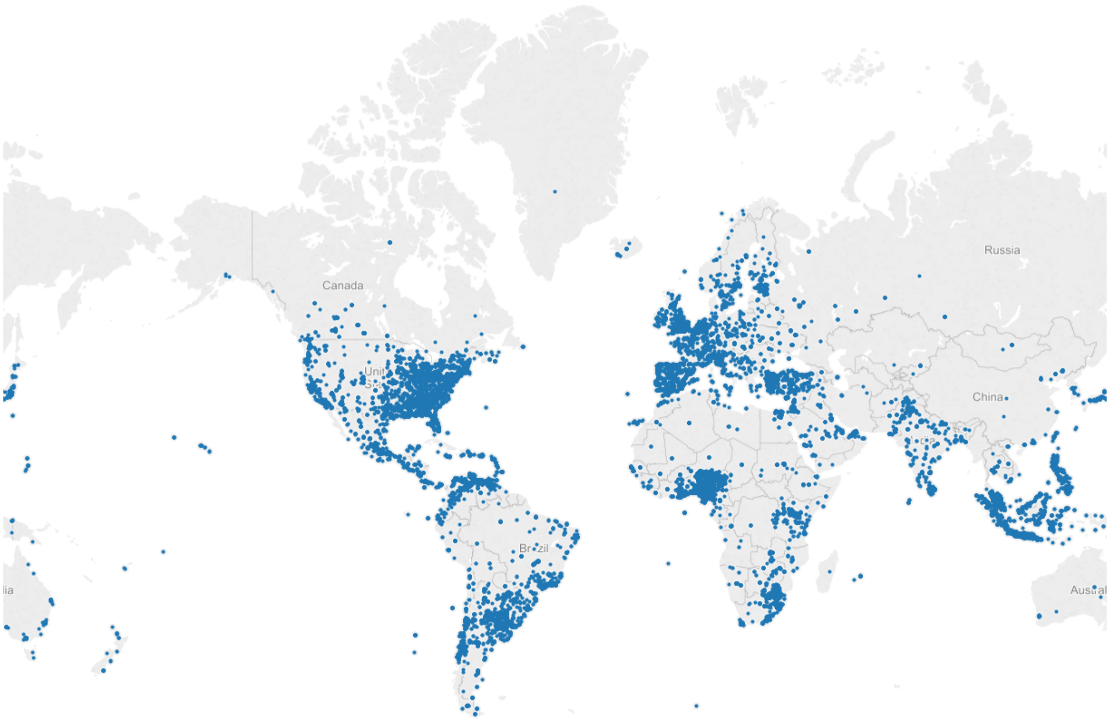
The dataset in total consists of 12.4 million tweets. Of these, 8.2 million (66 per cent) have location data (for example the name of a city) and 162,000 have precise geographic coordinates (1.3 per cent). The latter means users enabled the geo-tagging feature on their device, embedding in the tweet's 'meta-data' their precise longitude and latitude co-ordinates at the time the tweet was posted.

We analysed the 683 locations that occurred at least 1,000 times in the dataset. In total, this represents 4.1 million tweets, or half those tweets with location data. Of these 4.1 million tweets, 2.91 million were identifiable as being from Nigeria (71 per cent). The remaining 1.19 million (29 per cent) used descriptions that were either outside of Nigeria (London, NYC etc.) or were unverifiable (Worldwide, Under the Stairs etc.).

Of the 4.1 million tweets, 1.14 million came from users in Lagos (27 per cent), and 454,000 from Abuja (11 per cent).

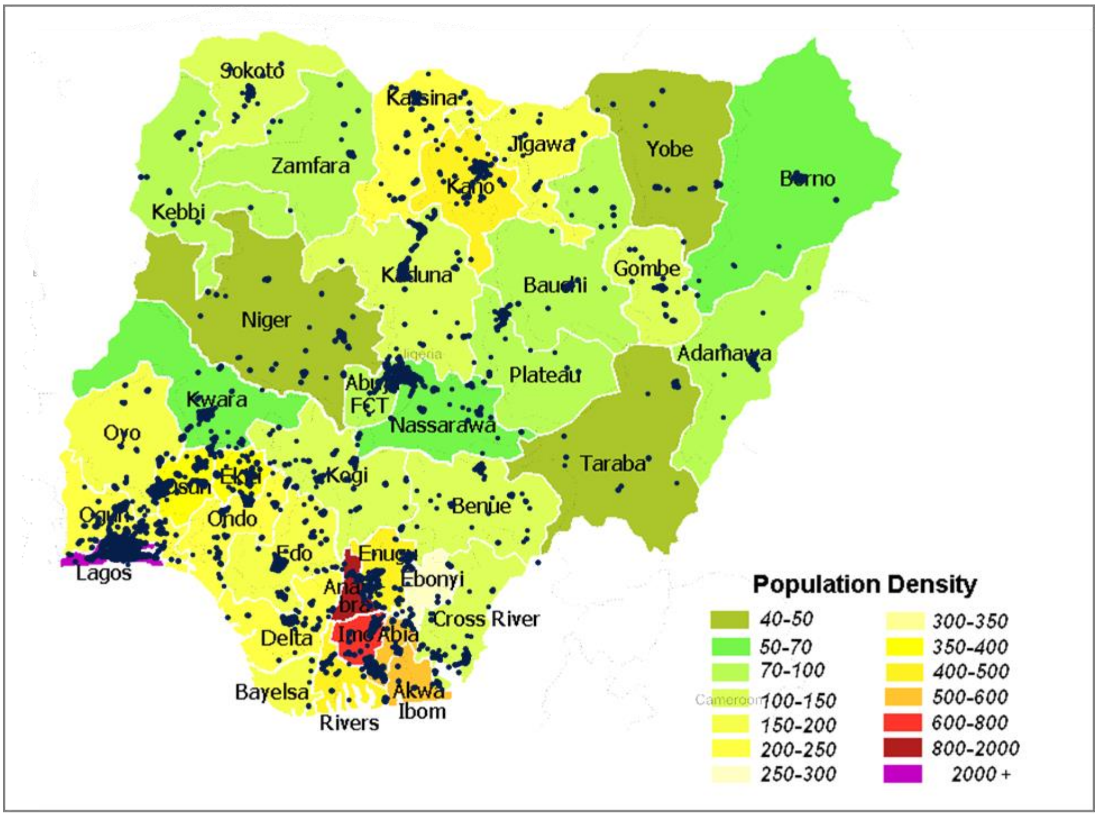
We downloaded every tweet in the 12 million dataset that contained a precise latitude and longitude. In total this produced 162,122 tweets sent by 50,281 unique users. The tweets were collected between 18 March – 23 April, peaking at the two elections as with the set as a whole. As is clear from figure 6 below, the Nigerian election was a subject of great interest around the world.

Figure 6 - Location of tweets around the world about the Nigerian election



We also focused only on those tweets geo-located in Nigeria, shown below. This accounts for 86,233 tweets (53 per cent of all geo-located tweets sent), sent by 11,466 unique users. This was then superimposed over a population density map. (Geo-located tweets are represented by the small blue dots.)

Figure 7 - Geo-located tweets in Nigeria



Broadly speaking Twitter activity does correlate with population, but there are interesting exceptions. The states surrounding Lagos in the south-west of the country tweeted a lot. There was also a concentration of Twitter activity in Nasarawa and Kaduna State. The tweets from Kaduna State are almost entirely from either Kaduna itself, Zaria, the second city, or the road connecting the two. Working out relative over- and under-representation by percentage is possible, but beyond the reach of this study.

Facebook

Despite Facebook having potentially more valuable demographic data about users, it is more difficult to extract that data. It is not possible to analyse the demographics of users based on their unique user ID via the public API (except manually).

In the Facebook data set, we identified 216,000 unique users from our sample of pages, who contributed 383,000 comments (around 1.7 per user on average) on these public pages. However, it is important to stress that this does not refer to all potential conversations on the network about the election – rather people who commented on the public pages for which we collected data.

Goodluck Jonathan's fan page account was by some distance the most active page; and had the most interacted with content.

Table 3 - Total users (Facebook)

	Users	Comments	Ratio
Bloggers	2,092	2,458	1.2
Candidates	113,331	173,250	1.5
News organisations	123,376	207,375	1.7

As some users comment on posts in more than one category, the total number of users is based on the number of individual users commenting on the top 100 posts, not the total individuals in each category. As shown, the majority of users posted just one comment on the pages.

Network analysis of users

Social network analysis is a distinctive research approach within the social and behavioural sciences. The approach focuses on the importance of the relationships between interacting actors or units of analysis.[33] This type of analysis allows researchers to visualise the users that are discussing a certain subject, or are part of a data set, and their relationship to each other – such as which accounts follow each other or share each other's content. It also creates a simple way to visualise the relative importance of users within the network, for example, whether their content is popular and widely shared among users.

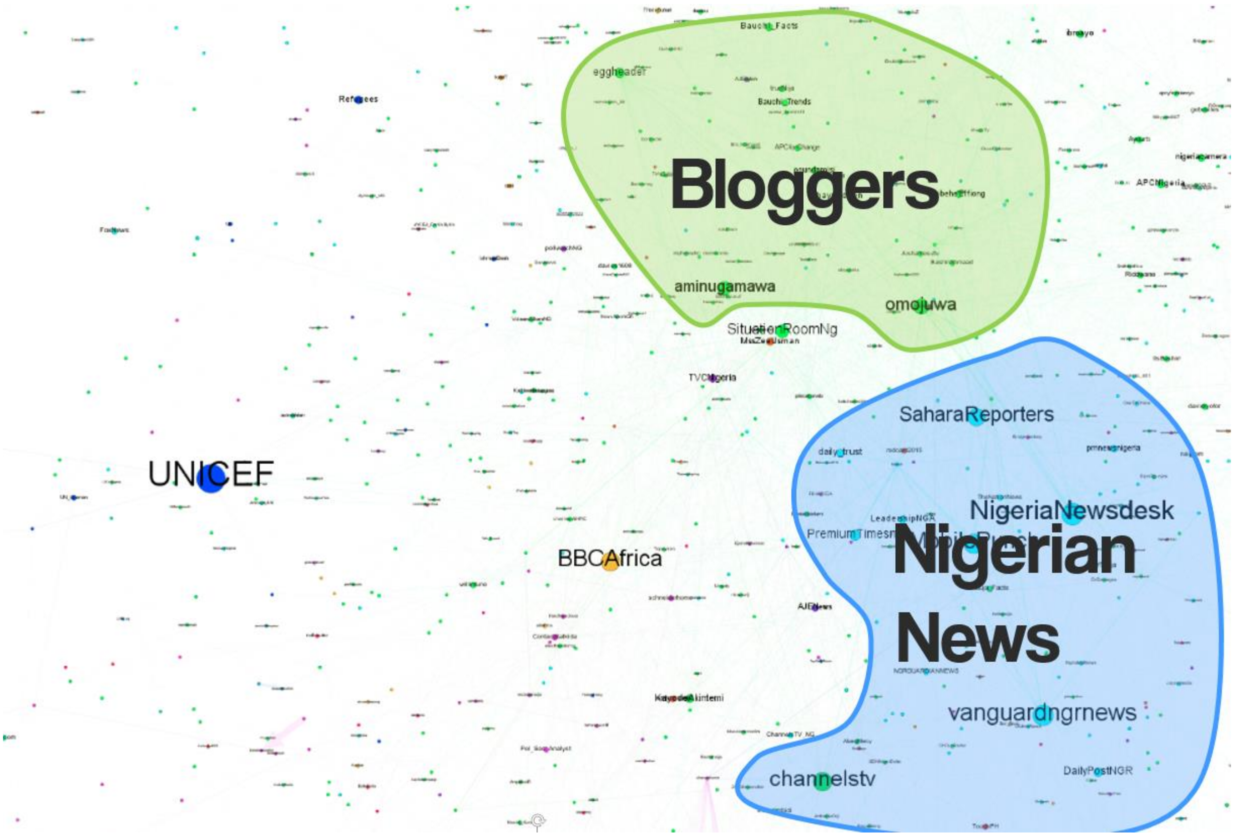
There are several ways to build network maps of users, since an analyst can determine what relationship between two users is being measured and presented. We used Method52, Python and the data analytics platform Qlik to explore the role of relationships in the communication of information about the election. We built a series of different networks to illustrate the different types of maps possible. This was done both on Twitter and Facebook data. We used the open source software Gephi to construct network maps. This is free software, but it is not able to handle the volume of our complete data set. Therefore, we built networks using a number of smaller data sets within our total sample.

Twitter

Violence data set

We built a network map of users based on whether a user's tweets had been retweeted by other users within the violence data set (approximately 409,000 tweets; from 109,000 unique users). The size of the node represents the number of times they were retweeted (the larger, the more often) and closeness to other accounts denotes who was retweeting them (the closer they are, the more they retweet each other).

Figure 8 - Network map: retweets in the violence data set



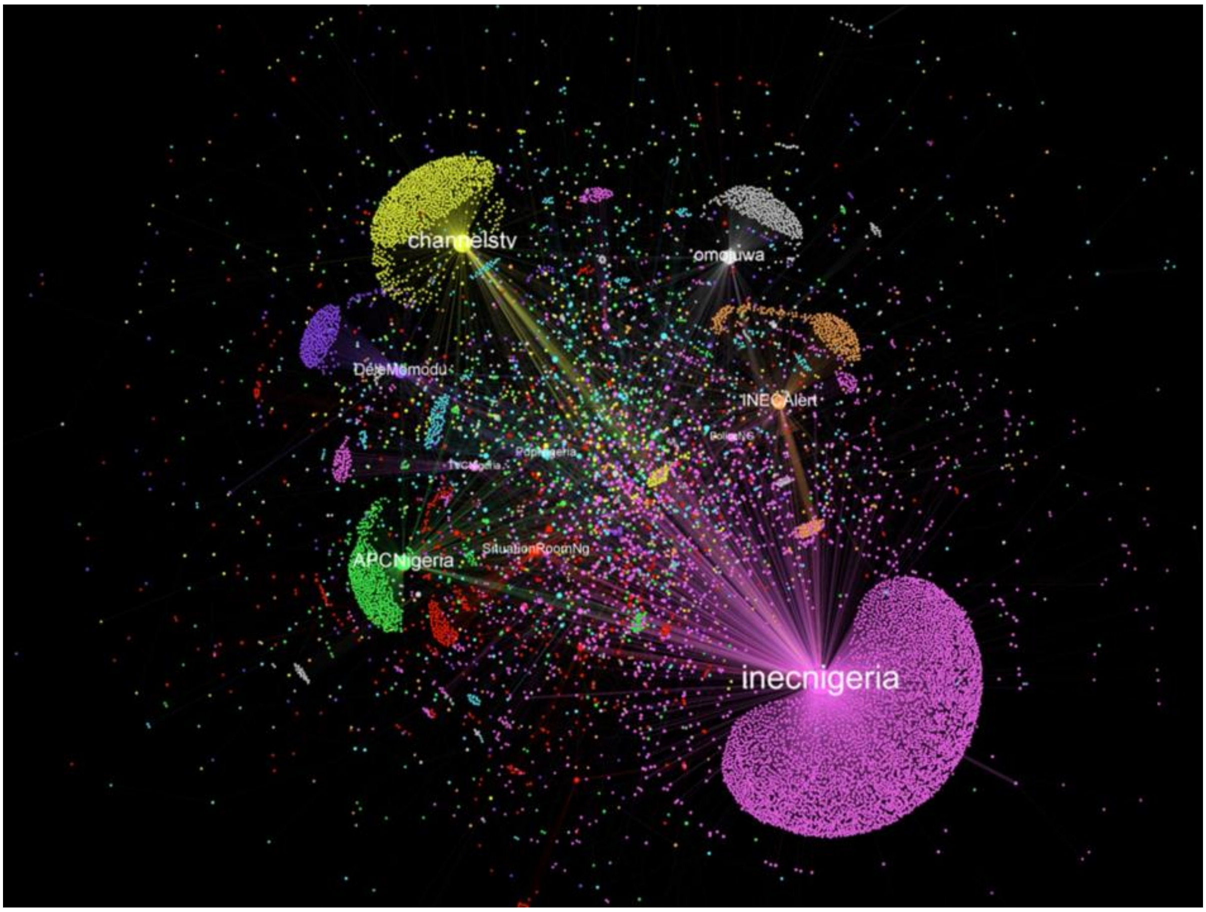
This map shows a number of things. First, there is was clear cluster of news outlet accounts (NigeriaNewsdesk, Vanguardngrnews) and bloggers (Omojuwa, Aminugamawa). (Boundary lines are added by us to illustrate the clusters.) Second, it shows how users' posts extended beyond their own network. Both UNICEF and BBC Africa accounts linked to both Nigerian and international users. Similarly, SituationRoomNG was a popular account among both Nigerian news outlets and bloggers. UNICEF was the largest account in terms of retweets – although that is a result of its very large number of followers based all around the world, and the organisation's interest in Boko Haram. While discussions of electoral violence were contained within the central cluster and referenced by INECnigeria, Nigerian bloggers and media outlets, discussions on Boko Haram tended to be limited to the international fringes. Non-Nigerian residents focused more on Boko Haram, while those taking part in the elections focused more on electoral violence and problems with the electoral process.

This is a relatively simple way to identify influential accounts within conversations online that are of interest. In this instance, it is possible to locate influential users within the network. At a higher degree of specificity, it is also possible to identify influential users within clusters. This image resolution is low, but it can be created to show every user in the network, and presented mathematically as a table.

Network of retweets for problems voting

Each network map looks slightly different according to the subject being discussed. Below, we conducted the same analysis, except applied to the problems voting data set (386,000 tweets, 84,000 unique users). In addition to it being a slightly different shape – for example with INECnigeria being very prominent – there were also accounts not found in the other networks. In particular, the APC Nigeria account was being widely retweeted, likely due to accusations it made of electoral misconduct. (We have presented a different visualisation, but the rules are the same: the larger the node the more retweets; and distance between accounts denotes how often they retweeted each other.)

Figure 9 - Network map problems voting data set

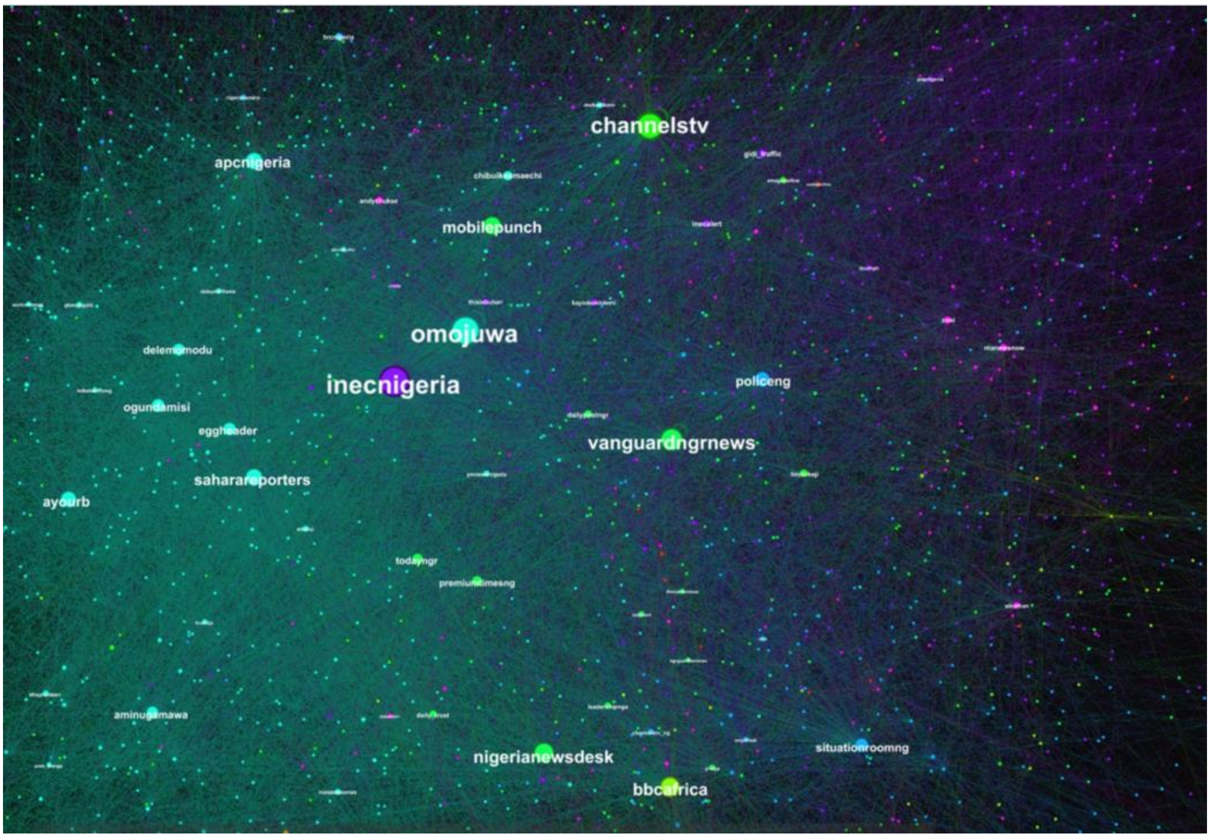


Most mentioned accounts

We built a second map using the same data set: this time, rather than measuring retweets, we measured user name mentions (when a user mentioned another user's Twitter username in their tweet). This is typically the way a user tries to send a public message to another user. As above, the size of the node represents the volume, and proximity to other accounts represents who is mentioning whom.

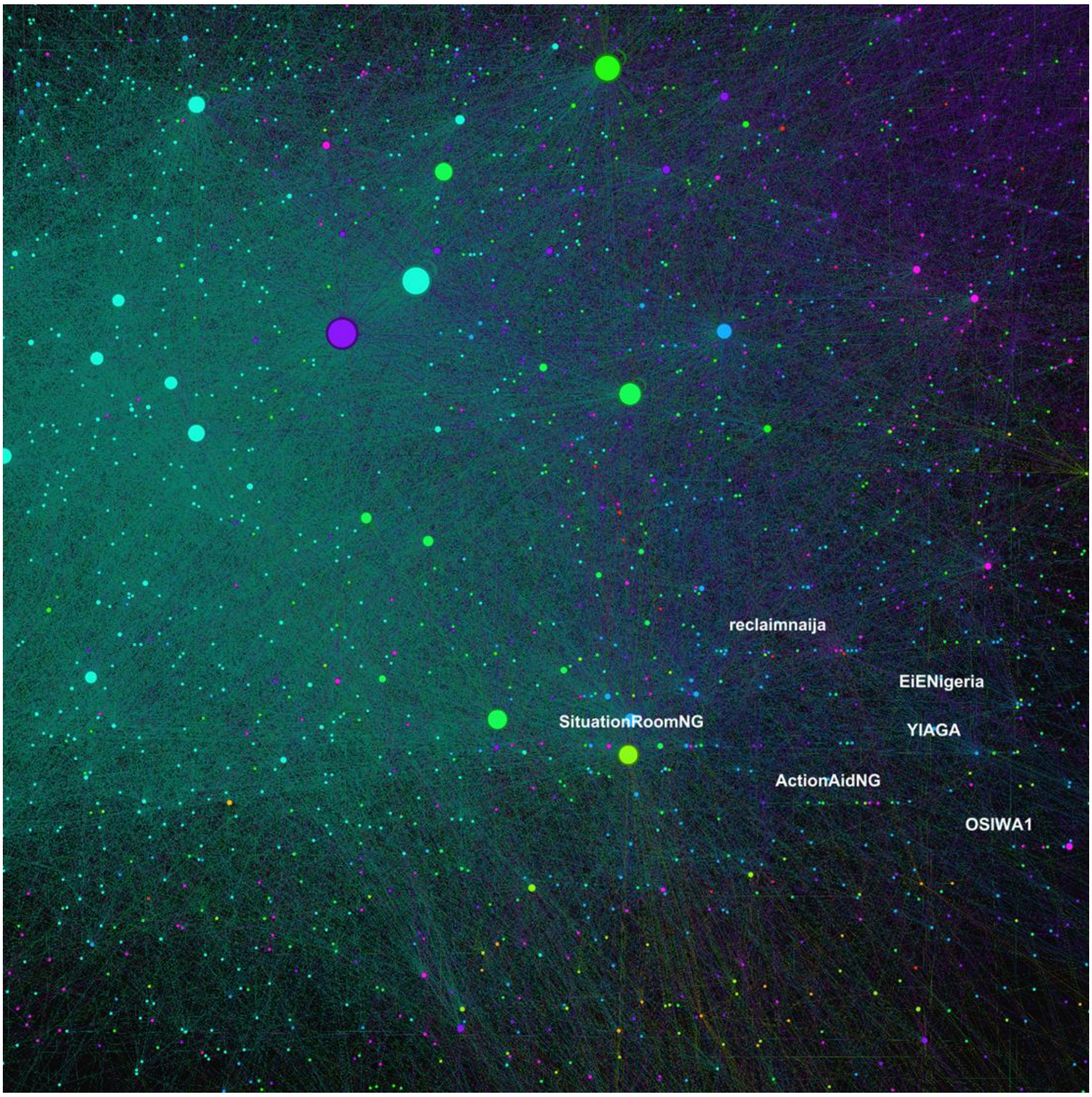
This shows some quite different accounts being influential. First, it shows that INECnigeria, while not retweeted often, is mentioned very frequently by other users and by a very diverse range of users, which is why it is in the centre of the map. It also reveals, as above, a number of accounts that might otherwise not have been noted as influential in these conversations.

Figure 10 - Map: network of mentions



As part of this analysis, we focused in on a small number of accounts that worked with DfID during the election. Figure 11 below is a close-up of the bottom-right-hand site of the network map of mentions, outlined in white.

Figure 11 - Map: key DfID accounts



This shows – in our judgement – that these accounts are fairly closely clustered together, suggesting they often mention each other, or are all mentioned by similar accounts. That, in turn, suggests they are not reaching as wide and varied an audience as they might if they were more actively trying to engage with users from other parts of this network.

Reach of tweets

It is extremely difficult to accurately determine the reach of any individual user, or the reach of any individual tweet that is posted.[34] It is possible, however, to quickly and easily find out how many retweets, favourites or replies an individual tweet has received (this is available in the meta-data attached to each tweet). However, this does not calculate the total possible audience, since that depends on how many followers other users have.

Figure 12 - Spread of tweets posted by INECNigeria

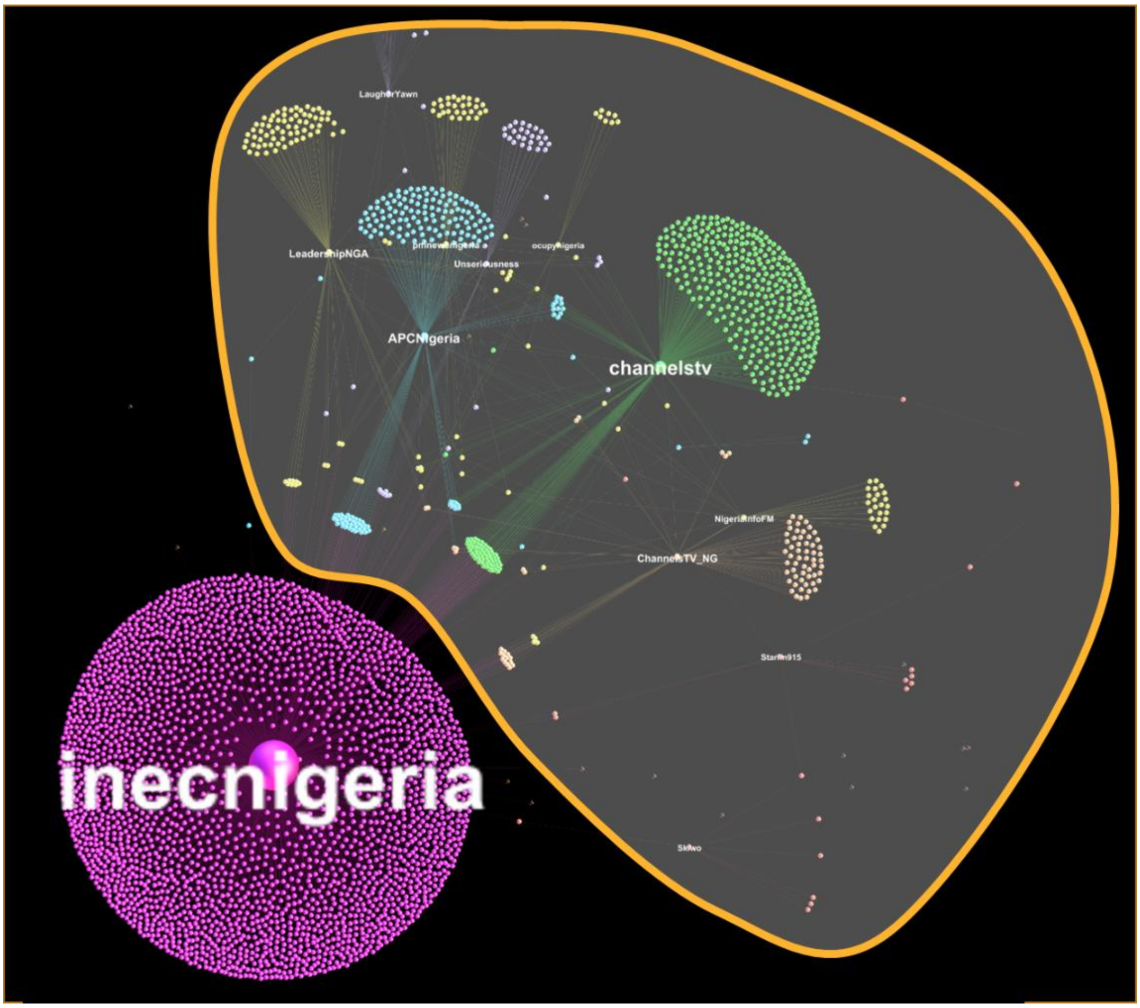


Figure 12 above shows nine tweets posted by INECNigeria's account. We can calculate a very approximate understanding of reach by combining the total followers of those who retweeted the tweet. By this method, these tweets could potentially reach around 3 million users (3,090,068). However, this doesn't account for secondary and tertiary retweets (i.e. retweets of retweets), as shown in the orange-bordered area above. If we follow the tweets beyond the original account and look at secondary and tertiary retweets, the 'reach' for the nine tweets then increases to an approximate 5.6 million. (There are 1.3 million users in the dataset. Reach here includes double counting of accounts, which gives an idea of how much duplication there is likely to be.) Further, this does not indicate how many of those might have actually read it, or acted on it.

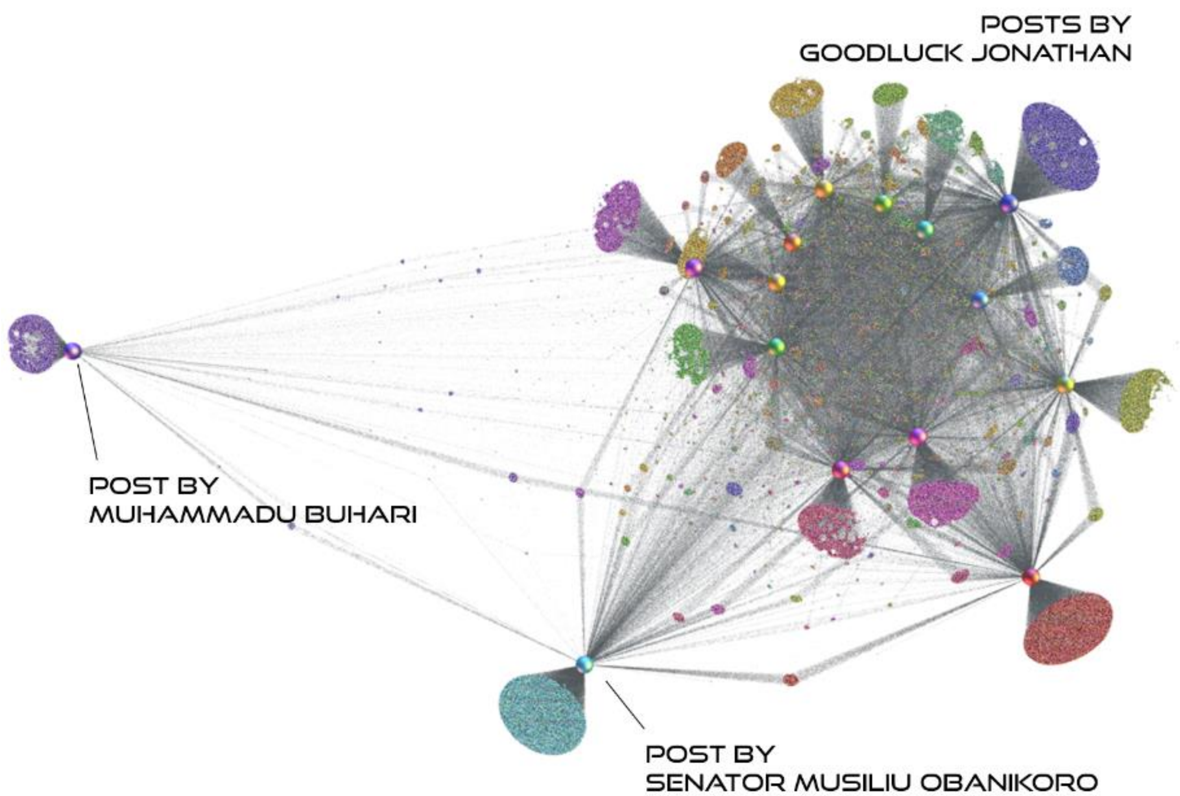
We also subjected the key DfID accounts to some analysis based on reach (SituationRoomNG, YIAGA, reclaimnaija, EiENigeria, OSIWA1, and ActionAidNG). SituationRoomNG (1640 tweets) was the most active by far over the period, followed by YIAGA (862). We also undertook a simple follower count.

By combining follower count and activity, we have some general sense which accounts are likely to reach wide audiences. SituationRoomNG has a large number of users and is very active, while EiENigeria has a very large following but is less active.

Facebook

Facebook results in slightly different types of networks being created, because the relationship between users is different (and so is the data available). On Twitter, networks can be built based on the relationship between two users; with Facebook it is based on interactions with the public page for which we have collected data.
To identify these communities of interest, we recorded the unique users who engaged with the most popular posts. In figure 13 below, large spheres represent the most 'liked' posts on candidate pages. The small spheres represent the individuals who like that specific post. This finds that most people only liked a single post. Where there are multiple posts from a single page (e.g. Goodluck Jonathan) these can be seen as a group of spheres – this is because users that like one post are more likely to like another post from the same page. We did this to examine the extent to which users tended to like a variety of posts from various candidates' pages. This suggests they do not.

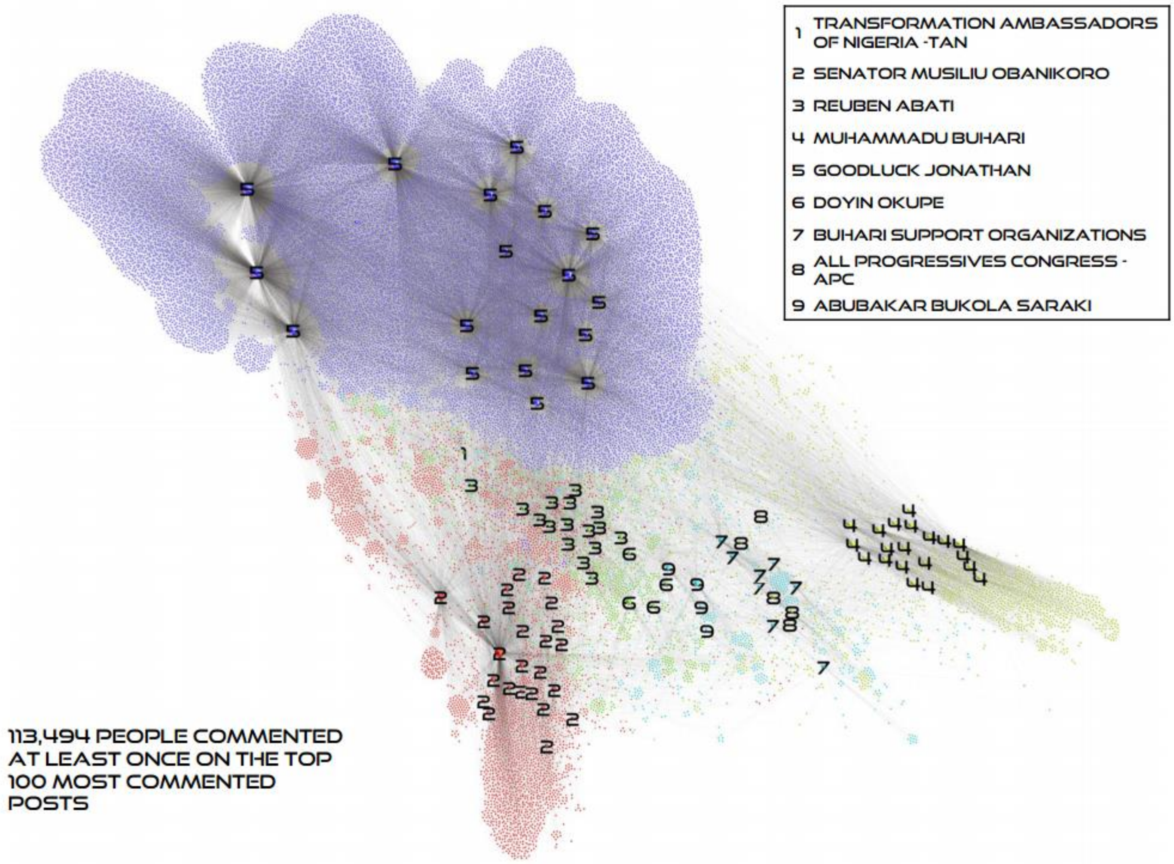
Figure 13 - Network of users that 'like' popular posts



Similarly to Twitter, this is a quick way of identifying top commenters on Facebook pages (although unlike Twitter, it gives no indication of how influential they might be). We did not seek to identify individual users, although this can be done manually.

We produced a separate map of people who 'commented' rather than 'liked' a page. As seen in the network of likes, the users commenting on the top 100 candidate posts tends to show 'loyalty' or engagement with a particular candidate's fan page. Around 113,000 users commented at least once on the top 100 most popular posts. Figure 14 shows the 2,475 users that commented more than five times on the top 100 posts.

Figure 14 - Network map of commenters



Page loyalty is particularly evident amongst the users who engage most frequently. These politically engaged Facebook users are much more likely to engage with a single candidate's page than comment on the pages of numerous candidates. This means users form small clusters which can mutually reinforce a particular perspective. Future projects around elections may have to consider how this clustering would influence the information individuals receive and how they interpret it.

Comparing online and offline data (CASE2015 data)

In order to compare online and offline data, we compared a selection of our twitter data with data from 796 CASE reports collected during the election and inputted them into Qlik. CASE stands for the 'Content Aggregation System for Election' and was a multi-stakeholder election system, which allowed citizen monitors to submit electoral observation reports via SMS / apps, to a central online system, allowing for real-time election analysis.. It was run by the Yar Adua Foundation, and sponsored by a number of organisations, including the MacArthur Foundation, the Open Society in West Africa, the Canadian Fund for Local Initiatives and DfID.

We had hoped to match the two datasets by time and find spikes in Twitter activity that matched CASE reporting. However, it was more challenging to make comparisons between the two data sets than had been anticipated.

A key difficulty was the difference in specificity: Twitter data was frequently very specific, but it was not always an easy task to link that to a category in CASE. For example, Twitter reported that on election day in Akwa Ibom armed political thugs fired shots and absconded with ballot boxes. Up to a hundred tweets (prior to retweets) reported this incident. CASE holds 26 records for those two days in Akwa Ibom. Parts of this marry up well: there are three reports of killings, a figure that is reported on Twitter too. However, whether 'disorderly conduct' (seven entries) or 'materials unavailable' (three entries) are referencing this incident specifically is not clear. This is because CASE data is based on a series of multiple choice questions that monitors answer, relating to events in and around the polling unit. For example, whether materials have arrived on time; whether voting has started; and whether there are incidents and/or intimidation in the vicinity of the polling unit.

Twitter, in contrast, provides an enormous amount of data – some of which might not be accurate – and also provides a great deal more 'colour'. Twitter data, for example, hints that APC supporters and journalists were targeted by violence in certain parts of Nigeria, it provides photographic evidence and suggests that the violence may have been two-way. However, as this is impossible to verify without the use of external sources, a method for verifying these details is imperative (and beyond the scope of this report).

In order to test the similarities and differences in these data sets, we took 15 prominent CASE reports and compared them to Twitter data. We then took 15 Twitter reports, and compared them to CASE data.

Eight of the 15 incidents reported in CASE did not seem to appear on Twitter, although as made clear, linking a report of threats to the myriad Tweets surrounding violence is not easy. (CASE has pre-determined categories, and Twitter of course does not.)

Table 4 - CASE2015 reports cross-referenced against Twitter

	Reports in CASE	Where?	Reported on Twitter?	Twitter Report
1	Arson	Zamfara	No	
2	Killing	Katsina	Yes	in katsina 2 shot, youths disrupting election materials distribution
3	Destruction of Property	Katsina	Yes	independent national electoral commission office in katsina was attacked and electoral materials vandalized
4	Disorderly Conduct	Taraba	No - only in News Reports	
5	Arson	Abia	No	
6	Disorderly Conduct	Jigawa	Yes	election violence in jigawa state at gwiwa local government.they are pdp thugs.
7	Disorderly Conduct	Bauchi	Yes	breaking: n'east poll - gunmen attack bauchi polling unit
8	Vote Buying	Bauchi	No	
9	Destruction of Property	Kaduna	No	
10	Delayed Voting Process	Kano	Yes	45 minutes after polls: no pvc readers at 3 polling stations in kano. no inec officials at one. everyone waits
11	Physical Assault	Kano	Yes	#nigeriadecides: ballot boxes destroyed as fight break out at tudun makera unit, dala lg, kano state [photo]
12	Ballot Stuffing	Kano	No	
13	Voter Disenfranchisement	Kebbi	No	
14	Threats	Lagos	No - only in News Reports	this just came in : sporadic gunshot around agege area of lagos. #besafe #votenotfight
15	Underage Voting	Benue	Yes	underage voting! [photo]

We then took a random sample of tweets we felt to be first-hand reports of violence that were circulated and compared the sample to the CASE data for a possible match. Fourteen of the 15 events we compared to Twitter data did not appear as CASE reports.

Part of the difficulty in making this comparison is that CASE is designed to collect data on polling units and so not all of the above Twitter data would appear on the CASE data set. However, of the five tweets below that mention booths (1, 3, 7, 8 and 15), only one appears in the CASE data set.

Table 5 - Twitter reports cross-referenced against CASE2015 data

	Reports on Twitter	Where?	Reported in CASE?
1	man running away with inec ballot box	Zamfara	No
2	gunmen kidnap zamfara emir from palace	Zamfara	No
3	pdp thugs disrupt distribution of materials	Katsina	Yes
4	violence rages in taraba, three killed	Taraba	No
5	taraba under attack as inec office, ssgs house geta burnt	Taraba	No
6	serious shooting in front of our house in umuahia, abia state. ochendo by pass. bende road. kindly rt and help us contact the police.	Abia	No
7	army carting voting boxes away all over. are they safeguarding the materials for us to continue voting tomorrow? #ukwa #abia @inecalert	Abia	No
8	a pdp thug who snatched a ballot box yesterday in jigawa. just look at his face. ðÿ˜†ðÿ˜, http://t.co/7hadfxdday	Jigawa	No
9	soldiers just shot a young chap in front of inec bauchi. allah rest his soul, we have to calm nerves and avoid a violent reaction.	Bauchi	No
10	just spoke to someone in bauchi. pdp deliberately created a false attack by bh to impose a curfew as the gov massively lost at pus @abinjnr	Bauchi	No
11	boko haram heading towards bauchi frm gombe, aim is to disrupt collation.. na are on ground heading towards gombe road..@ogundamisi @ayourb	Bauchi	No
12	one dead as military opened fire on innocents in markaz tudunwada kaduna just now	Kaduna	No
13	unknown gunmen in camouflage shot 2 to death in t/wada area of kaduna	Kaduna	No
14	gunmen open fire in market killing 14 in kaduna	Kaduna	No
15	all i heard was gunshots,den screams........well,ballot boxes have been carted away..makurdi lga benue state.smh	Benue	No

In order to provide a more granular analysis, we also took the CASE data for one region – Sokoto – as the basis for a second short comparison study. In total, 76 reports of violence were filed in CASE for Sokoto; and we identified 946 tweets containing the word 'Sokoto' in the violence dataset. This ratio of 1:12 is the fourth lowest by region, suggesting that either Twitter users were not tweeting about Sokoto as frequently as elsewhere, or they were using different words, perhaps referencing sub-divisions of the region; and / or that there were fewer Twitter users tweeting in Sokoto than in other states. In total, 762 unique users contributed to the dataset.

The Sokoto dataset on Twitter was dominated by one event, the kidnapping of the State governor's brother. This was widely reported, at first as conjecture and then in news reportage quoting the police. Once the dataset was filtered to unique references to incidents of violence, Twitter picked out 11 events. They are presented below (where an incident was retweeted or reported multiple times, only the first instance is quoted).

Table 6 - Tweets about Sokoto

sokoto state governor aliyu wamakkoâ€™s brother, salihu barde, has been kidnapped in sokoto.
oh my god fire outbreak at neighbouring house at polling unit. thinking it is a bomb #sokoto #nigeriadecides #nigeria2015 @carmenmccain
a ballot box was snatched by political thugs 17h51 at pu 002 rum far sharu, maikusa, sokoto state @inecnigeria #nigeriadecides @sanidtee
bad news......someone in a car just crushed number of people to death in my area in sokoto while celebrating gmb's victory @apcnigeria
16 people died and many more injured after two cars collided with each other, during buhari's victory celebration in sokoto.
gunshots in sokoto. meke faruwa ne?
many peoples lost their lives while jubilating gbm's victory and many more injured in sokoto yesterday...may their soul rip ameen....
military presence on major roads at sokoto causes confusions
not only two die in lugbe airpot road abuja 2 die there n in shuni lga of sokoto 2 die there again @daily_trust @thisisbuhari
crowd trouble in stations in #sokoto #nigeriadecides
i was attacked yesterday in sokoto by pdp area boys they damaged my car and wanted to kill me god save me i escaped!

Matching these to CASE reports isn't a perfect science: we do not know if the same incident is being referenced in both sets. Nevertheless, the biggest disparity in this Sokoto data is in the number of cases of violence reported in CASE that were not referenced on Twitter. This could be explained by the types of incident being reported in CASE: it is possible that incidents of 'disorderly conduct' (40 incidents in CASE), 'physical assault' (21 incidents in CASE), 'threats' (8 incidents in CASE) and 'intimidation' (4 incidents in CASE) are not deemed by eyewitnesses as worthy of reportage, with their focus instead on electoral fraud and major incidents of violence like abductions and killings. Similarly, there is no abduction in Sokoto in the CASE data, despite it being reported hundreds of times on Twitter. Again however, the CASE reporting format is not designed to report with this kind of specificity.

4 IMPLICATIONS

We address four key areas, which together cover the practical application of these technologies and approaches to election work; and the technical aspects of methodology.

How effective is social media as a voter education / monitoring tool?

Twitter and Facebook can both play an important role in a number of ways, not all of them obvious. It is difficult to determine how 'effective' social media is as a monitoring tool, but more useful to set out what likely uses of this type of software and techniques might add value. Based on this pilot, we consider there are five broad capabilities that social media and social media monitoring can provide:

1) *Better understand the network of influencers / journalists / commentators.* Network maps are relatively easy to construct, and provide a useful illustration of influential accounts within a data set – either to measure one's own position or to identify other important stakeholders talking on a subject. In the research, we found it a relatively simple task to identify influential and vocal accounts on the subjects researched - some of whom may not have been identified without using these techniques.
2) *Ability to detect and characterise unexpected events quickly as they occur (for example, violence).* In any major event, social media users now routinely share information and real-time commentary. Although in the 2015 Nigerian elections there was little violence, this is not always the case. Social media is an increasingly important way in which information and insight is available into these events as they take place. Classifiers allow for the rapid identification and classification of relevant information, including data which is in-country. The research has found there is a vast volume of data about elections, but only a small proportion might be of value, depending on the task. Classifiers on Twitter are useful for identifying very specific examples of data within the larger data sets (i.e. they were accurate in terms of finding and extracting content on a given subject from within a much larger data set). In the event of a major or significant disturbance, this is likely to be extremely valuable. The other advantage is that the data is real-time, which marks it out from almost any other form of reportage or mode of research. But formidable challenges remain in verifying and corroborating the data, especially with other information flows present during elections.
3) *Use to monitor electoral misconduct.* Social media is likely to be a very rich source of data relating to misconduct, violence and other election related issues not

recorded elsewhere. The system of classification of data provides a method for monitors to triage reports and identify where to direct resources to verify reports. While useful in real time, this may also be a useful longer-term research tool to better understand the general issues identified during an election, which can inform commissioning decisions for subsequent elections.

4) *Track and respond quickly to rumours or misinformation.* Rumours – especially those about electoral misconduct and violence – can be spread quickly and be seen by significant numbers of people. These rumours can be fairly quickly identified using automated methods. Given the potential damage that inaccurate rumours can do, one application of this technology is to spot and (then potentially) respond to rumours via official accounts or channels.

5) *Evaluation tool to complement existing evaluation systems.* With careful use of classifiers, it is possible to use social media as a new source of evaluation data to better understand the impact of certain projects. For example, classifiers would be able to identify relevant data about a given project, which (unless of exceptionally high volume, which we do not anticipate if it refers to a specific project) could then be manually and qualitatively analysed, and used as a supplementary source for existing evaluation methods. It can also help measure the formation of new collaborations and partnerships, such as whether groups speak more to each other, and whether companies cooperate more in terms of sharing information, supporting each other's campaigns and communications.

How effective is communication by social media at reaching an offline community?

On Twitter, over the period 18 March – 22 April 2015, researchers collected 13.6 million tweets posted from 1.38 million unique users, of which 12.4 million were about the Nigerian elections.

We are confident that 2.91 million tweets were identifiable as being from Nigeria (71 per cent of those which included some location data). On Facebook there were 216,000 unique users who contributed around 383,000 comments (around 1.7 per user on average) to the public pages we identified. When compared against known demographics of social media users, it is possible to draw a number of general conclusions about how representative users are:

- Awareness and use rates are much lower among older and less educated Nigerians.
- Men are significantly more likely to use social media than women.
- Although breakdowns of social media penetration by region are not available (we have not been able to locate any), the north of the country has less internet availability than the south, suggesting that there will be an intrinsic bias in social media data towards southern states.
- There is a significant proportion of 'organisational' accounts in the Twitter data set.
- 2.91 million tweets were identifiable as being from Nigeria (71 per cent of those which included some location data). Of the 4.1 million tweets, 1.14 million came from users in Lagos and 454,000 from Abuja. Only around one per cent of all tweets included a precise latitude-longitude geo-tag.

However, representativeness is mainly of importance and value when conducting large scale survey research, where population level inferences are made on the basis of smaller samples. As we have argued above, this is not necessarily the best way of using social media for monitoring and communication purposes.

Specifically in relation to Nigerian social media users (although not a perfectly representative sample) it appears there are a number of quite different types of users, and it does not make sense to consider them simply as citizens. We think they can be separated, at least in respect of election related data, into the following general categories:

- Citizens
- Citizen-journalists
- Media bloggers
- Official news outlets
- NGO account / Community and Voluntary Organisation accounts
- Political campaigners and supporters

It is our view that social media – both Facebook and Twitter – will remain a significant part of all future elections. The involvement of social media is increasing quickly, partly driven by mobile access. The elevated position of social media in Nigerian society and public life can also be seen from the changing nature of news websites. The third most visited site in Nigeria, Sahara Reporters, relies heavily on reporting by citizen-journalists for its content and has been at the forefront of publishing multimedia content on social platforms including Twitter, Facebook, Instagram, Tumblr and YouTube. Its popularity as a news platform (with over 1.5 million likes on Facebook) is testament to the influence that social media can have.[35]

How reliable and robust are the data?

The research was able to find that there was a very large volume of potentially useful data available on Twitter that was not available through other sources. In very general terms, most social media data tends to share the following broad features, which are useful in terms of deciding when and where it is (and where it is not) a potentially useful source. There are several beneficial features of social media data sets.

- *Relational:* Because most social media platforms are premised on curated networks of users, most data include some information about the relationship between users. This can take several forms: for example, if a user follows another user, has posted to another user, has interacted with another user, or has shared another user's content. What these relationships mean remains an open research question.
- *Real or near-real time*: Many social media platforms allow data to be collected as soon as it is posted. For example, on Twitter, researchers can access tweets as they are posted by users, making real-time research work possible.
- *Reactive and indirect:* Social media is often a reactive source of data; a space where people react to an event – either online or offline. This creates a dynamic relationship between media reports and stories and broader conversations which take place afterward, and creates new challenges in respect of accurately determining opinions and attitudes, which are often indirectly expressed.

However, there are a number of weaknesses in relation to reliability and robustness that need to be considered.

- *Unpredictable:* It can be extremely difficult to predict in advance the likely volume, data quality and subject matter of social media data on any given subject. This can make it difficult to plan in advance what topics and subjects can be researched. Before researching this project we did not know how much data to expect, which had implications for server space required to host the data.

- *Lacking in measurable demographic variables:* Unlike traditional research work, social media data is variable and unpredictable in the availability of demographic data relating to users. Facebook has more demographic data about users – but this is not available via the public API. Twitter contains very little in the way of demographic data, although location and gender can often be either directly measured or inferred using algorithms.

- *Reliability*: There are a number of difficulties in verifying social media data sets. First, much of the data is user-generated (including some of the demographic data) which cannot easily be independently verified. We are not able to estimate the extent to which users accurately describe themselves on social media accounts, for example. Second, it can be difficult to determine the likely accuracy of any claims made by users. In our data set, we found cases of inaccurate stories or information being shared among users. In some cases – such as when investigating various claims of electoral misconduct on Twitter – it was very difficult to independently verify the claims that were made. (There is a large literature on misinformation and inaccurate stories being shared on Twitter.) Third, where classifications of data are concerned using automated systems, results tend to be probabilistic in nature. The NLP algorithm categorises data into categories based on 'training data' provided by an analyst. Because of the nature of language, it not able to do this with 100 per cent accuracy for every piece of data. High performing classifiers can be expected to perform at around 70 per cent accuracy. This means that, over an aggregate of thousands of pieces of data, broad patterns are still likely to be accurate. However, it also means there will be many occasions where an individual piece of data (in this case a tweet) will be inaccurately classified.

- *Comprehensiveness*: The major weaknesses relating to Twitter and Facebook data is how data are collected. The research team conducted a very careful and extensive data collection effort using different key word searches during the election period, and carefully selected pages on Facebook. Just using hashtags or obviously election related words would have resulted in a far smaller data set. We collected around half the tweets (6.2 million) from the 'generic' collection which included search terms like 'Nigeria' and 'Lagos'. Despite this effort, we are not able to determine how much other data we might have missed by virtue of the search terms we employed.

Generally speaking, we think social media offer a valuable supplementary source of insight and information, alongside traditional (and better established) sources of research such as polling or focus groups. It is imperfect, but still incredibly useful if used for the right purposes.

What are the methodologies that prove effectiveness and how do they work?

Automated data analytics, key word sampling and natural language processing (NLP) classifiers create a number of new methodological challenges. In this section we set out how well the technology performed, and some ethical considerations when conducting work of this nature.

Each classifier trained and used for this paper was measured for accuracy. In each case, this was done by (a) randomly selecting 100 tweets, (b) coding each tweet using the classifier, (c) coding each tweet manually with a human analyst, (d) comparing the results and recording where they agreed and where they did not.

There are multiple outcomes of this test. Each measures the ability of the classifier to make the same decisions as a human – and thus its overall performance - in a different way. 'Recall' is number of correct selections that the classifier makes as a proportion of the total correct selections it could have made. If there were 10 relevant tweets in a dataset, and a relevancy classifier successfully picks 8 of them, it has a recall score of 80 per cent. 'Precision' is the number of correct selections the classifier makes as a proportion of all the selections it has made. If a relevancy classifier selects 10 tweets as relevant, and 8 of them actually are indeed relevant, it has a precision score of 80 per cent. All decisions (categories) within classifiers are subject to a trade-off between recall and precision. Decisions with a high recall score tend to be less precise, and vice versa. Each category has an 'F1 score', which reconciles precision and recall. The classifier as a whole has an 'overall' score, which is an average of the F1 scores of each decision, creating one holistic metric. This is displayed by the 'overall' score in the table below. 'Overall score' describes the proportion of the dataset that is expected to be correctly classified. We have not included recall and precision.

Below are the results of all the classifiers built for this research. A classifier performance of over 70 per cent is generally considered to be 'best practice' in the field of natural language processing algorithms. The results below suggest that, on the whole, the classifiers performed the tasks well.

Table 7 - Classifier performances

Reports & Comments Classifier		
Labels	Accuracy	Prior Multiplier
Reportage	0.81	1
Comment	0.71	0.6
Irrelevent	0.67	1
Overall	0.75	

Political Comments Classifier		
Labels	Accuracy	Prior Multiplier
Political	0.78	1
Other	0.73	1
Overall	0.74	

Reports; Problems Voting Classifier		
Labels	Accuracy	Prior Multiplier
Problems Voting	0.82	1
Other	0.81	1
Overall	0.81	

Problems Voting; Problems Voting Classifier		
Labels	Accuracy	Prior Multiplier
Card Reader	0.92	1
INEC	0.71	1
Accreditation	0.62	1
Other	0.57	1
Overall	0.69	

Reports; Results Classifier		
Labels	Accuracy	Prior Multiplier
Results	0.73	1
Other	0.71	1
Overall	0.72	

Reports; Violence Classifier		
Labels	Accuracy	Prior Multiplier
Violence	0.82	1
Other	0.79	1
Overall	0.81	

Violence; Violence Classifier		
Labels	Accuracy	Prior Multiplier
Boko Haram	0.74	1
Bombing	0.88	1
Electoral Violence	0.57	1
Other Violence	0.46	1
Other	0.57	1
Overall	0.69	

Generic: Demographic Classifier		
Labels	Accuracy	Prior Multiplier
Female		
Male		1
Institution		
Overall	0.75	

The challenge of this technology is that performance can be highly differential. In some cases, NLP of this type performs extremely well in classifying data sets. In other cases, for example with more event-specific data (such as 'electoral violence', 'other violence' or 'accreditation'), it is more difficult for NLP tasks to be performed accurately. The reason for this is that language and meaning is highly contextual, and the scale of the technical challenge posed depends on the specific context in which it's being applied. In the context of this study, it was possible to identify and accurately categorise relevant data within what is considered best practice for the types of task tested.

Ethical considerations

Research ethics are an increasingly important consideration when undertaking research or 'monitoring' using social media. They are not legally binding, rather a set of commonly agreed principles by which academic research institutions undertake research.

The distinction between surveillance (usually covered by law, in the case of the UK, the Regulation of Investigatory Powers Act 2000), monitoring, and research is not always entirely clear. The extent to which legal or ethical considerations come into play is likely to be driven by which organisation is conducting the work, and for what purposes. For example, professional regulatory bodies – such as market research guidelines – also adopt similar principles, albeit with slightly different rules that govern practice.

Research ethics (typically the most relevant for this type of work) aim to measure and minimise harm, and in this instance balance the need to undertake socially useful research against possible risks to those involved. In the UK, the standard best practice is the Economic and Social Research Council (ESRC) ethical framework, composed of six principles. Social media research is a new field, and the extent to which (and how) these ethical guidelines apply practically to research taking place on social media is as of yet unclear. Because the nature of social media research is highly varied – ranging from large quantitative data analysis down to very detailed anthropology – there is no single approach that can be applied.

We consider, drawing on the ESRC model, that the most commonly applied principles for human subject research are: a) any possible harms to participants must be measured, managed, and minimised and b) informed consent should be sought when and where possible. The issue of whether 'informed consent' is required on open public social media data sets, and how that can be reasonably achieved, remains perhaps the biggest and as yet unresolved debate in social media research.

Generally speaking, while such research should not be undertaken without particular caution and consideration, research without informed consent can be justified when no details about an individual are likely to be divulged, and where the risk of harm to research subjects is fully minimised. Even with open source data, however, certain conditions still ought to be met. Therefore, any use of this type of technology or monitoring capability should be done with careful consideration of its ethical use; ideally with reference to Government Social Research Service's Professional Guidance: 'Ethical Assurance for Social Research in Government'.[36]

Recommendations

How social media should be used for election monitoring and other activities (for Nigeria and elsewhere)

Based on this pilot study, we recommend that social media research would be best used in a number of ways for election monitoring and education. We have separated out 'election related' uses and 'generic' uses for this type of work, although stress that the distinction is in some cases misleading, since social media uses during an election will also be valuable in other contexts, and vice versa.

Election uses

- Given the likely presence of social media in future elections in Nigeria and elsewhere, social media provides an excellent opportunity for citizens to contribute to election monitoring. Campaigns in the lead-up to elections or other flash points to educate people on how to use social media to monitor / communicate (for example, using certain hashtags, accounts, adding location data) could increase the volume of citizen engagement in the election process. We believe active and supported citizen involvement in election monitoring could dramatically increase the volume and accuracy of data collected.

- Supplement existing media monitoring research work, by analysing social media to identify multimedia citizen generated information about events. Most organisations undertake some kind of media monitoring work, and it is relatively easy to include social media in that effort – for example accessing Facebook's public API and collecting posts from selected pages.

- In respect of the CASE data (or any other monitoring system), several reports of violence and electoral misconduct were reported on Twitter which were not recorded in the CASE data set (although in many cases, it was not to be expected they would be recorded). We did not verify these accounts – although *prima facie* there were thousands of reports on Twitter, and it is likely a proportion of them would warrant further investigation. This is particularly the case where certain users of social media become established and trusted sources of information on a subject. (We highlighted one such user, above.) We believe that automated Twitter analysis can, to a high degree of accuracy, identify citizen reports about electoral misconduct. This can be either a) new, unrecorded reports or b) colour and detail for already recorded reports. Each would require a slightly different collection effort: the first would require a large-scale collection and classification task, the second would be post-hoc searches. Both will require careful manual cross referencing and verification.

- Identify citizen attitudes and concerns – but as a rough guide and indication of general opinion, rather than a conclusive account. It is important not to see Twitter or Facebook as a representative sample of adults. However, it is a valuable resource to supplement existing research methods. For example, analysing these data sets in the way set out above would allow researchers to identify the most common complaints or concerns about electoral misconduct; or better determine citizens' experience of voting. Twitter in particular offers a way to assess citizen views on a scale hitherto not possible. That can be used, with accompanying demographic data and caveats, as a valuable research tool.

Generic uses

- Identify potentially influential voices and accounts which might be important to engage with or listen to. This can be done quickly and easily using freely available software: using 'R' or Python to collect data from Twitter's free API and using Gephi to produce the network maps. This could result in maps that identify key users in clusters that could build a valuable picture of political tweeters, citizen-journalists, bloggers, and other users likely to be useful to engage with. One specific example would be to investigate some of the rumour-peddling accounts to determine the extent of their relationship to the electoral process; are they regular citizen reporters, members of CVOs, or affiliated to the political process in any way? Roundtable discussions with the identified clusters of social media users could yield fruitful results.

- Identify and understand emerging events. 'Event' in this case does not only mean a physical 'real world' event; it can also refer to a purely digital event, such as a rumour. During an election, it is very difficult to predict what will occur. One valuable use will be to set up a very large data collection effort and use classification to split data into more manageable categories which can then be subjected to manual analysis. It is extremely difficult to 'spot' an event on Twitter earlier than with simple manual Twitter searches, but the use of classification will allow for better characterisation of an event to quickly understand what is happening on the ground.

- Better understand the possible reach and activity of certain organisations or movements – particularly those directly supported by DfID. Network maps and analysis allow DfID and others to determine, at a fairly general level, how far their online campaigns have reached and where these campaigns exist in the larger network. This could become part of evaluation processes, although it is important this is used critically. Although useful, it does not necessarily equate simply to offline impact.

Practical next steps

What extra research or investment can most contribute?

This research has been conducted using proprietary software, developed by the research team in partnership with the University of Sussex. Although the techniques used would be the same – natural language processing and network analysis – other research groups would likely have their own methods and may have reached different conclusions. This reflects that there are no accepted research methods and approaches in this field, and no widely used software. Nor, to our knowledge, are there any technological suites specifically dedicated to electoral monitoring.

There are a lot of 'off-the-shelf' technologies provided by commercial companies, and some free software such as 'R' and 'Gephi'. These are already available often at low or reasonable cost, typically a monthly subscription model, and they tend to be reasonably sophisticated at a number of tasks that are of value in this context. Specifically:

- Understanding the spread of messages;
- Very broad sentiment analysis;
- Network analysis.

However, there is an additional layer of possible capability that remains at the technological and academic coalface which we do not think is currently available as a market offering. In our view, the sort of nuanced and granular analysis required for election monitoring work would be best served by research collaboration creating bespoke software; potentially with one of a growing number of academic departments that specialise in social media research methods, or specialised consultancies, where social media analysis experts work alongside development and electoral monitoring subject matter specialists. Specifically:

- Situational awareness;
- Detecting events;
- Bespoke and finely tuned classification which can be applied to electoral monitoring.

Steps for commissioning work

In the event that DfID considers commissioning a capability (or a consultancy arrangement, where insight is delivered rather than a capability) in order to use social media research for election monitoring and education, we advise that the following steps be taken:

- Invest in developing internal expertise in how social media monitoring works, ideally combining a number of disciplines, most principally machine learning, computer programming, and social sciences. In particular, we recommend that analysts at DfID invest in internal knowledge building about: basic principles of API data access; how to use open source software 'R', and 'Gephi'; fundamentals of natural language processing, network analysis, social media research ethics, and programming.

- Review the range of social media monitoring tools currently available (both free and paid), and the current API access rights to the major social media platforms. Determine which of these capabilities can be satisfactorily done using free software, in house.

- Any tenders written or commissioning decisions taken should include input from individuals with a good knowledge of the strengths and weaknesses of social media monitoring software and methodology. We recommend a healthy level of scepticism about these offerings – especially those that are not open and transparent about every stage of the research / monitoring process. In particular, attention should be paid to: which platforms might data be collected from, and how comprehensive is the access; degree of control and flexibility over how data is to be classified; extent of demographic insight likely to be available with any data; how performance will be measured and shared (for example, with a 'gold standard' score); how any data will be stored and shared.

- Carefully review all research ethics guidance likely to be relevant for this type of work, and set out some broad guidelines to ensure any activity is conducted to high ethical research standards.

ANNEX: technical methodology

This paper collected tweets via Twitter's 'stream' and 'search' application programming interfaces (APIs). These allow researchers to collect publically available tweets. The 'search' API returns a collection of relevant tweets from an index that extends up to roughly one week in the past. The 'stream' API continually produces tweets that contain one of a number of keywords to the researcher, in real time as they are made.

The Twitter data that was collected was too large to be manually analysed or understood in its totality. We use NLP 'classifiers' that are trained by analysts to recognise the linguistic difference between different categories of language. This training is conducted using a technology developed by the research group to allow non-technical analysts to train and use classifiers called 'Method 52'. Method52 is a software suite developed by the project team over the last 18 months. It is based on an open source project called DUALIST which enables non-technical analysts to build machine-learning classifiers. The most important feature is the speed wherein accurate classifiers can be built. Classically, an NLP algorithm would require roughly at least 10,000 'marked-up' examples to achieve 70 per cent accuracy. This is expensive and takes days to complete. However, DUALIST innovatively uses 'active learning', an application of information theory that can identify pieces of text that the NLP algorithm would learn most from. This radically reduces the number of marked-up examples from 10,000 to a few hundred. Overall, this allows social scientists to build and evaluate classifiers quickly, and therefore to engage directly with big social media datasets.

Table 8 - Search terms

Harvest	Harvest2	Harvest3	Victory Search
godson_okoye	#Nigeria2015	PollwatchNg2015	#BabaNowThatYouAreThere
presgoodluck	#Nigeria15	Supportaworkinggovt	BabaNowThatYouAreThere
ThisIsBuhari	#NigeriaElection2015	WhoElseIfnotBuh	#LetPeaceReign
DanielAkinlami	#NPFElectionSecurity	GoVotePDP	IVotedAPC
ArcNamadiSambo	#VoteNotFight	IfBuhariWinsNigeriaLoses	IVotedGMB
ProfOsinbajo	#BokoHaram	SituationRoomNg	#IVotedAPC
officialnedg	#APCPropaganda	YIAGA	#IVotedGMB
EiENigeria	#gejwinsit	ChibuikeAmaechi	#NigeriansElect2015
elnathan	#QuestionsForBuhari	realFFK	#ChangeisComing
texasinafrica	#Nigeriab4GEJ	abati1990	#SaiBuhari
Karo_Orovboni	#GEJ	doyinokupe	#ByeToJonathan
omojuwa	#GEJisit	delemomodu	#Sitroom2015
ImadMesdoua	#IwillvoteAPC	bukolasaraki	#FreshNigeria
sirvic2011	#VoteGMB	NuhuRibadu	#Verdict2015
Sara__Firth	#AskJega	elrufai	#NigeriahasDecided
eggheader	#GMB2015	AmplifiedR	
PoliceNG	#ThingsBuhariWillChange	ClementNwankwo	
tmgng	#UDP2015	CDDWestAfrica	
inecnigeria	#WhyIVoteBuhari	electmonitorng	
INECAlert	#BringBackGoodluck2015	cleenfoundation	
VanguardNgrNews		PDPNigeria	
BBCAfrica		APCForChange	
Ynaija		RethinkNigeria	
NTANewsNow		kowaparty	
channelstv		UDPNigeria	
NigeriaNewsdesk		APCNigeria	
#TellYourNeighbour			
#JonaMustGo			
#Bauchi4GMB			
#DoyouKnowGEJ			
#IChooseGEJ			
#VoteGoodluck			
#NigeriaDecides			
#LagosDecides			
#NigeriaElection			

Harvest Weekend			
Nigeria	Sokoto	Warri	Ijero
Chinedu	Okene	Ekiti	Otukpo
Owuru	Calabar	Okpoko	"Kisi"
Adebayo	Osogbo	Awka	Buguma
Okorie	Katsina	Sapele	Funtua
Sonaiya	Akure	"Ila"	Abakaliki
Galadima	"Ife"	Shaki	Gbongan
Okoye	Bauchi	Onitsha	Lafia
"Goodluck Jonathan"	Iseyin	Owerri	Igboho
Ahmad	Minna	Ugep	Amaigbo
Onovo	Makurdi	"Ijebu Ode"	Gashua
Buhari	"owo"	Gboko	Offa
Salawu	"Ado Ekiti"	Ikare	Jalingo
Eke	Ilesa	Bida	Bama
Akinlami	Gombe	Abeokuta	Borno
Anifowose-Kelani	Umuahia	Maiduguri	Uromi
"Kano"	Ondo	Benin City	Nsukka
Abuja	Damaturu	Zaria	Okigwe
Ibadan	Jimeta	"Jos"	Modakeke
Kaduna	"Ikot Ekpene"	Ilorin	"Osun"
"Port Harcourt"	Gusau	Enugu	"Enugu"
"aba"	"Mubi"	"oyo"	"Akwa Ibom"
Ogbomosho	Sagamu	Nasarawa	

NOTES

[1] See: http://www.civilservice.gov.uk/wp-content/uploads/2011/09/ethics_guidance_tcm6-5782.pdf

[2] Africa Practice (2014), *Social Media Landscape in Nigeria.*

[3] For an overview of these trends, see Bartlett, J. et al (2014), *Vox Digitas,* Demos.

[4] See: http://www.internetworldstats.com/stats1.htm

[5] See: http://www.ncc.gov.ng/index.php?option=com_content&view=article&id=125:subscriberstatistics&catid=65:industry-information&Itemid=73

[6] Miniwatts Marketing Group (2012), *Internet Usage Statistics for Africa*

[7] See: http://www.bbg.gov/blog/2012/08/20/new-bbg-gallup-data-shows-dramatic-rise-in-mobile-use-in-nigeria/

[8] See: http://www.itnewsafrica.com/2015/04/study-reveals-african-mobile-phone-usage-stats/. The most common mobile activities reported in Nigeria were using Facebook (58 per cent), browsing the internet (47 per cent), sending SMS (39 per cent), listening to the radio (36 per cent), instant messaging (34 per cent), playing games (34 per cent), downloading apps (28 per cent) and using Twitter (14 per cent).

[9] See: http://america.aljazeera.com/articles/2015/3/19/social-media-plays-key-role-in-nigerian-elections.html

[10] See: http://www.punchng.com/feature/nigeria-third-tweeting-nation-in-africa-report/

[11] See: http://www.bbg.gov/wp-content/media/2014/05/Nigeria-research-brief.pdf

[12] See: http://www.cmdconf.net/2015/pdf/2.pdf

[13] See: http://www.socialbakers.com/statistics/facebook/pages/total/nigeria/page-10-14/

[14] See: http://america.aljazeera.com/articles/2015/3/19/social-media-plays-key-role-in-nigerian-elections.html

[15] See: http://www.africapractice.com/wp-content/uploads/2014/04/Africa-PracticeSocial-Media-Landscape-Vol-1.pdf

[16] See: http://www.bbc.co.uk/news/blogs-trending-27026755

[17] See: http://www.academia.edu/8880871/The_Evolution_of_Social_Protest_in_Nigeria_The_Role_of_Social_Media_in_the_OccupyNigeria_Protest

[18] See: http://www.africapractice.com/wp-content/uploads/2014/04/Africa-Practice-Social-Media-Landscape-Vol-1.pdf

[19] Ajayi R.A., 'Social media - Weapons of choice for the Nigerian revolution', http://www.pambazuka.net/en/category.php/comment/72597

[20] 'Social Media and Elections: The case of Nigeria 2011', ICT4Peace Paper Series, http://mikeb.inta.gatech.edu/uploads/papers/Nigeria%20Best%20Maclay%20111011.pdf

[21] See: http://america.aljazeera.com/articles/2015/3/19/social-media-plays-key-role-in-nigerian-elections.html

[22] See for example the rise and fall of hashtags in the elections here: https://www.premiumtimesng.com/features-and-interviews/179304-nigerias-election-who-is-winning-the-twitter-war-by-tobi-oluwatola-2.html/attachment/3-6

[23] See: http://www.reuters.com/article/2015/04/02/us-nigeria-election-technology-idUSKBN0MT25I20150402

[24] See: http://penplusbytes.net/wp-content/uploads/2015/01/SEE-THIS-SOCIAL-MEDIA-AND-THE-2013-KENYA-GENERAL-ELECTIONSS.pdf

[25] Miller C. (2015) , 'Social Action on Social Media', Nesta Working Paper: Working Paper Series.

[26] It is worth noting that some news organisations and blogs operate local language portals – for example, Premium Times, the BBC and many international news organisations operate Hausa language services with a large following on Facebook.

[27] Given the very large volume of data, we occasionally used subsections or random samples from the overall data set.

[28] See: https://developers.facebook.com/docs/chat

[29] This may change as Facebook extends the way news is delivered to users.

[30] In this case, the report was dismissed by INEC as false, for reasons we do not know.

[31] See: http://www.vanguardngr.com/2015/04/lagos-poll-backlash-over-akiolus-meeting-with-igbos/

[32] This was taken from a random sample of around 400k users that had contributed to the data set.

[33] Wasserman, S. and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge; New York: Cambridge University Press.

[34] The account holder, however, receives this information automatically.

[35] http://america.aljazeera.com/articles/2015/3/19/social-media-plays-key-role-in-nigerian-elections.html

[36] See: http://www.civilservice.gov.uk/wp-content/uploads/2011/09/ethics_guidance_tcm6-5782.pdf

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1 Definitions

a 'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.
b 'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.
c 'Licensor' means the individual or entity that offers the Work under the terms of this Licence.
d 'Original Author' means the individual or entity who created the Work.
e 'Work' means the copyrightable work of authorship offered under the terms of this Licence.
f 'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work,or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2 Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3 Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive,perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:
a to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;
b to distribute copies or phonorecords of, display publicly,perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised.The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4 Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:
a You may distribute,publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phonorecord of the Work You distribute, publicly display,publicly perform, or publicly digitally perform.You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder.You may not sublicence the Work.You must keep intact all notices that refer to this Licence and to the disclaimer of warranties.You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement.The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licencor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.
b You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation.The exchange of the Work for other copyrighted works by means of digital filesharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

C If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works,You must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5 Representations, Warranties and Disclaimer

A By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

i Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;

ii The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

B except as expressly stated in this licence or otherwise agreed in writing or required by applicable law,the work is licenced on an 'as is'basis,without warranties of any kind, either express or implied including,without limitation,any warranties regarding the contents or accuracy of the work.

6 Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental,consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7 Termination

A This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence,however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

B Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8 Miscellaneous

A Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

B If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

C No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

D This Licence constitutes the entire agreement between the parties with respect to the Work licensed here.There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You.This Licence may not be modified without the mutual written agreement of Demos and You.

This material has been funded by UK aid from the UK government; however the views expressed do not necessarily reflect the UK government's official policies.