

# Inferring User Attributes and Extracting Political Sentiment from Nigerian Social Media

Clay Fink

The Johns Hopkins University Applied Physics Laboratory

The last few years have seen enormous growth in the use of social media worldwide. The developing world, especially, has seen high rates of adoption, with countries like Indonesia, the Philippines, and Mexico now in the top-10 of countries represented on Facebook. While much of the user generated content in social media relates to people's everyday activities, it also captures reactions to major events, opinions about public affairs, and as was clearly seen during the 2011 Arab Spring events, communications related to collective action.

There is significant interest in social media as a means to collect public opinion, especially in locations where the timely collection of such information is difficult. To maximize the utility of these data, however, it is essential to be able to characterize the population represented in social media from these countries or regions in order to determine whether the authors of this content are representative of the population, or represent a sample skewed by geography and demographic categories such as ethnicity and religion. There is also reason to suspect that opinion captured in social media is qualitatively different from that captured by surveys and polls, given that the expressed opinion is typically volunteered, rather than in response to a question.

We will describe work we have done toward building a picture of the population of social media users in Nigeria. Focusing on data collected from Twitter, we describe techniques for inferring a user's regional location within the country, as well as their ethnicity. We then focus on Twitter content about the 2011 Nigerian Presidential election, looking at its geographic distribution, extract sentiment toward the three major candidates, and compare these sentiments to traditional survey and polling results collected during the election campaign.

Our initial results for Nigeria show that social media use, at least as captured on Twitter, is skewed toward the predominantly Christian, south west region of the country (site of Nigeria's most populous city, Lagos), and toward the capital city of Abuja located in the central region of the country. The predominantly Muslim northern regions are underrepresented. The results of machine learning experiments for predicting ethnicity from user names correlate with ground truth ethnicity data from Nigeria, suggesting that such hidden attributes can be inferred from social media data. Tweets mentioning one or more of the major candidates for the presidential election were drowned out by other topical content and, as with the rest of the Twitter data from Nigeria, under-representative of the northern region of the country. Sentiment analysis of these tweets was predominately negative as well, and generally not in agreement with surveys and polls taken during the same period, supporting the argument that sentiment captured from social media may be capturing phenomena different from what is captured by traditional polling.