Article

International Journal of Market Research

International Journal of Market Research
2018, Vol. 60(2) 169–189

DOI: 10.1177/1470785317753025
journals.sagepub.com/home/mre
SAGE

Using interaction networks to map communities on Twitter

Kyle Findlay and Ockert Janse van Rensburg
Kantar Innovation Team, South Africa

abordagem baseada em rede para mapear os principais públicos que discutem tópicos específicos no Twitter

Abstract

This article summarizes our work using a network-based approach to mapping the main constituencies discussing specific topics on Twitter. The approach gives researchers unique insight into the main groups involved and their agendas. While the individual pieces of the methodology are not new, the way in which we combine them will be novel to many market researchers. By connecting Twitter users that interact with each other into an "interaction network" or "conversation map" and using community detection algorithms to isolate distinct groups, we are able to identify the main constituencies discussing a specific topic such as a national election. Once the main constituencies have been identified, it is possible to profile them in more detail such as in terms of their demographics, their influencers, and the type of content that resonates with them. This article specifically focuses on roughly one million tweets about the 2014 South African national election to illustrate our approach. The authors believe that the approach described has wide-ranging applications and that it can be used to give researchers unprecedented insight into the public discourse surrounding specific topics and events. This article is adapted from an earlier article presented at the 2015 Southern African Marketing Research Association's (SAMRA) annual conference.

Keywords

social network analysis, social media, Twitter, interaction networks, community detection, community mapping, constituency mapping, data-led segmentation, influencers, key opinion leaders (KOL), political research, South African politics, media partisanship

Introduction

The world has changed. The battle for hearts and minds is no longer just fought on the street by politicians canvassing door to door or at rallies. These days, social media play an important role in convincing potential voters who to vote for.

As market researchers, we are still playing catch-up when it comes to measuring these conversations, but, as they are becoming an increasingly important part of any holistic campaign or brand strategy, we would be remiss if we did not attempt to measure them properly.

In this article, we demonstrate the value gained from our analysis of roughly 1,000,000 tweets relating to the 2015 SA national elections, covering a period of roughly 2.5 months in the run-up to

Corresponding author:
Kyle Findlay, Kantar TNS Global Brand Equity Centre, Roggebaai Place, 13th Floor, 5 Jetty Street, Foreshore, Cape Town, 8001, South Africa.
Email: kyle.findlay@kantar.com

the elections (and roughly a week after). In order to flesh out the agenda within each community, we share insights such as descriptions of the main communities of interest that discussed the elections, the content of their discussions, and we highlight some of the influencers driving the discussions. The results provide a better understanding of the constituencies involved in the South African digital political landscape at the time and their concerns. This article should prove interesting to any engaged South African citizen and, even more so, to social and political market researchers. The approach described can be applied to any topic, including brands and campaigns.

Many researchers and clients alike view social media as akin to the Wild West. We become overwhelmed by the volume and messiness of the data, which prevent us for seeing the wood for the trees. By mapping a topic such as the South African elections using this methodology, we demonstrate that such conversations really are tractable. We show how the methods can be used to identify the specific constituencies and their concerns, so that parties (or brands) can more effectively surface hot-button issues and address them (or simply take cognizance of them). In addition, we show how these methodologies can be used to gain a better understanding of constituencies in areas such as their topics of conversation, their language use, their influencers and the type of media that resonates best with them in order to better tailor communications toward them.

Case study background

The South African political landscape

We evaluate the 2014 national elections in South Africa for the purposes of this article; therefore, it is worth giving the reader a brief overview of South African politics. South Africa is a representative democracy, wherein parties vie for 400 seats in Parliament. Prior to 1994, South Africa was ruled unopposed by the National Party who instituted apartheid in 1948 soon after coming into power. However, by the 1980s, it was clear to the then government that the apartheid system was not sustainable and so negotiations were opened with the African National Congress (ANC), and its leader, Nelson Mandela, who was still in prison at that time. Over a period of several years, talks were held to facilitate a transition of power from rule by the white minority-dominated National Party, to the majority's elected representatives through democratic elections. These elections were held in April 1994 with the ANC winning a landslide 63% of seats in Parliament. The ANC has been in power ever since. In recent years, however, the ANC's grip on power appears less solid. While most black voters initially gravitated toward the "liberation party" in the post-apartheid years, as South African politics has matured, we have seen some fragmentation of the landscape, which is now dominated by three main parties: the ANC, the Democratic Alliance (DA), and the Economic Freedom Fighters (EFF). Figure 1 summarizes each party's presence in parliament after the 2014 national elections.

Nelson Mandela's party, the ANC, is a leftist organization that has shifted toward the center over time as it has enjoyed unimpeded access to power and the economy. The party is currently led by President Jacob Zuma, a former head of intelligence for the ANC who spent 10 years imprisoned during the apartheid era. Under Zuma's watch, the party has experienced steadily eroding support and he has received much criticism for allegedly corrupt dealings (he is currently accused of over 700 corruption charges). Critics blame Zuma for tarnishing the legacy of "Mandela's ANC" and for driving former supporters to new parties such as the EFF. In addition, the party, which has traded on its role as liberator of the masses, finds itself faced with a younger electorate who were not alive during the time of apartheid (known as "born frees"), and therefore, this role does not resonate with their lived experiences. What is more important to many of them is access to the economy and a better life.

The DA is the second largest party in Parliament with 22% of seats. It evolved from a white liberal tradition and has come to be the home of middle-class voters of all races who align

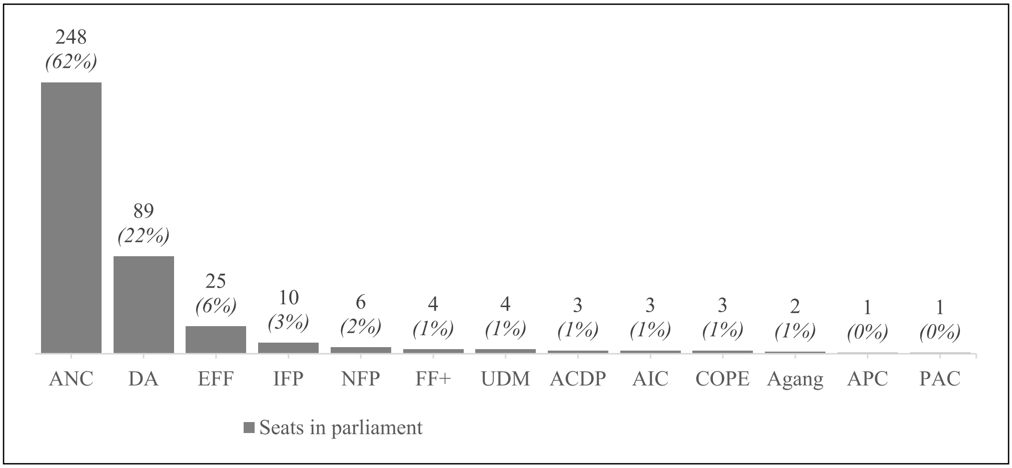

Figure 1. A distribution showing how many of the 400 parliamentary seats each party won in the 2014 elections.

themselves with a Western, free market ideology. They are sometimes pejoratively referred to by left-leaning critics as “neo-liberals” dominated by “white monopoly capital,” although their track record of good governance has made them attractive to a wide variety of South Africans. The DA leader is Mmusi Maimane, the organization’s first black leader. Maimane is a charismatic individual that has overtly modeled himself and his party’s communications after those of Barack Obama in the United States. Maimane has proven to be an articulate and vocal critic of President Zuma’s government in Parliament and this appears to have attracted many previously unreachable (i.e., black) voters to the party.

Finally, we have the EFF, a far left, socialist party whose message has found fertile ground due to the high level of inequality in South African society (South Africa has one of, if not, the highest income Gini coefficients in the world). The EFF, like many far left parties around the world, is dogmatically focused on the redistribution of wealth and resources, and their populist rhetoric has struck a chord with both poor, marginalized voters and newly middle-class, black voters. The EFF was formed less than a year before the 2014 elections and yet it still managed to garner over a million votes to capture 6% of seats in Parliament. Post the 2014 elections, they continue to enjoy strong growth in the South African political landscape. Their leader is Julius Malema, a firebrand activist from a humble background. Malema was formerly the head of the ANC’s Youth League, during which time he helped bring Jacob Zuma into power. He subsequently was expelled from the ANC and formed the EFF.

The South African Twittersphere

More than ever before, the 2014 South African national elections were fought on social media platforms like Twitter, where upstart parties such as the EFF were able to leverage instantaneous communication to make significant inroads in the battle for the minds of South African voters.

According to Fuseware’s *SA Social Media Landscape 2015 report* (Goldstuck & Wronksi, 2014), South Africa had 11.8 million Facebook users, 6.6 million Twitter users, 4.9 million Mxit users, and 1.1 million Instagram users around the time of the elections, while WhatsApp was the most popular app in the Android, Apple, and Windows app stores. The *TNS Sunday Times Top Brands Survey 2014* (TNS, 2014) reported similar proportions as summarized in Figure 2.

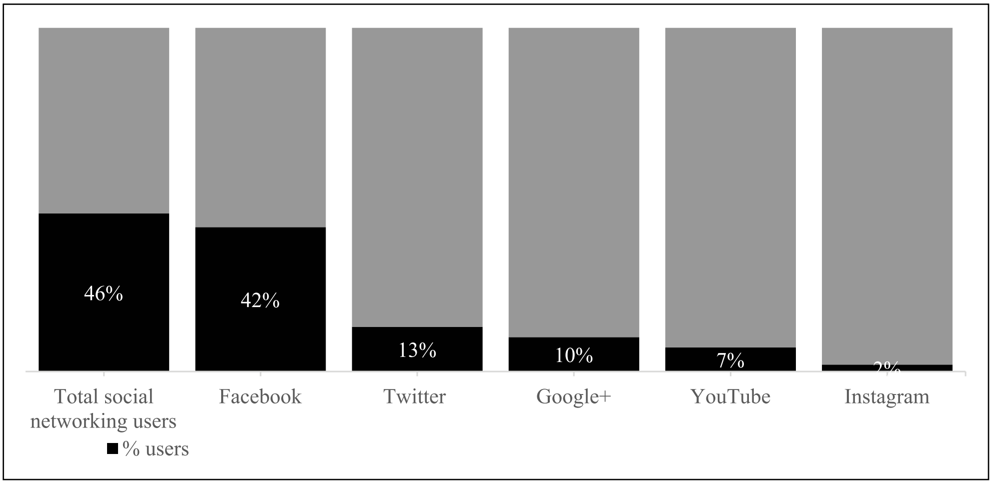

Figure 2. Almost half of South Africans (46%) used some form of social media in 2014 with Facebook being the most popular (42%), followed by Twitter (13%) (TNS Sunday Times, 2014).

We have focused on Twitter data in this article for two reasons:

1. It is the platform that is the most open with its data, which allows us to perform the most complex analyses on it.
2. It is the most popular venue for public political discussions including among influential users such as politicians, pundits, and news agencies.

While Twitter does not necessarily represent the largest active user base of all social platforms in South Africa, it has seen massive growth in recent years. In 2013, it grew by 129% to 5.5 million users, while in 2014 it grew a further 20% to 6.6 million users (Goldstuck & Wronksi, 2014). This obviously represents a slower, but still robust, growth rate, pointing to its continually growing relevance in a South African context. This corroborates the authors' own experience: having mapped the South African Twittersphere since 2011 (Findlay & Ooserveld, 2012), there has been a marked increase in activity and volumes. In addition, the Fuseware report (2014) found that Twitter, despite having fewer users, sees greater engagement per user (Facebook average per post: 19 likes and 1.5 shares; Twitter average per post: 0.4 favorites and 23 retweets).

Finally, Kantar TNS has found that the nature of Twitter users differs from other platforms (see Figure 3). According to their Connected Life 2014 report (TNS, 2014), Twitter has a higher proportion of "Leaders" (highly engaged users who are more digitally influenced) than Facebook (42% vs. 37%). Conversely, Facebook has more "Functionals" (users who are less engaged with social media and influenced by it) than Twitter (33% vs. 27%). This points to the differing nature of Facebook and Twitter users and corroborates the view of Twitter as the natural home of opinion leaders and generally opinionated people.

Previous research by Findlay and Ooserveld (2012) found that the top influencers in the South African Twittersphere were accounts such as *@helenzille* (then DA leader, Helen Zille), *@julius_s_malema* (EFF leader, Julius Malema), @sapresident (President, Jacob Zuma). Celebrities and sportspeople such as *@bonang_M* (radio and television personality, Bonang Matheba), *@dineoranaka* (radio and television personality, Dineo Ranaka), *@bryanhabana* (rugby player, Bryan Habana), and *@johsmit123* (rugby player, John Smit) rounded out the list (see Figure 4).

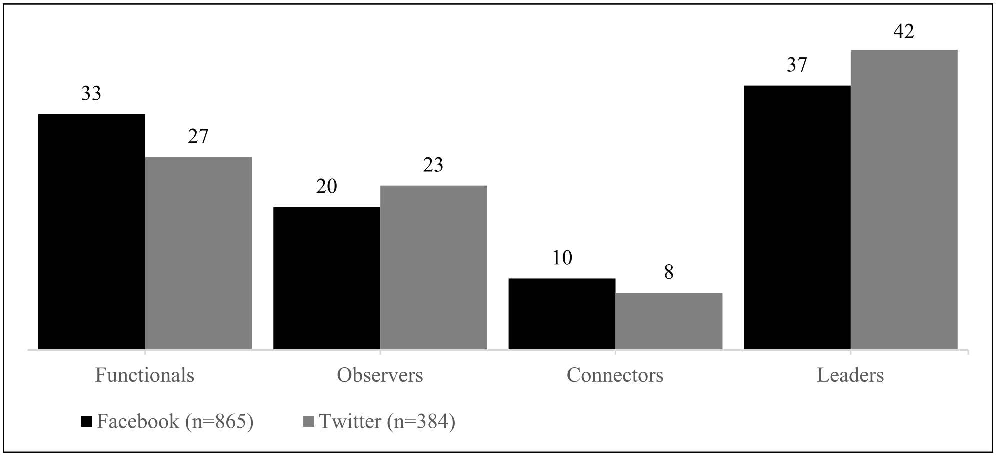

Figure 3. Summary of Connected Life segmentation of *TNS Sunday Times Top Brands Survey* (TNS) for Facebook and Twitter in South Africa (TNS, 2014).

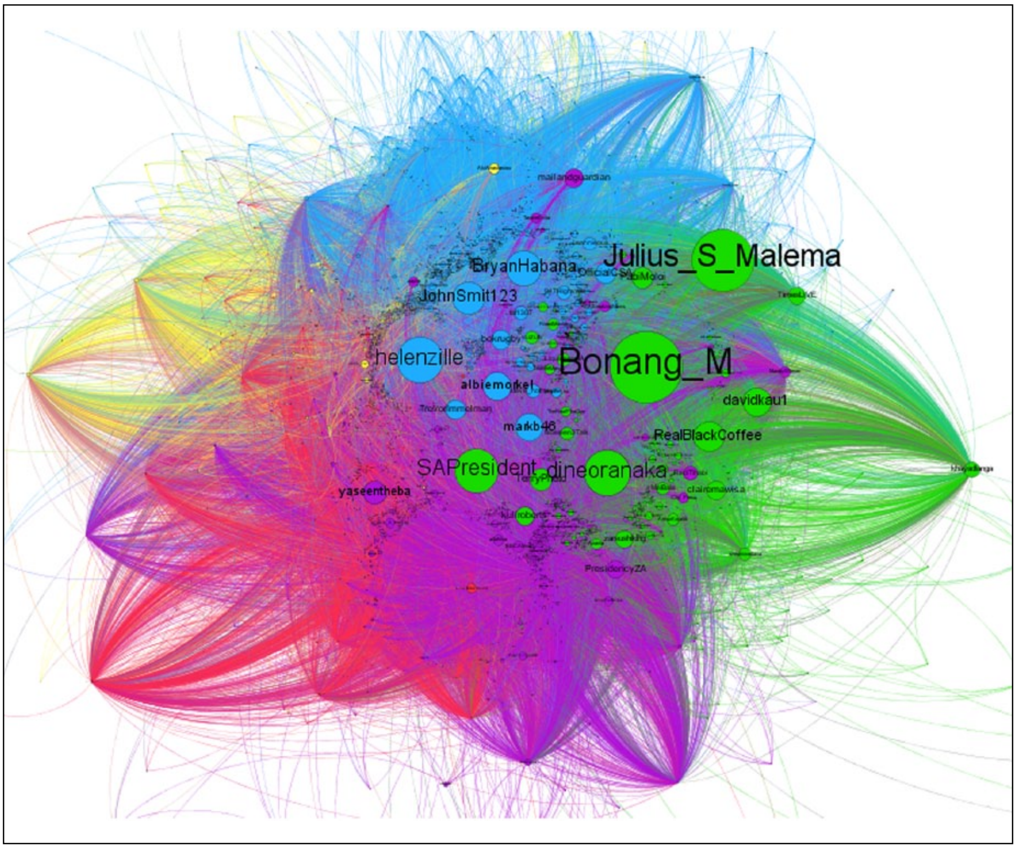

Figure 4. South African follower network from Findlay and Ooserveld (2012). Node size is based on number of followers. Blue: sportsmen and celebrities; purple: news media and political commentators; preen: ANC, related commentators and celebrities; yellow: tech pundits and entrepreneurs; red: proudly South Africa and DA supporters.

Table 1. Summary of the South African Twitter population versus the general consumer population.

	Total consumer market, %	Twitter, %
Survey, *n*	3500	442
Platform usage	N/A	13⇧
Race		
Black	77	70⇩
White	10	21
Colored	9	6⇩
Asian	3	3
Age, years		
18–24	23	35⇧
25–34	30	35⇧
35–49	30	26
50+	17	4⇩
Gender		
Male	48	56⇧
Female	52	44⇩
LSM		
LSM 1–4	13	1⇩
LSM 5–6	51	32⇩
LSM 7–8	21	35⇧
LSM 9–10	14	31⇧
Province (containing 3 major metropolitan areas)		
Gauteng	24	42⇧
Western Cape	11	9
KwaZulu-Natal (KZN)	20	15⇩

Source: TNS Sunday Times (2014).
LSM: living standards measure.
Values with ⇧ are significantly above the Total Consumer Market at the 95% confidence level.
Values with ⇩ are significantly below the 95% confidence level.

That study was based on a general look at the South African Twittersphere rather than specifically within the context of a national election. However, what is clear when one contrasts these two studies is that in the intervening years, large, robust, and distinct communities have coalesced around key political figures and parties as South Africans' Twitter usage has matured. Much of this would appear to be due to both the ANC and EFF recognizing the importance of the platform, which has allowed them to catch-up with the DA who were an early adopter of the platform and who dominated it in the early days. From the data presented in this article though, South Africa's political conversation is now clearly a three-horse race between the ANC, DA, and EFF.

Twitter representativity and its use in predicting election results

The Twitter population is not representative of the general population. As summarized in Table 1, it skews toward younger users, and it is these users who introduce the new dynamics into an otherwise staid political landscape.

Despite the lack of representativity, Tumasjan, Sprenger, Sandner, and Welpe (2011) found that the volume of tweets aligned with final political wins. Similarly, Sang and Bos (2012) found that counts can reflect final votes if the data are carefully collected and analyzed. Bermingham and

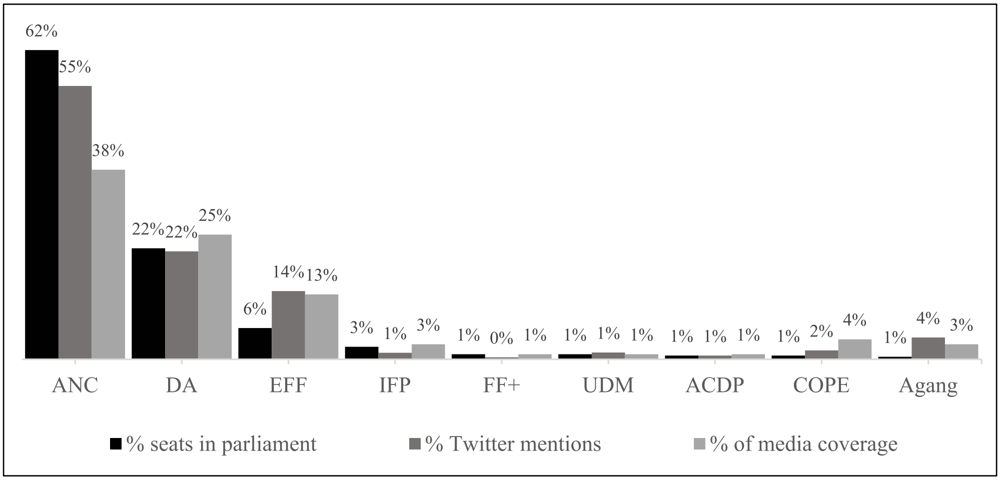

Figure 5. Strong correlations between the proportion of parliamentary seats that a party gained in the elections and the party's (1) share of Twitter mentions (R=.99) and (2) share of media coverage (R=.95) are seen. Media coverage figures come from Media Monitoring Africa's (2014) Elections Coverage 2014.

Smeaton (2011) also find that volume-based metrics can be predictive. However, Gayo-Avello, Metaxas, and Mustafaraj (2011), Gayo-Avello (2012), Chung and Mustafaraj (2011) and Mitchell and Hitlin (2013) are more skeptical about Twitter's ability to predict election results, warning that there is no silver bullet solution to doing so. Thus, the question remains open as to whether or not Twitter data can be used to predict election results.

A softer version of this question might be whether or not Twitter data can be used as an indicator of party momentum within the general electorate? The general Twitter population skews toward younger voters, and since South Africa's population is heavily skewed toward younger voters, can we use it to accurately gauge shifting sentiments among this part of the population?

It is difficult to evaluate this objectively, primarily because of the large volume generated by the EFF without any prior historical data to compare with (they had only been in existence for a few months prior to the election). However, if we consider that the EFF managed to gain over one million votes in this short time, and we look at the volume of chatter around them on Twitter, it would seem to be that their electoral momentum was presaged by the high volume of conversations mentioning them on Twitter.

Despite the above caveats, and despite this not being the purpose of this article, we share a distribution of tweet volumes versus seats won in Parliament as an interesting artifact from our research. A comparison of party tweet mention volumes (based on keyword matching of party-related terms) versus seats won in Parliament shows good alignment between the two sets of results (see Figure 5).

The main differences lie in the ANC being under-represented and the EFF being over-represented. These results could form the basis of an article themselves but for discussion purposes we simply point out that these differences appear to align with the current dynamics in the South African political landscape, that is, the ANC appears to be losing votes while the EFF appears to be gaining votes, mostly at the expense of the ANC rather than the DA.

Methodology

In broad terms, the following process was followed to produce the analyses in this paper:

1. Data was collected via Twitter's public REST API

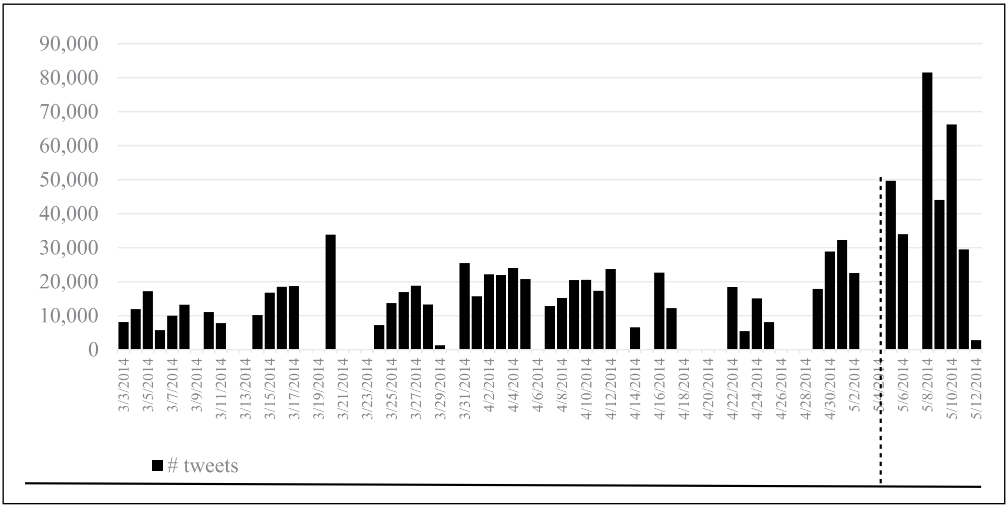

Figure 6. Volume of tweets collected in the run-up to the elections, which occurred on the 7 May 2014.

2. An “interaction network” was created whereby Twitter users were connected into a network graph when they interacted with each other by retweeting or @mentioning each other
3. The Louvain modularity algorithm was run on the network graph in order to identify community clusters
4. Individual communities were filtered on in order to identify the topics of discussions and key influencers in each community

The subsequent sub-sections will delve into the methodological specifics in more detail.

The data set

The data for this article were collected via Twitter’s public representational state transfer application program interface (REST API), which is subject to rate limits (i.e., you are limited to downloading x number of tweets per hour), so there is no guarantee that we collected every relevant tweet; however, we did end up with a substantial data set of 1,461,909 tweets. These tweets were garnered using the following search query that was designed to capture tweet mentions of the major political parties and their figureheads:

> *IEC, independent electoral commission, sa election, south african election, @elections_sa, #anc, #VoteANC, @MyANC_, @ANC_YOUTH, @SAPresident, @PresidencyZA, @Mgigaba, african national congress, anc, zuma, #DA, @DA_News, @helenzille, @LindiMazibuko, @DA_Youth, @MaimaneAM, @mbalimcdust, democratic alliance, zille, maimane, lindiwe mazibuko, mbali ntuli, #VoteEFF, #VoteDA, @EconomicFreedom, @Julius_S_Malema, @EconFreedomZA, malema, IFP, Inkhatha, Mangosuthu Buthelezi, @TeamCOPESA, @YourCOPE, @TerrorLekota, #COPE, Lekota, @AgangSA, @MamphelaR, Ramphele, Mamphela, Agang, UDM, Bantu Holomisa, United Democratic Movement, @Pieter_Mulder, @VFPlus, Freedom Front Plus, Pieter Mulder, Vryheidsfront, @A_C_D_P, ACDP, African Christian Democratic Party, dali mpofu, #BelieveGP*

We started collecting data just over 2 months before the elections (beginning of March 2014) and continued until a week after the elections (7 May 2014) (see Figure 6 for daily volumes). Unfortunately, there were certain days during this time where our system crashed and so we missed tweets for some days.

Our final data set consisted of 981,878 tweets specific to the South African elections. Reaching this final data required a large amount of data cleaning. Due to the ambiguous nature of social media text data, our search query captured many irrelevant tweets, including tweets relating to Indian, Turkish, and American politics. Bearing in mind that social data are inherently messy, and that one needs to accept a certain level of ambiguity that is inherent in the data, we used the following methods to isolate as much South Africa–only content as possible.

We started off by focusing on the metadata that is included with every tweet. We used the *user.location* field in each original tweet object to identify locations that fall outside of South Africa and removed all tweets containing these. Similarly, we isolated the South African time zones in the *user.time_zone* field and removed all the rest. These heuristics got us far but only about half of all tweets include data in these fields; the rest were blank.

Our next step was to remove irrelevant languages using the *user.lang* and *tweet.lang* fields. Twitter uses its own machine learning algorithms to predict the language of a user and of a specific tweet. Its predictions are not perfect though, as is the case with all kinds of machine learning classification tasks. In a South African context, this meant that many indigenous languages such as isiXhosa and isiZulu were misclassified as other languages, which prevented us from simply removing every non-English tweet. Instead, we focused on certain languages that were clearly not likely to be misinterpreted such as Japanese and Russian.

After this, we began building an extensive list of exclusion keywords. If a tweet contained one of these words, we could be sure that it was not relevant to us. To do this, we looked at the most popular hashtags and removed tweets that included clearly irrelevant hashtags. Similarly, we looked at the most popular influencers and retweeted tweets. We extracted unambiguous exclusion terms from these usernames and tweet content, which we used to further reduce the data.

Our pen-ultimate step was to cluster all the data using *K*-means clustering to group tweets together on the basis of the words they contained. A human analyst then evaluated the largest of these clusters to identify irrelevant ones for removal.

Finally, we created an interaction network out of the reduced data set where users who interact with each other are linked together in a network. We then ran a community detection algorithm on this network to identify distinct communities of users. Using a mixture of automated and manual evaluation, we identified the main groups of people talking in a non-South African context and removed these communities from our data.

This is how we arrived at our final data set of just under a million tweets. Despite this extensive cleaning process, we do accept that there are still some irrelevant, non-South African tweets in our data; however, we also believe that the data set is about as clean as it is likely to get without significant additional investment in terms of time and resources, which falls outside the scope of this article. Be that as it may, the reader would do well to keep this in mind as they read the article.

Interaction networks

Once we had our final, cleaned data set, we created an "interaction network" out of it by extracting usernames and connecting users together who interacted with each other via retweets or @mentions (interaction networks are not a new idea; see, for example, Conover et al., 2011; Kwak, Lee, & Moon, 2010; Larsson & Moe, 2011). In other words, we created an overall network of all the users who were speaking to each other within the context of the elections.

We used the Louvain Modularity algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) to identify distinct communities within the conversation network. We chose this method because it is widely used within the literature and it is scalable on large data sets. The basic idea behind most community detection algorithms is that they look for regions of the graph where interconnections between nodes are relatively dense compared to other regions of the network (see Figure 7 for an

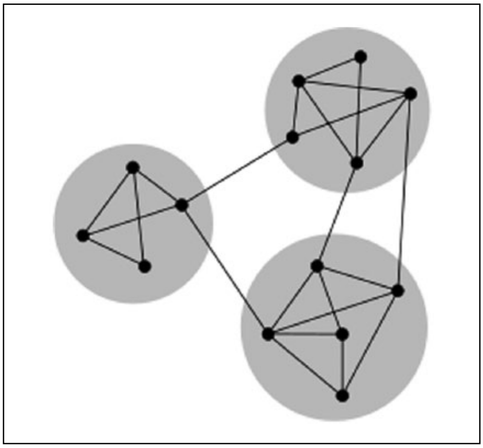

Figure 7. "A sketch of a small network displaying community structure, with three groups of nodes with dense internal connections and sparser connections between groups" (Wikipedia, 2015).

illustrated example). In the context of our data, areas of relative density occur when groups of people interact with each other more than with other groups. We essentially take advantage of the homophily effect to identify people who are talking together inside a "virtual room," bonded by their common interest or agenda. For example, you are more likely to speak to friends that you share a political ideology with and all those who share the ideology are more likely to speak to each other as well, creating an area of the graph with relatively dense interconnections. Community detection algorithms identify these regions of the graph. One could think of such an approach as an organic, bottom-up, data-led approach to segmenting a data set based on how people sort themselves. The Louvain Modularity algorithm in particular only allows a user to be placed into a single community. It does not detect overlapping community memberships.

Using the community detection algorithm, we were able to identify distinct constituencies within the elections conversations, most of which centered on specific political parties, although some parties had more than one community focused on it.

Demographics

We made use of a third party provider, Demographics Pro, to evaluate the demographics of our various constituencies. Demographics Pro maintains a large database of Twitter users against which it applies proprietary machine learning classification models in order to identify a variety of demographic information for those users. While their results are largely a black box, they do provide clear distinctions that have face validity. Demographics Pro describes their methodology as follows:

> ***About our methodology***
>
> *Demographics Pro relies on our core ability to estimate or infer the likely demographic characteristics (gender, age, marital status, ethnicity, occupation, location and etc.) of local consumers based purely on their social media presence/usage.*
>
> ***How we infer demographics***
>
> *Our methodology is data-centric, relying on multiple data signals from three primary areas: networks, consumption and language. Data signals are filtered and amplified using large, proprietary knowledge*

bases of established correlations between data points and demographic characteristics. Finally, we combine multiple amplified signals using a series of algorithms to estimate or infer likely demographic characteristics.

Coverage and accuracy

We have used our methodology to infer demographics for some 300 million local consumers to date. In terms of accuracy, we typically require confidence of 95% or above to make an estimate of a single demographic characteristic for any given local consumer. Our success here relies on the relatively low covariance of multiple amplified signals: iterative evaluation using established samples allows us to calibrate the balance between depth of coverage (i.e. the number of demographic estimates we make) and our required accuracy.

Case study results

Breakdown of communities

After cleaning the data and applying the methodologies described above, we were able to create the interaction network for the 2014 South African elections (see Figure 8). We use node size to quantify how influential a user was based on the number of interactions that they had with other users (i.e., how often they caused action in others in the form of @mentions and retweets). A node's color is based on the community that it belongs to. Regions of consistent color represent a single community. Thus, at a glance, we can see the most influential accounts and the various communities that make up the conversation.

We are now in a position to profile the top communities, or constituencies, that make up the overall conversation around the 2014 South African elections. By evaluating the influencers and the content that most resonated (i.e., was most retweeted) within each community, we were able to qualitatively describe the communities with the most members as seen in Figure 9.

The top 10 communities collectively represented 46.7% of all unique users in the data but generated 81.5% of all tweets about the elections. This breakdown clearly shows that communities tended to coalesce around specific political parties (including nascent party, Agang, lead by academic and businesswoman, Mamphela Ramphele, which does not feature prominently in the rest of the article), although there are exceptions such as a community centered around comedians, Chester Missing (*@chestermissing*) and Trevor Noah (*@trevornoah*), and around specific news outlets such as eNCA, Eyewitness News (EWN), News24, and Times Live.

It is interesting to compare the number of unique members that each community has versus the proportion of tweets that was generated by each community's members. The top four communities were the most vocal, accounting for 71.3% of all tweets generated. In the interest of space, these four communities will be the focus of the rest of the article. The remaining top communities quickly drop off in terms of size and will not be unpacked further here.

Based on the size of the top two communities and the fact that they collectively generated almost half of all of the gathered tweets (42.3%), we can in some sense say that the 2014 elections conversation on Twitter was dominated by the DA and EFF. This is particularly impressive on the part of the EFF which had only been in existence for several months at this point. This could perhaps be in indicator of the party's future momentum.

The ratio of users to tweets is interesting in and of itself, but an immediate question that arises is what type of conversations were happening within these communities? Were the communities dominated by one-way broadcasting via retweets of popular articles or quotes, or, were they dominated by more organic tweets by users in the form of either "isolated tweets" where users said something relating to the elections in an unprompted fashion without mentioning any other Twitter users, or "@mentions," which refer to user-generated tweets that explicitly tags or "mentions" another account(s), and are potentially indicative of two-way conversations?

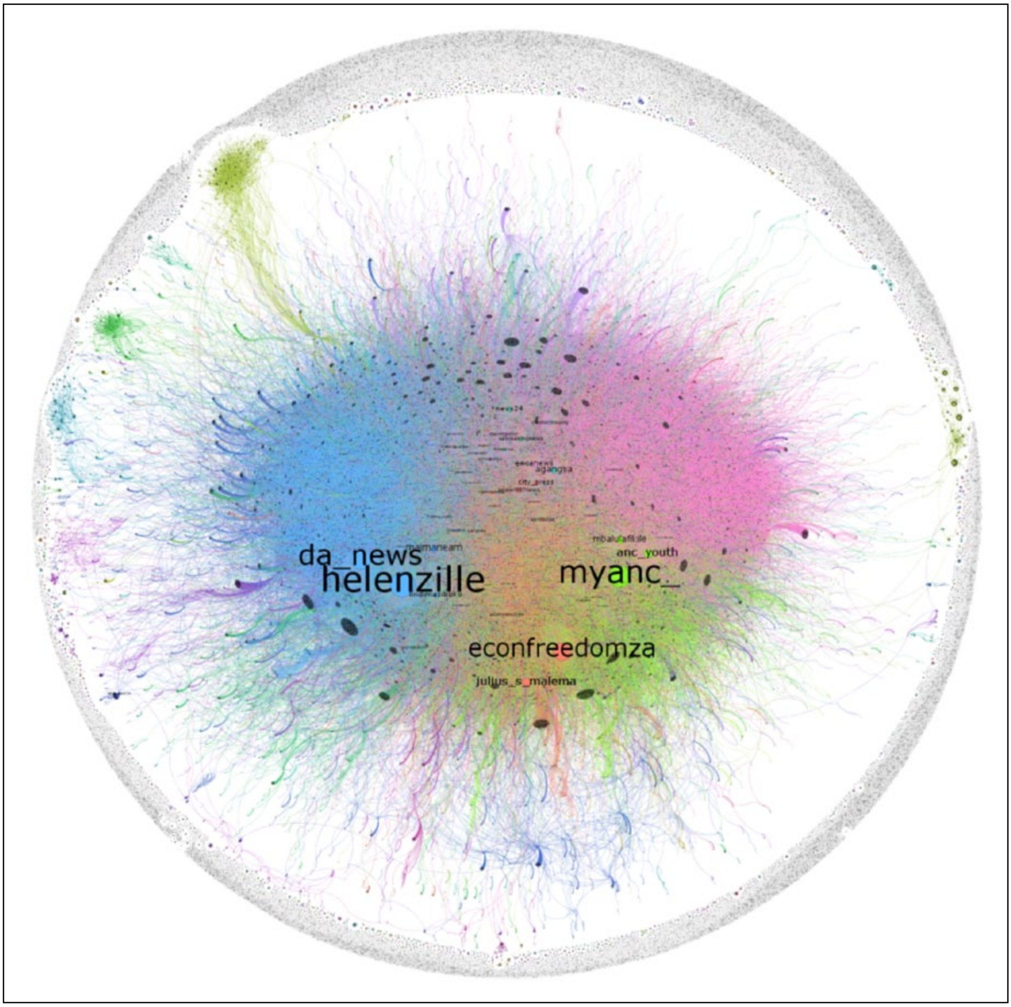

Figure 8. The overall interaction network map exhibits many dense connections between nodes (users) pointing toward the density of conversations between users across the board even though several clear communities are apparent in the data. User node color represents community membership. Colored regions of the map represent groups of nodes clustered into a distinct community.

Figure 10 breaks down the types of tweets generated by each of the top 10 communities. Most tweets generated by each community were one-directional retweets. This is unsurprising as the nature of the Twitter platform lends itself to information sharing. The proportion of retweets is particularly high in the Comedians community as their followers share their humorous tweets with their own followers.

The top four communities hovered around the 20% mark for the proportion of isolated tweets generated. A high number of isolated tweets is a good sign of a healthy community with engaged members. The DA community had the lowest proportion of isolated tweets (14%), but this was offset by a high proportion of @mentions (33%), perhaps implying more two-way conversations than within the other top communities. The Traditional ANC community appears to have generated the most organic conversations—it had the lowest proportion of retweets among the top four communities (46%) and the highest proportion of @mentions (36%). This might point toward more healthy,

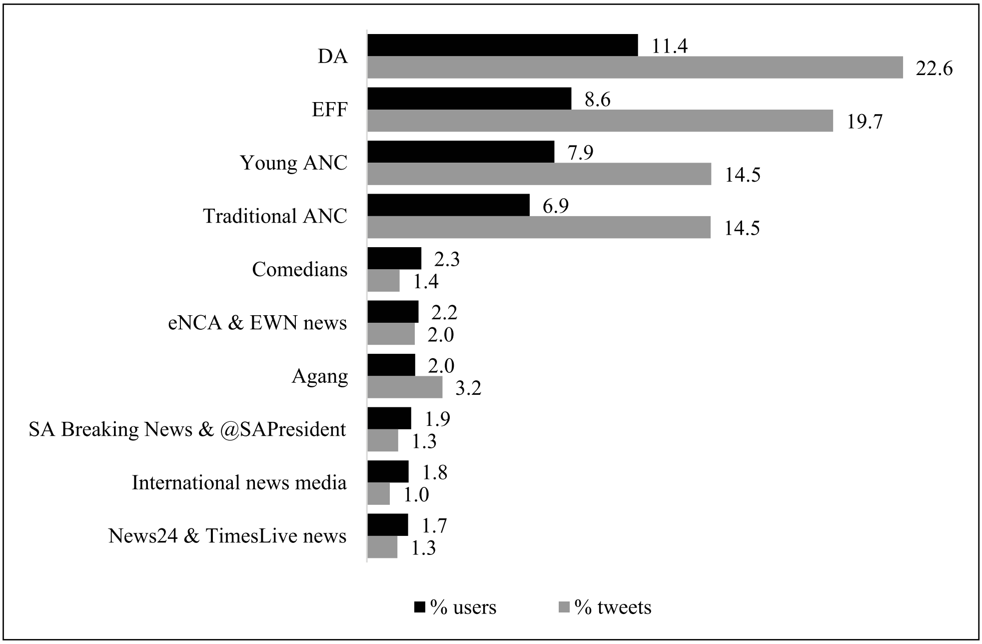

Figure 9. Breakdown of top communities in terms of (1) proportion of unique users and (2) proportion of tweets generated by each community.

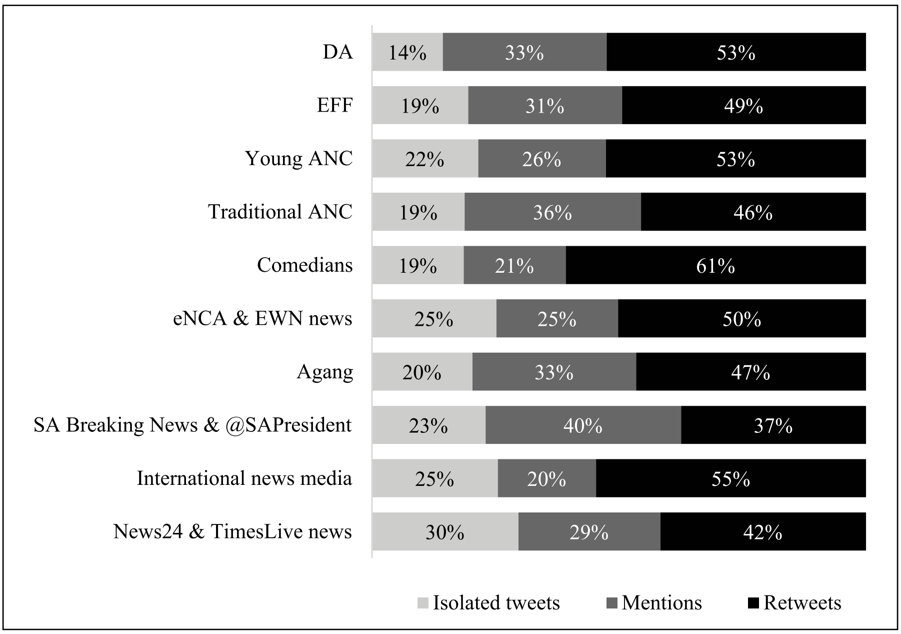

Figure 10. A breakdown of the types of tweets generated by each community.

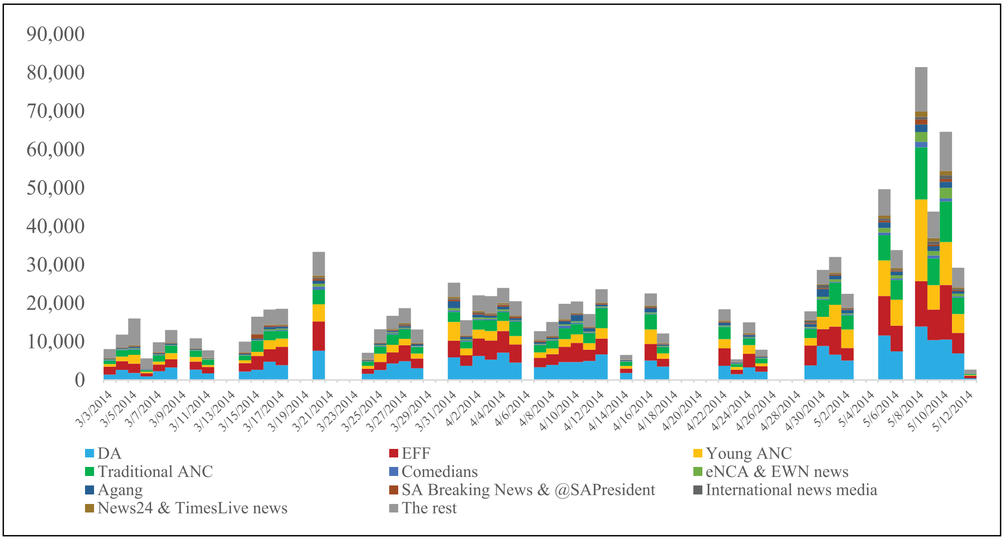

Figure 11. A breakdown of daily volumes across the top 10 communities in the data.

robust conversations. Much of the coverage of President Jacob Zuma around the elections was critical of the president (including the Nkandla scandal where Zuma spent over R200 million—over US$ 12 million in 2016 terms—on upgrading his private residence, much of which was public money). Could the Traditional ANC community members have had more to talk about among themselves than their DA and EFF counterparts in rationalizing their support for the party and the man?

Finally, we look at when each of the top communities in our data was active (see Figure 11). Barring the days that we are missing data for, we see that most of the top communities were active for the entire duration of the data (3 March to 12 May) implying that conversations persisted on social media in the months leading up to the elections. There were few notable spikes from any one community with chatter generally increasing across the board in the days after the election.

Influencers

In describing each community, we first start by looking at the key influencers that "anchor" each community. These were the accounts that had the largest impacts on the conversations within each community as they were retweeted and @mentioned the most.

The DA community coalesced around official party accounts such as the then-leader Helen Zille (*@helenzille*), the official party account (*@da_news*), current leader Musi Mmaimane (*@maimaneam*), and the then-speaker in Parliament Lindiwe Mazibuko (*@lindimazibuko*).

Key influencers within the EFF community were again official accounts, including the party account (*@econfreedomza*), leader Julius Malema (*@julius_s_malema*), and Sentletse Diakanyo (*@sentletse*), the "Self-appointed EFF Supreme Commander of the Twitter battalion."

The Young ANC community was an interesting community within our data. It was somewhat more decentralized than our other top communities indicative of a more broad-based engagement between community members rather than conversations centralized around a few key accounts. However, certain middle-class, black influencers—arguably the emergent face of the new South Africa—were still evident, including author and media commentator, Khaya Dlanga (*@khayadlanga*), young ANC spokesperson Mayihlome Tshwete (*@mtshwete*), and Twitter celebrity, TaxiDriverSipho (*@taxidriversipho*). These individuals have large followings among the youth and are tapped into youth culture.

Table 2. Top 10 most retweeted tweets in the DA community.

Author	Tweet content	Number of retweets
@helenzille	By saying that "only clever people" have problems with R246-mill Nkandla upgrade, Pres Zuma is implying that ANC voters are [stupid]	268
@TannieEvita	Confusius say: "He who knows nothing about his own house, knows even less about his own country." @SAPresident	235
@alexeliseev	Zuma on #Nkandla: "It's not an issue with voters. It's an issue with bright people. Very clever people." What is he saying . . .	207
@helenzille	Monitor the polls!! "@E_van_Zyl_17: "@JohnBiskado: IEC voting material found at house of anc party agent ward 77 http://t. . . .	206
@MaxduPreez	Defend this, ANC: Public Works Dept redirected service delivery funds to pay for Nkandla in contravention of constitution	176
@helenzille	Now that court has found it fair to say Zuma stole your money, when will HE appear in court? oh, and don't forget 763 count . . .	166
@helenzille	Shameful. Another abuse of power. RT @JacquesMaree73: Living with Zumability. #Zuma http://t.co/JZdILPK8fS	162
@helenzille	ANC campaign showing solidarity with the poor. "@adam_ McKendi: #Election2014 #SA #Zuma @Helenzille http://t.co/ Bld1pp8vSY"	161
@JacaNews	Maimane says most of the ballots found were votes for the DA #Elections2014 http://t.co/JWMEB9ktMw	155
@helenzille	ROFL. The ANC just sent me an sms saying that to stop corruption, I should vote ANC. A line straight from the Wolf in Red R . . .	154

The fourth largest community, and the last one that we look at in detail in this article, clearly made up of the ANC's traditional support base (Traditional ANC). It espoused the party line and its members readily consumed content from official party accounts @myanc_ and @anc_youth, and that of flamboyant then-Minister of Sports, Fikile Mbalula. Just by looking at the top influencers, we can already see that the communities are clearly drawn along party lines.

Resonant content

Twitter allows users to propagate tweets verbatim on to their own followers using the "retweet" function. We can get a clear idea of the kind of content that resonates with the personal feelings and agendas of users by looking at the content that is retweeted the most within a community. This gives us further insight into the type of sentiment and issues that resonate within a community.

Table 2 summarizes the top 10 most retweeted tweets within the DA community. A disproportionate number of the top retweets relate to President Jacob Zuma and his Nkandla corruption scandal. Overall, the conversations in this community do not present us with any major surprises. Many comments were pro-DA and against the ANC, although some were also derisory of the EFF. The Nkandla scandal in particular resonated incredibly strongly as did the banning of the DA's "Ayisafani" television advert, which openly criticized the ANC government. Again, by looking at the top retweeted content, we can see that this community has coalesced around the DA and its party politics.

Much of the top retweeted content that resonated within the EFF community related to alleged election rigging through lost and tampered with votes (see Table 3), perhaps pointing toward EFF supporters' particular anxiety as the underdog in the face of larger parties that might bring their weight to bear on election results. Overall, this community centered around the official EFF

Table 3. Top 10 most retweeted tweets in the EFF community.

Author	Tweet content	Number of retweets
@[]	EFF votes found dumped near Diepsloot cc @ Julius_S_Malema @Sentletse @EconFreedomZA with @ IECSouthAfrica stamp http://t.co/mIzvrow . . .	201
@Julius_S_Malema	We will soon announce the date of the march to the union buildings to demand Zuma's resignation as the president	187
@[]	@IECSouthAfrica apparently found in home of ANC party agent. https://t.co/OyoYIvI79Y	177
@Sentletse	The ANC and the IEC must be ashamed of themselves! http://t.co/STwypvLg8M	172
@[]	I.E.C voting material found in a house of an anc party agent in ward 77 @EconFreedomZA @EconFreedom_GP #VoteEFF http://t.c . . .	138
@POWER987News	More #IEC material and Ballot boxes found, allegedly, at the house of an #ANC party agent http://t.co/j8WTNk9QLg	126
@[]	Official Ballot Boxes found in a house in Alex full off marked ballot papers for DA and Eff.@EconFreedomZA http://t.co/ YWg . . .	125
@justicemalala	Mbeki's #ANC in 99: 66.35% Mbeki's ANC 2004: 69.6% Zuma's ANC 2009: 65.9% Zuma's ANC 2014: 62.84 (22.12/08 May) #justsay . . .	121
@POWER987News	Police officer inspects the #IEC material & Ballot Boxes found allegedly at the house of an #ANC party agent http://t.co/ . . .	112
@[]	ANC teasing us with flashy cars . . . a life which is mostly unrealistic for most of us . . . SMH! http://t.co/ ymHDPZaKo7	105

Some usernames have been excluded for privacy reasons. We have only included usernames of well-known public figures and entities.

mouthpieces; however, some ambiguity between the ANC and EFF is evident, perhaps as supporters struggled to reconcile the appeal of the EFF with their traditional allegiances. The DA seldom came up, perhaps implying that the EFF was primarily drawing its supporters from the ANC fold.

The top retweeted content in the Young ANC community (see Table 4) favored the ANC; however, there was content sympathetic to the DA as well (e.g., congratulating the DA's Lindiwe Mazibuko for being accepted at Harvard University). It is interesting to see the relative balance with which they regarded other parties as opposed to the clear partisanship that tends to dominate so much of South African political discourse. While they firmly sat in the ANC camp, they also appeared to make a clear distinction between the party and what it stands for, and President Jacob Zuma and the hundreds of charges of corruption against him (see, for example, the tweets by @justicemalala and @ThomasGumede in Table 4).

Sports Minister, Fikile Mbalula, dominated the top retweeted content within the Traditional ANC community (see Table 5). Overall, the most resonant content was unequivocally pro-ANC and pro-President Zuma.

Thus, by looking at the top influencers within each community and at the content that they retweeted the most, we get a clear insight into who leads the conversations within each community and the extent to which their agendas resonate within the community.

Table 4. Top 10 most retweeted tweets in the Young ANC community.

Author	Tweet content	Number of retweets
@MbalulaFikile	But you cant use the same BIS bought with NSFAS to tweet "ANC HAS DONE NOTHING." Uxokelani?	119
@justicemalala	Mbeki's #ANC in 99: 66.35% Mbeki's ANC 2004: 69.6% Zuma's ANC 2009: 65.9% Zuma's ANC 2014: 62.84 (22.12/08 May) #justsay . . .	110
@__Senz	Congratulations to Lindiwe Mazibuko for being accepted to Harvard University. Great role model for young South African women. G . . .	110
@ProudlySA	Congratulations to Deputy President Motlanthe & his beautiful bride, Gugu on their wedding today!! @ PresidencyZA http://t.co . . .	104
@MbalulaFikile	Biggest loser of the century DR Mamphela Ramphela,u Luzile shameeee	102
@MbalulaFikile	Yet you're on Twitter and can write that in perfect English, let's face it, you're ANC's good story. @ lusylooya: ANC ha . . .	95
@MbalulaFikile	Having Black colleagues under Apartheid never gave anyone struggle credentials. It's the fight that did, you don't featu . . .	80
@ThomasGumede	Zuma's thinking "I must REALLY suck, even Bafana aint getting booed!"	75
@IECSouthAfrica	National Assembly seats: APC–1; PAC–1; AGANG SA–2; ACDP–3; AIC–3; COPE –3; UDM–4; VF Plus–4; NFP–6; IFP–10; EFF–25; DA– . . .	74
@khayadlanga	The ANC lives. The ANC leads. Long live the ANC.	67

Media partisanship

Finally, we look at the various news outlets that resonated the most with each community. This analysis does not imply that these outlets consciously shape their content in order to appeal to a particular partisan group but it does imply that their content clearly resonates more with one specific community than with others since, in order for a news outlet to find itself included within a particular community, it must have been retweeted and @mentioned more by members of that community than any other.

It is important to note that many news outlets did not fall into our top four party-related communities, instead forming their own independent communities. For example, we have already seen that news outlets eNCA and EWN formed a distinct community, as did News24 and Times Live.

Here then are some of the news outlets that found themselves included within party-specific communities as their content clearly resonated more with that community than any other:

The news outlets that most resonated within the DA community were Jacaranda News (*@jacanews*), Business Day (*@bdlivesa*), and The Daily Maverick *(@dailymaverick*). These outlets have traditionally white audiences, espouse free market capitalism, and/or come from a liberal political tradition, all of which are elements of the DA ideology and heritage.

News outlets in the EFF community included City Press (*@city_press*), POWER987 News (*@power987news*), SABC News (*@sabcnewsonline*), SAfm News (*@safmnews*), and Radio 702 (@

Table 5. Top 10 most retweeted tweets within the Traditional ANC community.

Author	Tweet content	Number of retweets
@MbalulaFikile	I VOTE ANC: Retweet if you are.	139
@MbalulaFikile	But you cant use the same BIS bought with NSFAS to tweet "ANC HAS DONE NOTHING." Uxokelani?	111
@MbalulaFikile	Biggest loser of the century DR Mamphela Ramphela,u Luzile shameeee	107
@MbalulaFikile	Yet you're on Twitter and can write that in perfect English, let's face it, you're ANC's good story. @lusylooya: ANC ha . . .	101
@mgigaba	Kwangathi sesikhuluma sodwa nje la kwa twitter ma ANC! Baphi kanti abanye abebenemilomo emide, bekhulumela safuthi?	95
@ANCYLhq	Retweet to wish comrade president Jacob #Zuma a happy #birthday!	74
@[]	I haven't heard 1 person blame Zuma for the ANC's victory, but they blame him for everything else!	73
@MbalulaFikile	Having Black colleagues under Apartheid never gave anyone struggle credentials. It's the fight that did, you don't featu . . .	70
@mgigaba	The ANC now reaches a 7m mark #ThePeopleHaveSpoken	65
@ProudlySA	Congratulations to Deputy President Motlanthe & his beautiful bride, Gugu on their wedding today!! @ PresidencyZA http://t.co . . .	63

radio702). Some of these tend toward a more left-leaning ideology, while others, such as state (and thus ANC-controlled) news outlets, SABC News, and SAfm News are more surprising. Their inclusion might be due to adversarial interactions where community members disagreed with their content.

No specific news outlets were included in the top influencers for the Young ANC community, perhaps indicating that this is a latent group that was not catered for by the mainstream media. Also, given their more balanced approach to other parties, they might simply be consuming content from a wider variety of news outlets than the other communities.

The news outlets in the Traditional ANC community are particularly intriguing as they represent our first glimpse of overt attempts at state propaganda. It is interesting to note the appearance of the ANN7 news channel (@ann7v) and The New Age newspaper (@the_new_age) accounts among the top ten influencers for this community, as ANN7 and The New Age have come under heavy criticism for being funded by the Gupta family who have close ties with the ANC and President Jacob Zuma. Indeed, this relationship has formed one of the core scandals at the heart of the Zuma presidency. These data would seem to imply that the content released by these news outlets particularly resonates with this community, perhaps lending some credence to the suggestion of media partisanship; however, community membership alone is not enough evidence to draw a conclusive judgment given that interactions can be supportive or antagonistic and community detection is non-deterministic (although in our other runs of the data, these accounts also fell within the ANC community). Bearing these caveats in mind, Smit, Findlay, and Janse van Rensburg (2016) identified a similar dynamic in the global discussion around immigration politics, where conservative, partisan news outlet, Breitbart News, was firmly embedded within the US conservative,

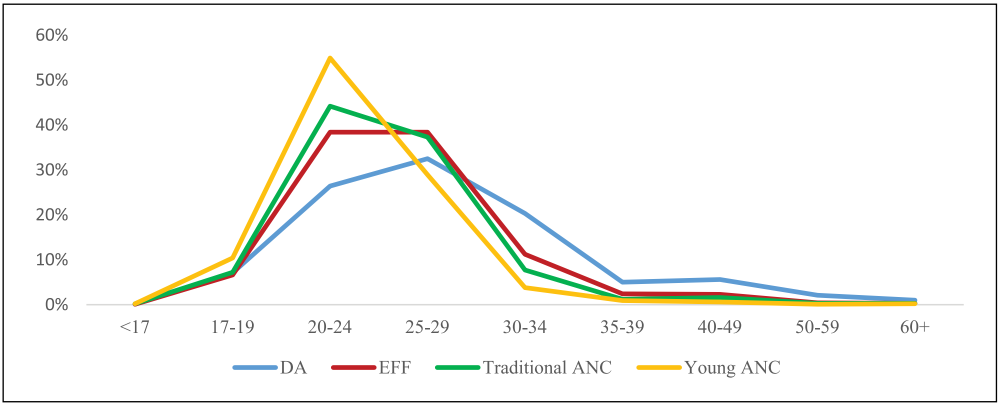

Figure 12. Distribution of community members' ages according to Demographics Pro.

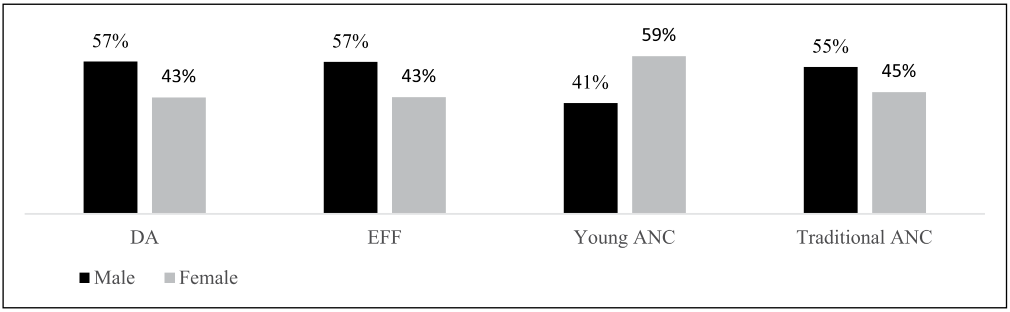

Figure 13. The gender distribution across our top four communities according to Demographics Pro.

Republican community and its sister account, Breitbart London, within the conservative UKIP community in the United Kingdom.

Demographics

We used Demographics Pro to infer the demographics of the four main communities in our data (see the *Methodology* section for a description of how they infer this information). Figure 12 shows that the split within the ANC was indeed between younger voters who we have already seen were more critical of President Zuma and older voters who continued to support him. Conversely, the DA appears to attract slightly older voters.

We also find that younger ANC voters critical of President Zuma were more likely to be women (see Figure 13).

These findings align with the broader social demographics at play in South Africa as described in the *Introduction* section. Namely, younger voters, or so-called "born frees," who did not grow up during apartheid, put less importance on the narrative of the ANC as a liberation organization. Instead, they are more interested in access to the economy and a better standard of living. It is possibly for this reason that they are critical of President Jacob Zuma. They likely consider the more than 700 corruption charges leveled against him as demonstrative of the key role he has played in the worsening South African economy, which in turn limits their access to a better life.

Future research

This paper has focused on the decomposition of network structure as the primary mode of analysis. These analyses could be further fleshed out through more comprehensive text mining than what was conducted in this paper for basic illustrative purposes.

Also beyond the scope of this current paper is investigations into electoral predictions. To what extent could this data have accurately predicted the election outcomes?

Finally, while Twitter is a very important form of public data, it is not necessarily representative of the general voting public. As such, it is interesting to consider how the data could have been expanded to include i) other social media platforms, such as Facebook, and ii) to include more data modalities beyond just text, such as image and videos, in the analyses.

Conclusion

The intent of this article is to demonstrate how interaction networks can be used to identify and describe the various constituencies that make up specific discussions occurring on Twitter. We have specifically focused on political discussions in this article but the approach can be applied to any defined conversation on Twitter, or similar public platforms.

Our approach relies heavily on the info gleaned from the structural features of interaction networks to quickly profile constituencies. We connect users together when they interact with each other (in the case of Twitter, via retweeting and @mentioning each other). This gives us a network of who is speaking to whom. We are then able to identify the subcommunities that exist within this conversation. By filtering our data on these communities, we can profile each in terms of their key influencers, the type of content that resonates with them, their demographics, and other useful information. Even deeper content analysis is possible than what we have presented in this article, but such investments in time and resources were beyond the scope of this article. Regardless, using interaction networks allows researchers and marketers to quickly isolate and segment their core constituencies in terms of their priorities, agendas, and vectors of influence using a bottom-up, data-led approach.

In the specific case of the South African 2014 national elections, we found that there were main four constituencies split along party lines that dominated the political discourse on Twitter. By unpacking them in more detail, we saw that President Jacob Zuma has forced a generational divide within ANC supporters due to his allegedly corrupt activities. We saw that the scandals surrounding President Zuma provided the main fodder for discussion among the other communities as well.

Finally, we briefly touched on the issue of media partisanship by identifying the media outlets whose content specifically resonates with one constituency more than others. This is an area ripe for further research.

While none of the individual methodological elements presented in this article are necessarily novel in execution, we believe that the applied ensemble approach presented in this article allows researchers and marketers to make sense of the seemingly messy and chaotic world of social media and gives us unprecedented insight to human behavior at scale and look forward to seeing how others systematically extend it.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Bermingham, A., & Smeaton, A. F. (2011, November 13). *On using Twitter to monitor political sentiment and predict election results*. Sentiment Analysis where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP), Chiang Mai, Thailand.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *10*, Article P10008. doi:10.1088/1742-5468/2008/10/P10008

Chung, J., & Mustafaraj, E. (2011, August 7–11). *Can collective sentiment expressed on Twitter predict political elections?* Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, CA.

Conover, M., Ratkiewicz, J., Francisco, J. M., Gonçalves, B., Flammini, A., & Menczer, F. (2011.). *Political polarization on Twitter*. Fifth International AAAI Conference on Weblogs and Social Media. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2847/3275

Findlay, K., & Ooserveld, P. (2012). *Adventures in big data land: Mapping the South African political Twitter landscape*. South African Market Research Association's 2012 Annual Conference. 16–17 August 2012, Rosebank Crowne Plaza Hote, Johannesburg, South Africa.

Gayo-Avello, D. (2012). *"I wanted to predict elections with Twitter and all I got was this lousy paper": A balanced survey on election prediction using Twitter data*. Retrieved from http://arxiv.org/abs/1204.6441

Gayo-Avello, D., Metaxas, P. T., & Mustafaraj, E. (2011). Limits of electoral predictions using Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 490–493). New York, NY: AAAI. https://www.aaai.org/Library/ICWSM/icwsm11contents.php

General Chair: Nicolas Nicolov and James G. Shanahan Program Cochairs: Lada Adamic, Ricardo Baeza-Yates, Scott Counts.

Goldstuck, A., & Wronksi, M. (2014). *SA social media landscape 2015 executive summary* (Fuseware report). Retrieved from http://www.worldwideworx.com/wp-content/uploads/2016/09/Social-Media-2017-Executive-Summary.pdf

Kwak, H., Lee, C., & Moon, S. (2010, April 26–30). *What is Twitter, a social network or a news media*. Proceedings of the 19th World-Wide Web (WWW) Conference, Raleigh, NC.

Larsson, A. O., & Moe, H. (2011). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society*, *14*, 729–747.

Media Monitoring Africa. (2014). *Media monitoring Africa election coverage 2014*. Retrieved from http://elections2014.mediamonitoringafrica.org

Mitchell, A., & Hitlin, P. (2013, March 4). Twitter reaction to events often at odds with overall public opinion. *Pew Research Center*. Retrieved from http://www.pewresearch.org/2013/03/04/twitter-reaction-to-events-often-at-odds-with-overall-public-opinion

Sang, E. T. K., & Bos, J. (2012, April 23). *Predicting the 2011 Dutch senate election results with Twitter*. Proceedings of the Workshop on Semantic Analysis in Social Media, Avignon, France.

Smit, C., Findlay, K., & Janse van Rensburg, O. (2016, April). *#borerlinewelcome: 'migration' on Twitter*. Paper presented at the 70th Market Research Society Impact Conference, London, England.

TNS. (2014). *Connected Life 2014*. Retrieved from http://www.tnsglobal.com/about-connected-life-2014

TNS Sunday Times. (2014). Top brands survey 2014 Sunday Times supplement, August 2014. URL: http://blackstaradroom.co.za/wp-content/uploads/2016/08/SundayTimesTopBrandsAugust2014.pdf. Accessed 9 January 2018.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*, *29*, 402–418.

Wikipedia. (2015, April 8). Community structure (The Free Encyclopedia). Retrieved from http://en.wikipedia.org/w/index.php?title=Community_structure&oldid=655476118