Int. J. Internet Marketing and Advertising, Vol. X, No. Y, XXXX

Competitive analysis of social media data in the banking industry

Ibukun Tolulope Afolabi*,
Azubuike Ansalem Ezenwoke
and Charles K. Ayo

Department of Computer and Information Sciences,
Covenant University,
Ota, Nigeria
Email: ibukun.fatudimu@covenantuniverity.edu.ng
Email: azu.ezenwoke@covenantuniverity.edu.ng
Email: charles.ayo@covenantuniverity.edu.ng
*Corresponding author

Abstract: Recently, most companies interact more with their customers through the social media, particularly Facebook and Twitter. This has made large amount of textual data freely available on the internet for competitive intelligence analysis, which is helping reposition more and more companies for better profit. In order to carry out competitive intelligence, financial institutions need to take note of and analyse their competitor's social media sites. This paper, therefore, aims to help the banking industry in Nigeria understand how to perform a social media competitive analysis and transform social media data into knowledge, which will form the foundation for decision-making and internet marketing of such institutions. The study describes an in-depth case study which applies text mining to analyse unstructured text content on Facebook and Twitter sites of the five largest and leading financial institutions (banks) in Nigeria: Zenith Bank, First Bank, United Bank for Africa, Access Bank and GTBank. Analysing the social media content of these institutions will increase their competitive advantage and also lead to more profit for the banking institutions in question. The results obtained from this research showed that text mining is able to reveal uncommon and non-trivial trend for competitive advantage from social media data, and also provide specific recommendations to help banks maximise their competitive edge.

Keywords: social media; Twitter; Facebook; text mining; banking; competitive intelligence; clustering; sentiment analysis.

Reference to this paper should be made as follows: Afolabi, I.T., Ezenwoke, A.A. and Ayo, C.K. (XXXX) 'Competitive analysis of social media data in the banking industry', *Int. J. Internet Marketing and Advertising*, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes: Ibukun Tolulope Afolabi is a Lecturer I in the Department of Computer and Information Sciences, Covenant University, Ota, Nigeria. She holds a BSc in Engineering Physics from Obafemi Awolowo University, Ile-Ife, MSc in Computer Science from the University of Ibadan, Ibadan, Nigeria, and a PhD degree in Computer Science from Covenant University, Ota, Nigeria. Her research interest is in the field of data mining with special focus on text mining. She enjoys reading, teaching and engages in creative arts. She is a Member of the Nigerian Computer Society and Computer Professional Registration Council of Nigeria.

Azubuike Ansalem Ezenwoke is an Assistant Lecturer in the Department of Computer & Information Sciences. His research interest includes software engineering, service oriented computing, software product line engineering, mobile computing, technology assisted education, learning and career guidance. He is a recipient of the 2012 Google Computer Science for High School Awards. He has collaborations/linkages with the Center of Excellence for Mobile e-Services, Department of Computer Science, University of Zululand, Kwa-Zulu Natal, Republic of South Africa.

Charles K. Ayo is currently the Vice Chancellor of Covenant University, Ota, Ogun state, Nigeria. He holds a BSc, MSc and PhD in Computer Science. His research interests include mobile computing, internet programming, e-business and government, and object-oriented design and development. He is a Member of the Nigerian Computer Society and Computer Professional Registration Council of Nigeria. He is also a Member of a number of international research bodies such as the Centre for Business Information, Organization and Process Management, University of Westminster (http://www.wmin.ac.uk/wbs/page-744) and the Editorial Board of the *Journal of Information and Communication Technology for Human Development*.

1 Introduction

Social media is a term used to refer to a wide range of internet-based and mobile services that allow users to participate in online exchanges, contribute user-centred content or join communities. Social media encompasses blogs, wikis, social bookmarking, virtual world content, media sharing sites, social network sites and status update services (Dewing, 2012). Online social network facilitate connection between people to communicate with individuals who are in their network using the web as their interface (Barnes and Ganim, 2011). Social network sites are web-based services that allow individuals to own a semi-public profile, share connections with other users, view and explore their connections and those of others and express and share personal opinions about any topic or subject of interest. Prominent examples of social networking sites are Facebook and Twitter (Dewing, 2012). Facebook, founded by Mark Zuckerberg, with his college and roommates in 2004 (Farooq et al., 2012), was initially developed for college and university students, but is now made available to anyone. People may register under various networks, such as school, place of employment, geographic region, etc. (Yang et al., 2007). Status update on the other hand allows people to share short updates about people or events and to view updates created by others. An example of status update service is Twitter (Fox et al., 2009). Twitter is a fast growing micro blogging service with over 320 million users as of January 2016. Twitter users tweet about any topic within the 140-character limit and follow others to receive their tweets (Kwak et al., 2010). The diffusion and ubiquitous use of social media in recent times has enormously increased the volume of User-Generated Contents (UGC) such as text, videos and photos, available on the web (Choudhury and Alani, 2014). Twitter alone boasts more than 500 million tweets generated per day; Facebook encourages textual content generation in the form of wall posts, comments, notes, etc., and millions of photos are uploaded to Facebook daily.

The growth and popularity of social networking trends has inspired social media to emerge as platforms for promoting business values. For example, many organisations have adopted social media to increase customer base, disseminate information about new products and/or services, increase sales, manage customer relationship, promotional advertising, build brand reputations and so on (He et al., 2013). Since social media provide platforms for consumers to express and share feelings and thoughts with a wider audience, the presence of organisations on social media gives these organisations access to useful information that can help them to provide better service delivery, connect with prospective customer/client, identify gap in the market for a target product/service or discover competitive marketing advantage through Competitive Intelligence (CI). In the light of this, the application of knowledge discovery techniques on social media sites is useful in revealing significant information embedded in interaction behaviours and in observing general opinions and inherent patterns in human thinking or feelings, in relation to any topic of interest (Aggarwal and Wang, 2011).

CI is a process in which an organisation uses publicly available resources to generate data on the business environment, and particularly the competitors' strategies. These data are then transformed into useful information to support business decisions, in anticipation of competitors' activities and for the organisation to understand their position relative to the competition (Agarwal, 2006; McGonagle and Vella, 2002; Nasri, 2011). In other words, CI is the analytical process that transforms scattered information about competitors and customers into relevant, accurate and usable strategic knowledge on market evolution, business opportunities and threats (Albescu et al., 2009). Therefore, organisations can, through CI, detect and/or anticipate changes in the business environment, such as emerging technological trends, consumer preferences, opportunities for new market and changes in demographics (Gray, 2010).

Gémar and Jiménez-Quintero (2015) itemised seven benefits of CI to an organisation; they include the following: (1) gain a better understanding of their business environment and industry; (2) learn about corporate and business strategies of competitors; (3) forecast opportunities and threats; (4) anticipate the research and development of competitors' strategies; (5) validate or deny industry rumours; (6) take effective decisions; and (7) act instead of reacting. Until now, the data required for CI are often limited due to insufficient data about the competitors. However, the proliferation of social media use leaves large volume of information about competitors, becoming a new platform to source CI (Gémar and Jiménez-Quintero, 2015). In their work, McGonagle and Vella (2002) posited that about 90% of the information an organisation requires to mine CI is already public while also noting that the information required for most CI projects is already available on the web (Bărbulescu et al., 2007).

It is now possible that contents about an organisation on social media can enormously affect an organisation's reputation, impacting on sales and the overall survival of that organisation (Gémar and Jiménez-Quintero, 2015). Therefore, an effective marketing strategy must take into cognisance knowledge about the customer and the competition (Jaworski et al., 2002), making it imperative to analyse social media data for CI. This is particularly true bearing in mind that the social dimension has become an important consideration in most decisions made by business executives. On the basis of this, there are increased research interests towards profitable use of social media contents from sites such as Twitter, Facebook, YouTube or LinkedIn (Gémar and Jiménez-Quintero, 2015) for competitive analysis.

The ability to analyse the richness of data from social media platforms is vital for CI to maximise the vastness of social media. Data analysis for CI is the most challenging part of the intelligence cycle. This is true because it involves detecting interesting patterns hidden in the data and developing different scenarios based on what the analyst has discovered (Kahaner, 1996). To perform data analysis, a variety of analytical models can be employed. Some examples include PEST (Political/legal, Economic, Socio-cultural and Technological analysis), Porter's five forces model, SWOT (Strength, Weakness, Opportunity and Threats) analysis and competitor profiling. These analytical models can convert disparate pieces of information into actionable intelligence (Nasri, 2012). According to Nasri (2012), the ability to perform appropriate analysis and interpretation is vital for the success of the process of CI. The fact that social media data come mostly in the form of text generates interest in application of approaches like text mining. Noting the recent trends of engaging social media platforms for customer engagement and its potential for CI, this paper explored the social media CI analysis by employing both statistical analysis and text mining techniques, using the banking industry in Nigeria as a case study.

2 Literature review

Apart from photos and videos, text is a major form of content on most social media sites, and the use of unstructured and semi-structured language for textual expressions is prevalent. This is not surprising, as most communications in everyday life involve the use of informal expressions (or slangs), incorrectly spelt words and irregular grammatical constructs. Such irregularities lead to lexical, syntactic or semantic ambiguities in analysing textual content (Sorensen, 2009); and a suitable knowledge discovery technique to handle such ambiguities is text mining. Text mining is an intersection of techniques from information retrieval, text analysis, Natural Language Processing (NLP) and information classification domains, and can be used to provide computational intelligence (Irfan et al., 2015). Some researchers (e.g. Liu and Xiong, 2011) regard text mining as an extension of traditional data mining technique, since data mining involves the extraction of logical patterns from structured database. However, text mining involves a more complex procedure, as it applies in automated discovery of useful and interesting knowledge from unstructured and ambiguous textual data (Basu et al., 2001; Han, 2005). Several techniques have been proposed for text mining including conceptual structure, association rule mining, episode rule mining, decision trees and rule induction methods.

Alqhtani et al. (2015) developed a method to detect events from a Twitter stream. The procedure detects events by using the term frequency-inverse document frequency method, histogram of oriented gradient descriptors, grey-level co-occurrence matrix and colour histogram and finally *k*-nearest neighbour classification. Mosley (2012) was able to identify keywords and concepts in the social media data to help insurers by applying correlation, clustering and association analyses.

In their work, Wang and Tou (2013) performed text normalisation of social media texts written in an informal style using a novel beam-search decoder. They improved other downstream NLP applications by including other normalisation operations, e.g. missing word recovery and punctuation correction.

Gattani et al. (2013) designed an application that extracts entities from social data, such as tweets. Their system uses a Wikipedia-based global ‘real-time’ knowledge base that is well suited for social data, and generates and uses contexts and social signals to improve task accuracy.

To predict the future with social media Asur and Huberman (2010) applied both regression model and sentiment analysis classifier. Their results reveal that the buzz from social media can be accurate indicators of future outcomes.

However, a survey on the application of text mining in social networks conducted by Irfan et al. (2015) revealed that textual information is not structured according to the grammatical convention. They noted that people do not care about spellings and accurate grammatical construction, therefore making the extraction of logical patterns a critical task.

Karanikas et al. (2009) were able to mine temporal textual data for CI in the biotechnology and pharmaceutical industry. Their approach contained the following three stages: information extraction phase, ontology-driven generalisation of templates and discovery of associations over time. Another approach by Chakraborty (2014) organised and analysed textual data for customer intelligence based on social media data. In their approach, text analytics and sentiment mining using SAS® Text Miner and SAS® Sentiment Analysis Studio were used. The information obtained was used to improve business operations and performance. In the same line of work, He et al. (2013) applied SPSS Clementine text mining tool and Nvivo 9 on the customer-generated content (or UGC) on social media sites to perform a social media competitive analysis. Their results reveal that text mining is an effective technique to extract business value from the vast amount of available social media data.

Tulankar et al. (2013) predicted the current market trends along with accuracy measures based on sentiment analysis using data mining and implemented it with RapidMiner 5.0.

He et al. (2015a) analysed social media data using sentiment analysis and proposed a framework for social media CI. Their findings reveal that there is a need to create a social media data application which will be used for real-time social media CI, marketing intelligence and for identifying specific actionable areas in which businesses are leading or lagging. He et al. (2015b) further developed an innovative business-driven social media competitive analytics tool called VOZIQ. Fu et al. (2012) proposed a graph-based sentiment crawler which outperformed the methods it was compared with and also proved that sentiment analysis improves web content analysis. Other works related to sentiment analysis of social media data include, but are not limited to the following. Blatnik et al. (2014) improved on the traditional machine learning techniques by introducing sentiment class, i.e. the neutral class. Nasukawa and Yi (2003) were able to achieve high precision by focusing on identifying semantic relationships between sentiment expressions and subject terms. Buche et al. (2013) surveyed and analysed various techniques for the key tasks of sentiment analysis and concluded that for producing better summary based on feature-based opinions as positive, negative or neutral, the expectation–maximisation algorithm based on naïve Bayes method is the most efficient method. Pang and Lee (2008) in a review of techniques and approaches for opinion-oriented information seeking systems discovered among other things that there is increasing interest in opinion mining and sentiment analysis which is partly due to its potential applications. D’Andrea et al. (2015) also reviewing approaches, tools and applications for sentiment analysis concluded that these tools can be classified into machine learning approach, lexicon-based approach and hybrid approach, which is a combination of both the machine learning and the lexicon approaches.

3 Case study

A report by Suvarna and Banerjee (2014), corroborated by another report from Accenture (2014), identifies social media as a top trend in the banking industry. These reports outlined the inevitable role of social media in the banking industry to retain the brand loyalty of particularly digitally savvy customers. According to Suvarna and Banerjee (2014), it was reported that with social media increasingly becoming a popular media outlet among consumers with over 89% of customers surveyed stating they had asocial media account, it becomes necessary for the banking industry to intensify the integration of social media trends with banking services. Through social media platform, customers can engage in activities such as opening new bank accounts, access to financial advice, simplified transactions (e.g. transfer funds, bill payments, etc.), learn about new promotions, participate in customer surveys and get real-time feedback on complaints and/or enquiries. Furthermore, banks can also use the social media platforms to understand preferences of customers, build brand awareness, monitor customers' sentiments and advertise new products and services.

In the future, banks would be required to provide the same quality of service via social media platforms with more flexibility, like they strive to do through all other traditional channels (online banking, ATM, POS, mobile money, etc.). This move would open a new vista, which would enable banks to understand their individual customers' preferences in order to provide customised banking experience, similar to what is obtainable from platforms like Google or Amazon (Accenture, 2014). Customers can also have more convenient, personalised, instant and ubiquitous access to banking information and services, as part of everyday life. Some case studies of the integration of social media include the following: DenizBank (Turkey) becomes the first bank to open a Facebook branch that enable customers' to access accounts and transfer money to friends. Royal Bank of Canada (Canada), via social media, enabled P2P payments between Facebook friends, while ICICI Bank (India) launched a Facebook-based app that allows Facebook friends pay each other, track group expenses and upload funds to prepaid accounts. ASB Bank of New Zealand (New Zealand) took the social media integration further by creating a virtual branch similar to traditional walking-branch services, where customers clicks on photos of bank staff to initiate one-to-one chat. GTBank (Nigeria) launched its innovative 'Social Banking' service on Facebook, being the first in Nigeria, with over 950,000 Facebook fans. Although the social account differs from a regular GTBank account, the new channel allows GTBank social account holders to transfer money, purchase airtime, pay bills and confirm their account balance on Facebook. As the social media trend grows, more banks are expected strategically to set up and maintain social media presence to engage their customers, as some banks in Nigeria (e.g. GTBank) have utilised social media in advancing what is termed 'social banking', with a considerable followership base.

In order to understand patterns in the use of social media platforms by banks in Nigeria, we conducted a social media competitive analysis using with the five most prominent banks in Nigeria as a case study. The banks, namely Zenith Bank, FirstBank, Access Bank, GTBank and United Bank for Africa (UBA), are the top five banks that made the top 1000 banks in the world in 2014 based on Tier-1 capital, according to *The Banker* (2014), a publication of *Financial Times* newspaper, which is regarded as the most influential financial newspaper in the world. Zenith Bank was the top among Nigerian banks at 293, followed by GTBank (415), First Bank (424), Access Bank (532) and UBA (539). Zenith Bank was established in 1990, went public in 2004 and has over 500 branches, customer base in excess of 1.6 million accounts, a shareholder base of about 1 million and shareholder fund of 509.25 billion Naira in 2013. GTBank's business outlay spans Anglophone/Francophone, West Africa, East Africa and the UK. With an asset base of over 2.54 trillion Naira and shareholder fund of over 385 billion Naira, GTBank currently employs over 10,000 people in Nigeria, Cote d'Ivoire, Gambia, Ghana, Kenya, Liberia, Rwanda, Sierra Leone, Uganda and the UK. FirstBank, established in 1894, is Nigeria's largest financial services institution by total assets and gross earnings. With more than 10 million customer accounts, total assets of 3.3 trillion Naira and customer deposits of 2.6 trillion Naira, FirstBank has over 750 branches providing a comprehensive range of retail and corporate financial services. With over 366 branches worldwide, over 830,000 shareholders, 203 billion Naira in revenue and total asset worth 1.835 trillion Naira, Access Bank's vision is to emerge as the most respected African bank. Access Bank serves its various markets through four business segments: personal, business, commercial and corporate and investment banking. UBA started operations in 1948, and was incorporated in 1961. UBA is the first Nigerian bank to undertake an initial public offering and listed its shares on the Nigerian Stock Exchange in 1970.

3.1 Research questions

To the best our knowledge, no work has investigated the use of social media in banking in Nigeria and how these banks are using social media to engage customers and drive customer loyalty base. Therefore in this study, we focus on and examine data from Facebook and Twitter accounts of the five leading banks in Nigeria. In particular, *k*-means clustering and sentiment analysis were used to analyse unstructured text content. The following are the research questions that this study attempts to answer:

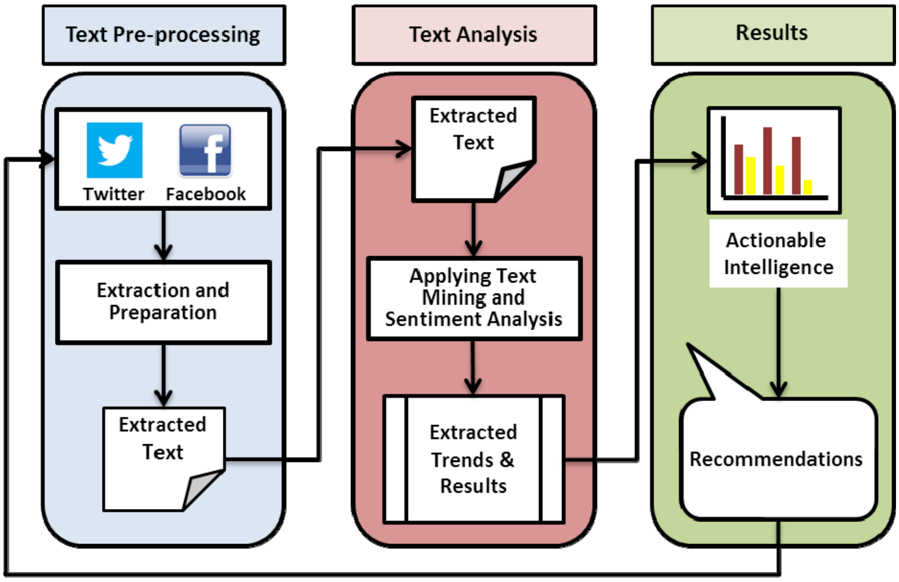
- What patterns can be found from the Facebook accounts of the respective banks?
- What patterns can be found from the Twitter accounts of the respective banks?
- Are there any differences in terms of the Facebook and Twitter patterns?

3.2 Methodology

In carrying out this research, data were gathered in two phases. In the first phase, account statistical data were gathered manually from the social media sites of the banks under investigation, including the number of fans/followers, the number of comments, the number of postings, the frequency of posting and the number of shares and likes. The data gathered in the second phase were actual textual contents available from the Facebook and Twitter accounts of these banks. These data were gathered in July 2015, which is usually the midyear for these banks when they want to evaluate what progress has been done so far in the year and re-strategise if need be (reference this). More specifically, the data and text posted were gathered between 1 and 31 July 2011. Text mining and sentiment analysis were performed on these data in order to harvest customers' opinion regarding the five best rated banks in Nigeria.

The methodology adopted in order to answer the research questions involves three major phases as captured in Figure 1. This method is based on He (2013) but modified to suit the specific purpose of this research.

Figure 1 Methodology (see online version for colours)



3.2.1 Text pre-processing

Textual documents were extracted from Facebook and Twitter accounts and were tokenised. We transformed the extracted tokens into lower case, followed by filtering of stop words. The textual corpus was stemmed using WordNet dictionary, in order to prepare the corpus for sentiment analysis. All the above-described process was achieved using the RapidMiner studio (https://rapidminer.com/). RapidMiner, which was formerly known as YALE, is one of the world's leading open-source data mining tool. It is a combination of leading-edge technologies with wide functional range. Its plug-ins provide more than 400 operators for all aspects of data mining. RapidMiner's meta operators automatically optimise the experiment designs and have a huge amount of visualisation techniques (www.rapidminer.com). The SentiWordNet plug-in used to enhance the sentiment analysis is a lexical resource in which each WordNet synset is associated with three numerical scores Obj(s), Pos(s) and Neg(s), describing how objective, positive and negative the terms contained in the synset are (Esuli and Sebastiani, 2006).

3.2.2 Text analysis

The text analysis stages include text mining using *k*-means clustering and sentiment analysis. According to Satish et al. (2012), *k*-means algorithm is one of the most widely used hard clustering techniques.

The algorithm works as follows:

- Specify the number of clusters (*k* in *k*-means).
- Randomly select *k* cluster centres in the data space.
- Assign data points to clusters based on the shortest Euclidean distance to the cluster centres.
- Re-compute new cluster centres by averaging the observations assigned to a cluster.
- Repeat above two steps until convergence criterion is satisfied.

The *k*-means clustering was used to discover the five major topics trending in Facebook and Twitter accounts. To achieve this, all the tweets for the five banks were combined and RapidMiner was used to cluster the data. This procedure was also repeated for Facebook comments too.

Sentiment analysis, also called opinion mining, analyses people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organisations, individuals, issues, events, topics and their attributes. It focuses on opinions which express or imply positive or negative sentiments (Bing, 2012). This research uses the 'Extract Sentiment' operator in RapidMiner studio to extract sentiment based on SentiWordNet 3.0.0, from the individual tweets and Facebook comment for each bank. This operator uses a WordNet 3.0 and a SentiWordNet 3.0.0 database to extract the sentiment of an input document. The sentiment value is in the range [–1.0, 1.0], where –1.0 means very negative and 1.0 means very positive. WordNet and SentiWordNet are connected by synset IDs. To determine the sentiment of a document, we compute the sentiment of each word, where the first meaning of a word has the most influence on a sentiment and each next meaning has less influence on a sentiment. Document sentiment is then computed as the average value of all word sentiments (https://rapidminer.com/).

3.2.3 Interpretation of results

The results obtained were visualised using line graphs and the interpretation on the basis for our recommendations.

3.3 Findings

3.3.1 Phase 1 findings

Figure 2 is the trend of tweets numbers in July 2015 (1–31 July) from the Twitter accounts of the five banks, available at the links below:

- First Bank (https://twitter.com/firstbankngr),
- GTBank (https://twitter.com/gtbank),

- Access Bank (https://twitter.com/myaccessbank),
- UBA (https://twitter.com/ubagroup)
- Zenith Bank (https://twitter.com/zenithbank).

A total of 746 tweets were collected from the five twitter account of all the five banks put together. Access Bank had the highest number of tweets (257) followed by GTBank (213), then First Bank (203), Zenith Bank (44) and finally UBA (24).

Figure 2 Tweet trends in July 2015

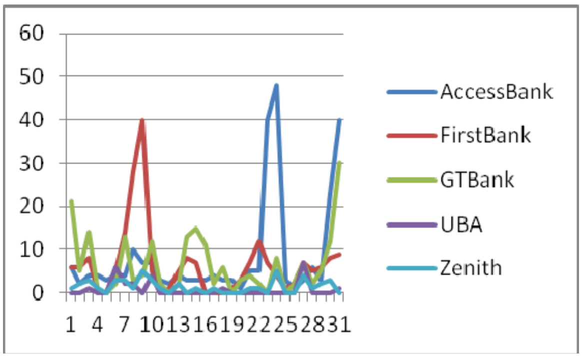


Figure 2 further expresses the fact that Access Bank has the highest number of tweets obtainable per day (48 tweets) which occurred on the 25th of July 2015. Also obvious from Figure 2 is the fact that UBA has the lowest engagement on Tweeter having 23 days of inactivity on Tweeter out of the 31 days investigated for the research. On the other hand, a total of 34,555 Facebook messages/comments were posted altogether. GTBank had the highest number of comments (16,999), followed by First bank (12,915), then Access Bank (2488), Zenith Bank (1649) and lastly UBA (504).

Figure 3 Access bank Facebook statistics

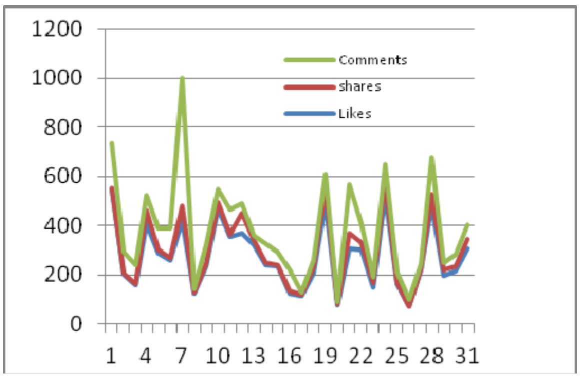
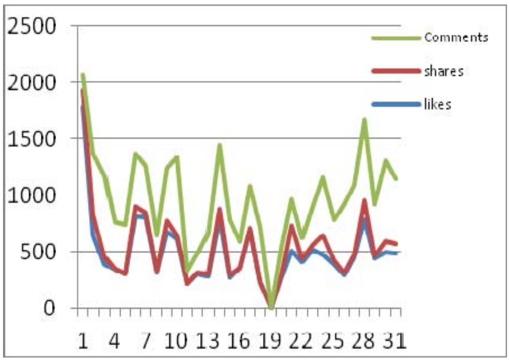


Figure 4 First bank Facebook statistics



According to McCorkindale (2010), Facebook comments can be posted by anyone who visits the site. Likes reveal the user who visits the site opinion of the current issue discussed. Figures 3–7 display the Facebook statistics as regards comments, likes and shares for the five banks in question for different days in July 2015.

Figure 5 GTBank Facebook statistics

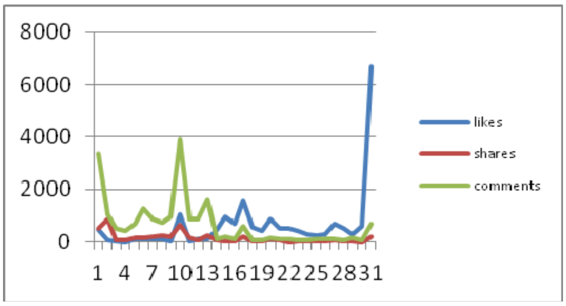
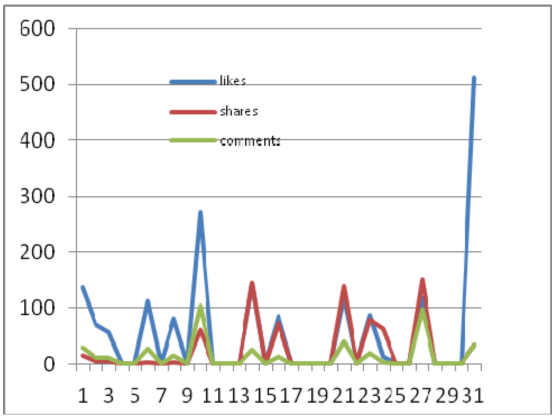


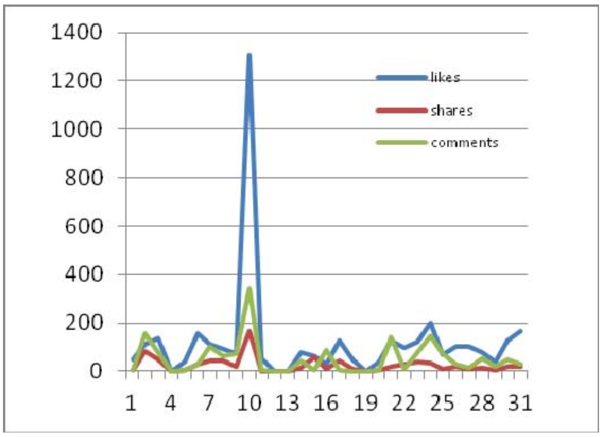
Figure 6 UBA Facebook statistics



From Figure 5, it is obvious the GTBank has the highest customer engagement via Facebook comments accruing to also the most comment a day (4000 on 10 July 2015). GTBank also has the highest likes on their post in a day (29 July 2015) though First Bank had the highest shares recorded in a day (Figure 4). Figures 2 and 6 reveal that UBA has the poorest customer engagement on both Facebook and Twitter.

In Figure 3, Access Bank had the highest comment per day of above 1000, highest shares per day of above 500 and highest likes per day of almost 500. In Figure 4, First Bank had the highest comments per day of above 2000 comments, highest shares per day and likes per day of almost 2000. GTBank had the highest comments per day of 4000 and the highest likes per day of above 6000, but the highest shares per day does not come up to 1000 (see Figure 5). From Figure 6, it is obvious that UBA had the highest comments per day of 100, the highest likes per day of about 500 and the highest shares per day of less than 200. Finally, Figure 7 shows Zenith Bank had the highest comments per day of above 300, the highest likes per day of above 1300, but the highest shares per day does not reach 200. Results therefore help understand the degree of activities of these banks on the social media.

Figure 7 Zenith bank Facebook statistics



3.3.2 *Phase 2 findings*

The result of phase 2 finding is divided into two parts: sentiment analysis and clustering.

(a) Sentiment analysis

The total tweets and the total Facebook comments for each bank in the month of July 2015 were combined for each bank and sentiment analysis was done for each of these combinations to generate the following results. The results are visualised in the graph shown in Figures 8 and 9, while the summary is contained in Table 1.

Figure 8 Sentiment analysis for the tweets of all five banks over the month of July 2015

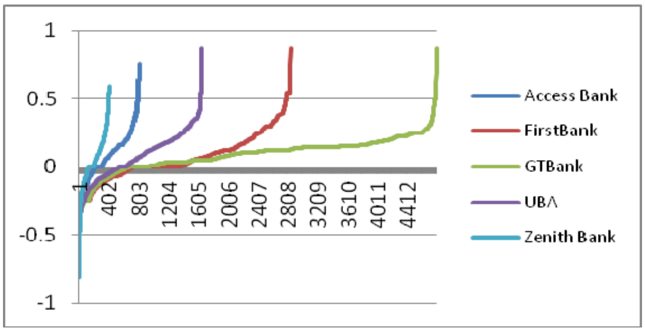
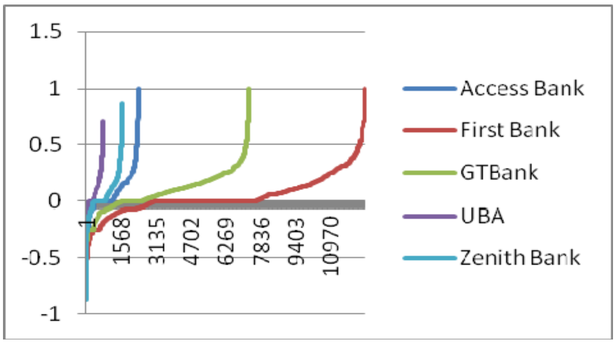


Figure 9 Sentiment analysis for the Facebook comments of all five banks over the month of July 2015



The summary of the sentiment analysis result from the graphs in Figures 8 and 9 is given in Table 1.

Table 1 Summarised sentiment analysis result

Bank	*Tweets*		*Facebook comments*	
	MNV	*MPV*	*MNV*	*MPV*
Access Bank	–0.5284	0.75	–0.625	1.0
First Bank	–0.75	0.875	–0.75	1.0
GTBank	–0.75	0.875	–0.75	1.0
UBA	–0.75	0.875	–0.5	0.709
Zenith Bank	–0.8182	0.5936	–0.875	0.875

Notes: MNV: most negative value; MPV: most positive value.

From Figure 8, it can be seen that Zenith Bank had the highest Most Negative Value (MNV), which is –0.8182 for the tweets obtained from just 44 tweets. It also had the lowest Most Positive Value (MPV) of 0.5936. This result requires further investigation to uncover the root cause warranting negative feedback in the tweets. Interestingly enough, the same pattern persists for the Facebook comments (see Figure 9), in which Zenith Bank was the highest MNV, which is –0.875.

(b) Clustering result

The clustering analysis result is presented as discovered from textual data from Twitter and Facebook.

Twitter. The tweets for all the five banks were combined into one file and were clustered. Table 2 contains the result of performing the *k*-means algorithm on 746 instances of data (total number of tweets), with 4092 attributes, i.e. number of distinct words that were clustered. The algorithm ran for ten iterations in order to get the five clusters shown in Table 2. *Cluster4* had the highest number of instances with 7175 items.

Table 2 Clustering Analysis (Twitter)

Cluster name	*Items*
Cluster 0	652
Cluster 1	1532
Cluster 2	325
Cluster 3	827
Cluster 4	7175

A summary of the five major themes talked about in the combined tweets of the banks include the following as discovered from the clusters.

- *Cluster 0.* Focused mainly on Tweeter followers making enquiry and also complaining about one thing or the other. This cluster also featured issues such as traffic and asking for help. Some individual followers' name such as *mubarack*, *henry*, *nelson*, *abbey*, *Jerome*, *Charles* and *bimbo* were also on focus.
- *Cluster 1.* Focused on expressing appreciation, featuring words such as *thank*, *kindly*, *hello*. It also clustered words such as *advise, Nigeria, guaranty, trust* and *bank*.
- *Cluster 2.* Focused on words such as *mobile*, *phone*, *apologise*, *number*; this clearly signifies an interest in following issues relating to mobile telecommunication network.
- *Cluster 3.* The focus of this cluster is not quite clear as it focuses on words such as *tweet*, *issue*, service, *help* and *hi*.
- *Cluster 4.* This cluster focuses on words such as *good custom* and *thanks*, while the names of some banks also came up, including *access*, *zenith* and *guaranty*. It might be safe to say that there is a form of appreciation for the good customs of these banks.

Facebook. Clustering was carried on the file containing a combination of the Facebook comments for all the five banks. Table 3 shows the result of performing the *k*-means algorithm on 24,651 instances of data (total number of comments); the attributes were 7543, i.e. the number of distinct words that were clustered. The algorithm ran for ten iterations in order to get the five clusters shown in Table 3. Cluster 1 had the highest number of instances with 17,406 items.

Table 3 Clustering analysis (Facebook)

Cluster name	*Items*
Cluster 0	2362
Cluster 1	17,406
Cluster 2	756
Cluster 3	821
Cluster 4	3306

A summary of the five major themes talked about in the combined Facebook comments of the banks include the following as discovered from the clusters.

- *Cluster 0*. Focused mainly on words such as registered and number. This might not be far from the fact that banks were compelling customers to register for their bank verification numbers at the time when the data were gathered.
- *Cluster 1.* Focused on positive words such as *joy*, *thanks*, *nice*, *ok* and *good*. Interestingly, the name of one of the banks, *access*, featured in this cluster.
- *Cluster 2.* Did not have a particular focus, but words such as *deposit*, *requirement*, *money*, together with a lot of individual names, were found in this cluster.
- *Cluster 3.* The focus of this cluster is on words such as *cards*, *account*, *debit* and *loan*, together with quite a lot of customers' names.
- *Cluster 4*. This cluster focused mainly on banking-related terms such as *atm*, *card*, *account*, *bank*, *interest*, *money*, *transfer*, *branch*, *withdraw*, *balance*, *loan* and *pin.*

4 Discussions

In answering the first research question (*What patterns can be found from the Facebook sites of the respective banks?*), we discovered that generally on Facebook, the activities of the banks can be categorised into three: very active, averagely active and least active. GTBank and First Bank are in the most active category, followed by the average active category which contains Access Bank and Zenith Bank. UBA is in the least active category. For the sentiment result, Access Bank has the lowest negative sentiment analysis result, while Zenith Bank has the highest negative sentiment analysis result. Access Bank, First Bank and GTBank have the highest positive sentiment analysis result which is 1.0. Also, the largest cluster, cluster 1, in the Facebook cluster analysis is focused on words such as *joy*, *thanks*, *nice*, *ok* and *good*, which are positive and appreciative words.

For the second research question (*What patterns can be found from the Twitter accounts of the respective banks?*), we discovered that Access Bank is the most active bank on Twitter with the least negative sentiment value. This invariably means that inasmuch as Access Bank is active on Twitter, their followers do not express as much negative opinion, which is in turn a plus on bank's reputation and business. This sentiment analysis result happens to be consistent with that of the Facebook sentiment result. UBA, though the least active on Twitter, still has a fairly high negative sentiment analysis result. On Twitter, Zenith Bank has the highest negative sentiment value and the lowest positive sentiment value. For the clustering result, the biggest cluster for Twitter analysis focused on words like *thanks*, *kindly* and *hello*, signifying that the best way to attract and keep customers is through appreciation.

To answer the third research question (*Are there any differences in terms of the Facebook and Twitter patterns?*), we discovered that the patterns on the Twitter and Facebook accounts for each banks are similar, which confirms that the result obtained from the social network accounts conforms to the true opinions of the followers or fans as the case may be.

Table 4 Summary of analysis results

	Research question	*Activeness*	*Sentiment pattern*	*Cluster explanation*
1	What patterns can be found from the Facebook sites of the respective banks?	• Most active *(GTBank and First Bank)* • Averagely active *(Access Bank and Zenith Bank)* • Least active *(UBA)*	• Highest negative sentiment *(Zenith Bank)* • Highest positive sentiment *(Access Bank, First Bank and GTBank)*	Largest cluster focused on *joy*, *thanks*, *nice*, *ok* and *good*, which are positive and appreciative words.
2	What patterns can be found from the Twitter sites of the respective banks?	• Most active *(Access Bank)* • Least active *(UBA)*	• Highest negative sentiment *(Zenith Bank)* • Highest positive sentiment *(Access Bank, First Bank, GTBank)*	Largest cluster focused on *thanks*, *kindly*, *hello*, signifying that the best way to attract your customers is through appreciation.
3	Are there any differences in terms of the Facebook and Twitter patterns?	Similar patterns revealed for UBA	Similar patterns revealed for Zenith Bank, Access Bank, First Bank, GTBank and UBA	Similar patterns revealed for Zenith Bank, Access Bank, First Bank, GTBank and UBA

5 Implications and recommendations

Based on the results of the analysis, it is evident that most of these banks have recognised and embraced social media trends in banking industry; however, for maximum effectiveness it is expected that a more holistic social media strategy is put in place if banks are to maximise the benefits of social media for competitive advantage. It is important that banks in developing countries like Nigeria should expedite the process of adoption and integration of social media channels that provide richer customer experiences via these channels. This would in turn increase the customer base, brand loyalty and ultimately boost the bottom line. More specially, we make the following recommendations:

- *Continuously mine social media for competitive intelligence*. An effective social media implementation and monitoring strategy, using text mining, can help banks to measure their own social media presence and its impact, while they benchmark this performance against the competitors to determine a new line of action and initiate strategic moves to maximise market share. Also, the large volume of UGC provides the basis to strive for personalised services and experiences targeted at specific demographic, enabling better customer service.

- *Integration of social media sites into banks' mobile apps*. The success of mobile banking has made it possible for many customers to perform transactions on the go via a mobile app. Integrating social media into these mobile apps provides customers greater access to connect with their banks via these channels. Banks can readily engage more requests and provide answers or advice to customers' requests, thereby encouraging customers to converse directly with their banks, rather than make general casual comments on social media. Such connections create room for more positive sentiments, particularly when customer's expectation has been exceeded.
- *Provide means for transactions on social media*. Apart from being able to access promotional information, ask questions and receive feedback via social media, banks in Nigeria should also explore the use of social media to allow its customers to perform bank transactions like fund transfer to their friends on social media, check account balance and possibly pay for a concert or event found while navigating social media sites.

6 Conclusion and future research

In conclusion, the patterns discovered from the analysis performed in this study are sufficient basis for the banks under review to pay more attention to their customers' preferences and concerns and address those issues more effectively. Furthermore, the results obtained are a foundation for further investigation by the banks. For example, there may be a need for Zenith Bank to dig deeper into the reason for the high negative sentiment value result obtained. In addition, banks such as UBA, though doing well financially, could do even better if they invest more to improve their social media presence.

References

Accenture (2014) *Moving Beyond Listening and Monitoring: Social Media Marketing for Financial Services*, Accenture, Chicago, IL.

Agarwal, K.N. (2006) 'Competitive intelligence in business decisions: an overview', *Competition Forum*, Vol. 4, pp.309–314.

Aggarwal, C.C. and Wang, H. (2011) 'Text mining in social networks', in Aggarwal, CC (Ed.): *Social Network Data Analytics*, Springer, New York, pp.353–378.

Albescu, A., Pugna, I. and Paraschiv, R.D. (2009) 'Business competitive intelligence: the ultimate use of information technologies in strategic management', *Informatica Economica*, Vol. 13, No. 4, pp.39–50.

Alqhtani, S.M., Luo, S. and Regan, B. (2015) 'Fusing text and image for event detection in Twitter', The *International Journal of Multimedia & Its Applications*, Vol. 7, pp.27–35.

Asur, S. and Huberman, B.A. (2010) 'Predicting the future with social media', *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 31 August–3 September, Toronto, Canada, pp.492–499.

Bărbulescu, A., Bucurestia and România (2007) 'Competitive intelligence and internet sources', *Revista Informatica Economică*, Vol. 3, pp.80–82.

Barnes, D. and Ganim, N. (2011) 'Society for new communications research study: exploring the link between customer care and brand reputation in the age of social media', *Journal of New Communication and Research*, Vol. 5, pp.23–37.

Basu, S., Mooney, R.J., Pasupuleti, K.V. and Ghosh, J. (2001) 'Evaluating the novelty of text-mined rules using lexical knowledge', *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 26–29 August, San Francisco, CA, pp.233–238.

Bing, L. (2012) *Sentiment Analysis and Opinion Mining: Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers, San Rafael, CA.

Blatnik, A., Jarm, K. and Meza, M. (2014) 'Movie sentiment analysis based on public tweets', *Elektrotehniški vestnik*, Vol. 81, No. 4, pp.160–166.

Buche, A., Chandak, M.B. and Zadgaonkar, A. (2013) 'Opinion mining and analysis: a survey', *International Journal on Natural Language Computing*, Vol. 2, No. 3, pp.39–48.

Chakraborty, G. (2014) 'Analysis of unstructured data: applications of text analytics and sentiment mining', *Proceedings of the SAS Global Conference*, 23–26 March, Washington, DC.

Choudhury, S. and Alani, H. (2014) 'Personal life event detection from social media', *Proceedings of the 25th ACM Hypertext and Social Media Conference*, 1–4 September, Santiago, Chile, pp.1–4.

D'Andrea, A., Ferri, F., Grifoni, P. and Guzzo, T. (2015) 'Approaches, tools and applications for sentiment analysis and implementation', *International Journal of Computer Applications*, Vol. 12, No. 3, pp.26–33.

Dewing, M. (2012) *Social Media: An Introduction*, Library Parliament Publication No. 2010-03-E.

Esuli, A. and Sebastiani, F. (2006) 'SentiWordNet: a publicly available lexical resource for opinion mining', *Proceedings of the 5th International Conference on Language Resources and Evaluation*, 22–28 May, Genoa, Italy, pp.417–422.

Farooq, F., Jan, Z. and Karachi, S. (2012) 'The impact of social networking to influence marketing through product reviews', *International Journal of Information and Communication Technology Research*, Vol. 2, pp.627–637.

Fox, S., Zickuhr, K. and Smith, A. (2009) *Twitter and status updating*, Pew Internet and American Life Project. Available online at: http://www.pewinternet.org/2009/02/12/twitter-and-status-updating/

Gattani, A., Lamba, D.S., Garera, N., Tiwari, M., Chai, X. and Das, S. (2013) 'Entity extraction, linking, classification, and tagging for social media: a Wikipedia-based approach', *Proceedings of the VLDB Endowment*, Vol. 6, No. 11, pp.1126–1137.

Gémar, G. and Jiménez-Quintero, J.A. (2015) 'Text mining social media for competitive analysis', *Tourism & Management Studies*, Vol. 11, pp.84–90.

Gray, P. (2010) 'Competitive intelligence', *Business Intelligence Journal*, Vol. 15, pp.31–37.

Han, J. (2005) *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers Inc., San Francisco, CA.

He, W., Zha, S. and Li, L. (2013) 'Social media competitive analysis and text mining: a case study in the pizza industry', *International Journal of Information Management*, Vol. 33, pp.464–472.

He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G. and Tao, R. (2015a) 'Gaining competitive intelligence from social media data: evidence from two largest retail chains in the world', *Industrial Management & Data Systems*, Vol. 115, No. 9, pp.1622–1636.

He, W., Wu, H., Yan, G., Akula, V. and Shen, J. (2015b) 'A novel social media competitive analytics framework with sentiment benchmarks', *Information & Management*, Vol. 52, pp.801–812.

Irfan, R., King, C.K., Grages, D., Ewen, S., Khan, S.U., Madani, S.A. et al. (2015) 'A survey on text mining in social networks', *The Knowledge Engineering Review*, pp.157–170.

Jaworski, B.J., Macinnis, D.J. and Kohli, K.A. (2002) 'Generating Competitive Intelligence in Organizations', *Journal of Market-Focused Management*, Vol. 5, pp.279–307.

Kahaner, L. (1996) *Competitive Intelligence: How to Gather, Analyse, and Use Information to Move Your Business to the Top*, KANE Associates International, Inc., New York.

Karanikas, H., Koundourakis, G., Kopanakis, I. and Mavroudakis, T. (2009) 'A temporal text mining application in competitive intelligence', *Proceedings of the 23nd European Conference on Operational Research*, 5–8 July, Bonn, Germany, pp.1–17.

Kwak, H., Lee, C., Park, H. and Moon, S. (2010) 'What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web*, 26–30 April, Raleigh, NC, doi:10.1145/1772690.1772751.

Liu, F. and Xiong, L. (2011) 'Survey on text clustering algorithm: research present situation of text clustering algorithm', *Proceedings of the IEEE 2nd International Conference Software Engineering and Service Science*, 15–17 July, Beijing, China, pp.901–904.

McCorkindale, T. (2010) 'Can you see the writing on my wall?: a content analysis of the Fortune 50's Facebook social networking sites', *Public Relations Society of America*, Vol. 4, pp.1–13.

McGonagle, J.J. and Vella, C.M. (2002) 'A case for competitive intelligence', *Information Management*, Vol. 36, pp.35–40.

Nasri, W. (2011) 'Competitive intelligence in Tunisian companies', *Journal of Enterprise Information Management*, Vol. 24, pp.53–67.

Nasri, W. (2012) 'Conceptual model of strategic benefits of competitive intelligence process', *International Journal of Business and Commerce*, Vol. 1, pp.25–35.

Nasukawa, T. and Yi, J. (2003) 'Sentiment analysis: capturing favorability using natural language processing', *Proceedings of the 2nd International Conference on Knowledge Capture*, 2–4 July, Graz, Austria, pp.70–77

Pang, B. and Lee, L. (2008) 'Opinion mining and sentiment analysis', *Foundations and Trends' Information Retrieval*, Vol. 2, Nos. 1–2, pp.1–135.

Mosley Jr., R.C. (2012) 'Social media analytics: data mining applied to insurance Twitter posts', *Casualty Actuarial Society E-Forum*, Vol. 2, pp.1–36.

Satish, G., Goutam, C. and Gary, G. (2012) 'Comparison of *k*-means, normal mixtures and probabilistic-d clustering for b2b segmentation using customers' perceptions', *Proceedings of the SAS Global Forum*, 22–25 April, Orlando, FL.

Sorensen, L. (2009) 'User managed trust in social networking comparing Facebook, MySpace and LinkedIn', *Proceedings of 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic System Technology*, 17–20 May, Aalborg, Denmark, pp.427–431.

Suvarna, V.K. and Banerjee, B. (2014) *Social Banking: Leveraging Social Media to Enhance Customer Engagement*, Capgemini, Paris.

The Banker (2014). *The Banker Top 1000 World Banks 2014 rankings*, the Banker. Available online at: http://www.thebanker.com/Top-1000-World-Banks/The-Banker-Top-1000-World-Banks-2014-rankings-UK-Press-release-For-immediate-release (accessed in November 2015).

Tulankar, S., Athale, R. and Bhujbal, S. (2013) 'Sentiment analysis of equities using data mining techniques and visualizing the trends', *International Journal of Computer Science Issues*, Vol. 10, No. 4, pp.265–269.

Wang, P. and Tou, N.H. (2013) 'A beam-search decoder for normalization of social media text with application to machine translation', *Proceedings of NAACL-HLT*, 9–14 June, Atlanta, GA, pp.471–481.

Yang, T.A., Kim, D.J. and Dhalwani, V. (2007) 'Social networking as a new trend in e-marketing', *Proceedings of the International Conference on Research and Practical Issues of Enterprise Information Systems*, 14 October–16 March, Beijing, China, pp.847–856.