

# The Use of Sentiment Analysis and Topic Modelling to Understand Online Communicative Ecologies in MobiSAM

Hannah THINYANE<sup>1,2</sup>, Mamello THINYANE<sup>1</sup>

<sup>1</sup>*United Nations University Institute on Computing and Society (UNU-CS)*

*Estrada do Engenheiro Trigo No. 4, Macau SAR*

<sup>2</sup>*Department of Information Systems, Rhodes University, Grahamstown 6139, South Africa*

*Tel: +853 2871 2930, Email: {hannah, mamello}@unu.edu*

**Abstract:** Communicative ecologies are a tool that can be used to understand the existing use of information and communication tools within a specific community. By using the ecology metaphor to understand the interaction between the technological, discursive, and social layers within the community, this research develops a holistic understanding of citizens' communication surrounding the MobiSAM project. This paper proposes the use of sentiment analysis and topic modelling to understand how citizens are currently using technology for political participation. The paper argues that this rich understanding of current use of technology for participation can support the embedding of further interventions into existing communicative ecologies.

**Keywords:** Communicative ecologies, sentiment analysis, topic modelling, m-participation, ICT4D.

## 1. Introduction

In their report, Closing the feedback loop: can technology amplify citizen voices, Gigler and Bailur [1] present a theoretical framework addressing the links between technology, citizen participation and improvements to well-being, based on Sen's [2] capability approach. This approach moves beyond assessing easily measurable outputs such as access to Internet, and instead measures the effects that access has on the person's well being. Gigler and Bailur's approach is based on an understanding of: existing information ecologies; access to and use of ICTs; and level of ICT usage. Once this information is understood, Gigler and Bailur argue that appropriate interventions and trainings can be performed to enhance the informational capabilities and then the human social capabilities of participants [1]. In other research, we have proposed the use of communicative ecologies as a way of identifying the first three links in Gigler and Bailur's ICT Impact Chain. Communicative ecologies allow the researcher to uncover within the confines of a specific locale, the social, technological and discursive context of the use of ICTs. This approach relies on building a holistic understanding of the nature of communication before seeking to formulate interventions. By negating assumptions of homogenous use of ICTs in communities, far more specific understandings of how best to structure interventions that aim to address current ICT use in the maintenance of daily communications networks.

Communicative ecologies are used to map both online and offline patterns of interaction and communication within a community. Other work has discussed a baseline study that was used to obtain a broad understanding of the ways citizens currently: have access to technology, how they use technology, and how they currently participate in local government. A similar study was undertaken to understand the ways that municipal

employees use technology, and how they communicate with citizens. This study looks more specifically at the MobiSAM project, to understand how citizens are currently using technology to communicate within MobiSAM. It performs this analysis by exploring the discursive and technical aspects of MobiSAM's online streams: Facebook, twitter, blog, MobiSAM page.

This paper begins with some background on the MobiSAM project, followed by an introduction to the concept of communicative ecologies. It then presents a discussion on the analysis evaluation techniques employed in this study, followed by the results of the evaluation.

## **2. MobiSAM**

MobiSAM was formed in 2011 to investigate the use of mobile phones to increase citizen participation in local government around the area of service delivery. The first phase was funded by Ford Foundation (from 2011 to 2014, referred to as MobiSAM v1.0) and now the second phase is funded by Making All Voices Count (June 2016 – June 2017, referred to as MobiSAM v2.0). MobiSAM is adapted from a mobile polling application that allows individuals to use their mobile phones to voice their opinion or provide information in response to a number of different types of questions (or polls).

From a high level, the purpose of MobiSAM is to empower citizens to participate in local government, particularly in the area of service delivery. Although MobiSAM is offered to all citizens, the largest proportions of the population that do not currently receive basic services are marginalized communities, and so they are the assumed ultimate beneficiaries of the project.

MobiSAM offers the following channels for citizens to engage with local government: Facebook, twitter, blog, MobiSAM page, MobiSAM client, and SMS gateway. The latter two channels have been temporarily suspended, while revisions are made to the system so are not included in the analysis detailed in this paper.

### **2.1 Methodology**

The MobiSAM project is based on a case study, which is defined as a “strategy for doing research which involves an empirical investigation of a particular contemporary phenomenon in its real life context using multiple sources of evidence” [3, p. 52]. This approach allows researchers to focus on the primary strength of case study methodology, that it “enables the uncovering of events or processes that one might miss with more superficial methods” [4, p. 238]. A case study approach allows an in depth study of specific factors that influence, in this case, the use of technology for citizen participation. In doing so, it emphasizes linkages between local processes and the broader socio political and economic context in which they occur [5], [6]. Critics of case study research cite the overt specificity and intense focus that is required when performing the in depth analysis. Drawing from this, critics note that case study findings are so specific, that they cannot be transferred to other contexts. Yin however argues “case studies, like experiments, are generalizable to theoretical propositions and not to populations or universes ... the investigator's goal is to expand and generalize theories and not to enumerate frequencies” [7, p. 10].

Within the MobiSAM project, a mixed method research design is used, as this has been found to improve trustworthiness (internal validity), transferability (external validity), reliability (dependability) of findings [6]. This mix of quantitative and qualitative data and analysis is used to reveal different aspects of a single reality [8]. The next section describes the context in which the MobiSAM project is situated.

## 2.2 Context

Makana Municipality is situated in the Eastern Cape Province; one of South Africa's poorest provinces. The area experiences a high unemployment rate (65.52%) and low levels of formal education (28.52% of residents have received a primary school education or none at all) [9]. As a result, 64.35% of Makana residents live below the poverty line [9] (defined by Statistics South Africa as earning less than R801 per month, which equates to the earning bracket above the poverty line [10]). Efforts to alleviate poverty are hampered by a sluggish economy and there is an increasing dependency on social grants, 45.5% of the total Makana population receiving some form of government grant in 2011 (compared with approximately 29% of the national population receiving some form of government grant in 2011 [9], [11]). This places a significant burden on the municipality, particularly in delivering basic services. The municipality is responsible for (amongst other things) municipal health services; municipal roads; sanitation; electricity reticulation; potable water; refuse removal, refuse dumps and solid waste disposal; child care facilities; and local tourism [9].

The situation in Makana Municipality is comparable to many local municipalities across South Africa, where “local government capacities are in short supply and financial sustainability is frequently in doubt. This hampers total government ability to perform traditional functions such as service delivery and regulation, collecting rates, user charges and fees” [12, p. 337]. However, this is not a problem that is unique to South Africa, but has been demonstrated on numerous occasions and across a number of different contexts. A recent cross-country, empirical evaluation found similar evidence across developing countries around the world [13].

## 3. Communicative Ecologies

Communicative ecologies consists of a “range of communication activities that link networks of individuals, groups and institutions in a specific community domain” [14, p. 360]. By using an ecological metaphor to understand the interaction between ICTs, formats (platforms) and activities which organize communication within the confines of a specific locale or geographic space, this theory seeks to recognise significant social processes that play a role in the social organization of the daily lives of individuals and communities [14].

What this theory allows is a nuanced approach to uncovering these “communication activities” because, “in order to understand one aspect of communication within a particular setting, you need to understand how it fits into the wider communicative ecology” [15]. This interrogation of the use and value of communication networks and the platforms which facilitate these ‘channels’ has been refined identifying three core aspects involved in both conceptualizing a “communicative ecology” as well as methodologically seeking to locate and organize the various ‘levels or strata’ of activity. This conceptual model therefore represents, “communication as a complex arrangement of technical, social, and discursive contexts in which communication processes occur” [16, p. 1]. By understanding these networks as an arrangement of social, technological, and discursive elements allows the research to interrogate the, “multi-modal communication connections, shaped by particular social and cultural conditions” that are used by people to “construct knowledge and achieve goals” [17, p. 328]. This approach is sensitive to the complexity of everyday life of communities and puts forth a method which supports the use of existing modes of communication as the most suitable avenues through which to connect with people and communities. Communicative ecology theory can therefore be regarded as to some degree a “corrective to technological determinism inherent in overenthusiastic speculation about the impact of these new technologies” [18, p. 211].

## 4. Evaluation techniques

Recently there have been a number of studies aimed at using online, big data to understand protest communication ecologies [19]. These studies have used a number of different techniques including: granger causality to determine if social media feeds can be useful in forecasting social protests [20]; connection action to understand how crowd organization is produced [21]; and sentiment analysis to monitor electoral campaigns [22].

Beyond basic descriptive statistics, sentiment analysis and topic modelling have been used in this research to understand both the attitudes of the communicators and the underlying topic structures across the various channels within the MobiSAM communicative ecology.

### 4.1 Sentiment Analysis

Sentiment Analysis is a technique that analyses written language to determine people's attitudes, opinions and emotions regarding a particular topic or issue. Basic analysis uses a classifier to categorize a given piece of text based on polarity as either positive, neutral or negative; or based on the subjectivity and objectivity identification [23]. More advanced techniques classify text into multiple categories of emotional states such as anger, sadness, trust and joy. Fundamentally sentiment analysis is a text classification technique, and therefore uses supervised learning methods (e.g. Naive Bayes classification, Support Vector Machines) for the classification at the level of a document, a sentence or a more fine grained feature level [24].

### 4.2 Topic Modelling

Topic Modelling is a text analysis technique that uses statistical topic models to determine the key topics and their associated terms from a set of documents. Topic modelling allows for a better understanding of the underlying semantic and topical structure from a text corpus. Two of the commonly used topic models and techniques are Probabilistic Latent Semantic Indexing (PLSI) [25] and the more generalized version Latent Dirichlet Allocation (LDA) [26].

## 5. Evaluation

The MobiSAM channels that were considered for analysis are the Facebook and Twitter channels, which provide a relatively more text-rich and unconstrained communication engagement from the citizen. The MobiSAM platform only allows citizens to express their opinion through a controlled polling system, while the blog is largely articles posted by the MobiSAM project team.

In total 434 tweets from the MobiSAM twitter handle posted between 19-05-2016 and 12-12-2016, 879 Facebook statuses posted between 10-10-2013 and 9-9-2016, as well as 1585 Facebook comments posted between 10-10-2014 and 9-9-2016 were analysed.

The analysis made use of both R and Python statistical and text analysis tools. In particular the *Syuzhet* package for R was used with the *NRC* dictionary method for the sentiment analysis, and both the *TfidfVectorizer* from the *sklearn* package, and the *LdaModel* from the *gensim* package for python were used for the topic modelling.

Preliminary data cleaning involved removal of stopwords, punctuation, and URLs from the texts. For the sentiment analysis, the individual multi-factor emotion valence scores were calculated for each text document (i.e. a tweet, a status update, or a comment) and a mean score calculated across each channel. For topic modelling *TF-IDF* was used for the vectorization of the text, before applying the *LDA* model. Topics each with five terms were determined across the different MobiSAM channels.

## 5.1 Results

The mean sentiment across the various channels was calculated at 0.77 ( $\sigma = 1.52$ ) for Facebook statuses, 0.24 ( $\sigma = 1.0$ ) for Facebook comments, and 0.43 ( $\sigma = 0.85$ ) for the tweets. Figure 1 below gives the mean emotion valence across the 10 factors including positivity and negativity classification. Beyond the actual sentiment values noted, the relative scoring between different channels and across the emotional factors provides a better consideration of the MobiSAM communicative ecology dynamics.

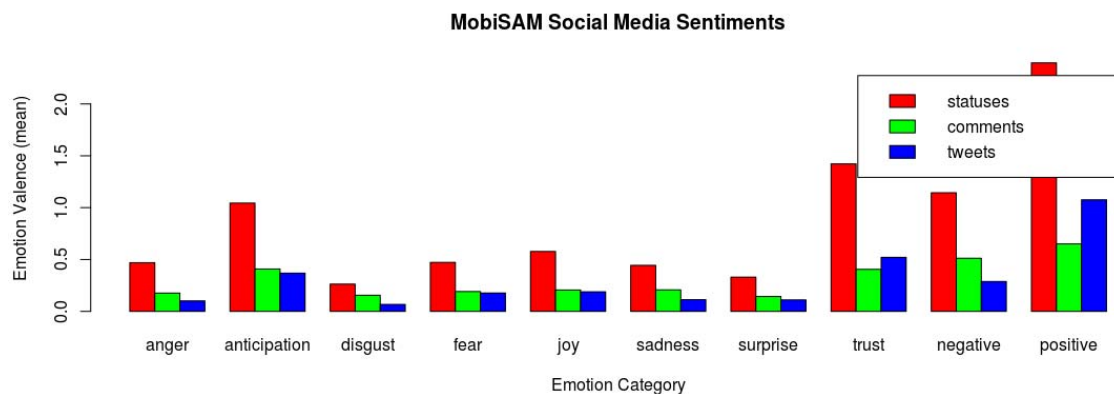


Figure 1- MobiSAM Social Media Multifactor sentiment

A more honed in analysis was undertaken on Facebook statues and comments to understand the sentiment trend over time (Figure 2).

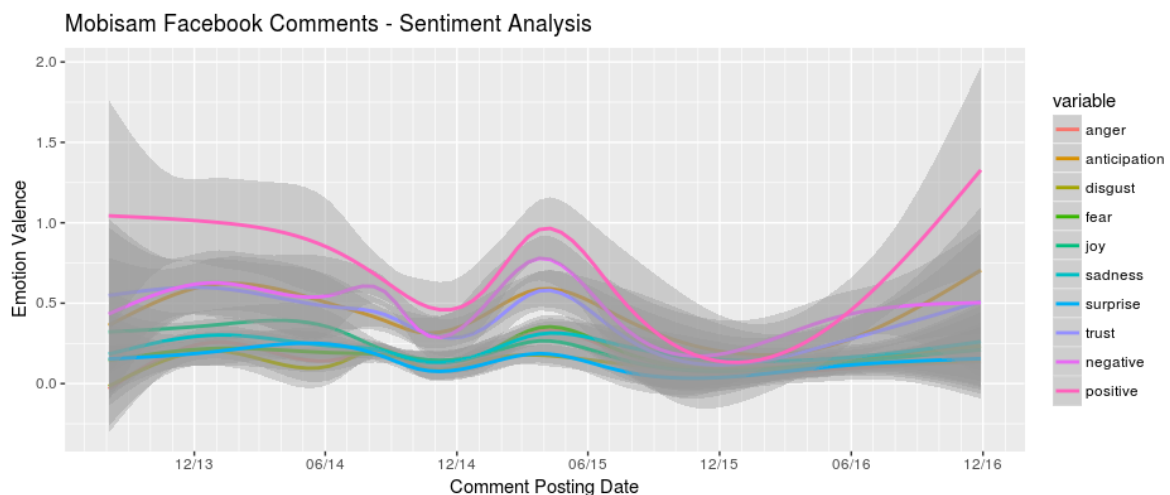


Figure 2- MobiSAM Facebook comments sentiment trend

Facebook represents the most interactive and active channel for MobiSAM, indicated by the high number of engagements (e.g. “likes”, ‘shares”, “comments”) with issues as shown in Figure 3.



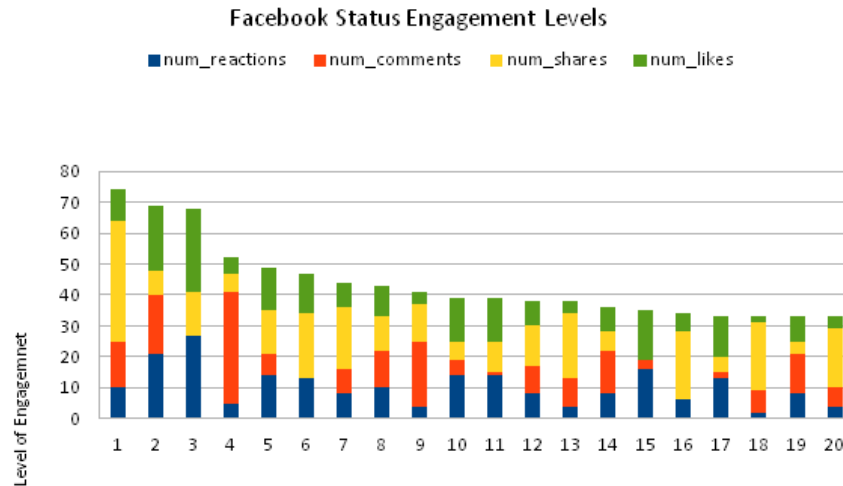


Figure 3- Top twenty most engaged Facebook statuses

The topic terms that have been noted for the various MobiSAM channels are given in Table 1.

Table 1- LDA topic model key terms

Facebook Statuses	Facebook Comments	Tweets
water : 0.01345	water : 0.0193	mobisam : 0.0090
makana : 0.0088	thanks : 0.0122	water : 0.0086
update : 0.00712	street : 0.0089	makana : 0.0085
municipality : 0.0061	thank : 0.0072	da : 0.0072
supply : 0.0044	update : 0.0058	says : 0.0062

## 5.2 Discussion

While there are a number of noted limitations of sentiment analysis and in particular a few levelled at the implementation in the *Syuzhet R* package, it still remains a useful tool for exploratory analysis of opinions within a text corpus. From the analyses of the MobiSAM channels the following observations can be made.

Facebook is the channel on which most engaged communication interaction takes place. This is reflected in the overall high number of engagements (i.e. likes, shares) within the top twenty most engaged status updates, compared to the other channels. Beyond the fact that the platform allows for free text comments as well as shares, for which Twitter has similar functionality, the high engagement level could be attributable to the fact that Facebook was the first platform (earliest post in 10-10-2013) to be used for sharing service delivery updates.

The status update with the most number of comments is “*A couple of reports coming in of no water this morning. Anyone else without water?*”, and the one with most shares is “*Easy to use calendar showing the different stages of loadshedding (according to Eskom's schedule)*”. Based on the analysis of the content of these updates as well as the resultant nature of the citizens engagement one can note a level of attentiveness and responsiveness from the citizens, both of which are important attributes of a communicative interaction. On the other hand, the status with the most number of likes is one that refers to the legislation (i.e. South African Water Services Act) and informs the citizen’s about their rights regarding provisioning of clean water and sufficient sanitation.

The main consideration for the topic modelling analysis employed in this research has been to highlight key topical structures across the different MobiSAM channels, and not necessarily to have the most optimized topic modelling e.g. the choice of the number of

topics and terms per topic wasn't informed by any optimization analysis such as perplexity evaluation. However, having also undertaken a further exploratory run of the LDA model with 10 topics and 15 terms per topic, the general observation is that the key service delivery issue mostly discussed across all the channels is related to "water". The nature of the engagement on this topic and other related topics is bidirectional between the citizens and government. This is indicated by the topic prevalence on channels that are geared more for government to citizen communications such as tweets and Facebook status updates; as well on channels that are more utilized for citizen to government, and citizen to citizen engagement such as the Facebook comments.

An understanding of the nature of the communication between the citizens and government, and among the citizen through the analysis of both the sentiments and the topics of the discussions sheds further light on and assists both government and civil society stakeholders with an understanding of the positioning and role of the MobiSAM platform within the wider communicative ecology.

## 6. Conclusions

Communicative ecologies is a theoretical perspective and a tool that has been used in this research to understand the use of the MobiSAM social accountability monitoring platform in context of a discursive engagement between citizens and the government. The paper has presented and highlighted how the textual analysis techniques of sentiment analysis and topic modelling can enrich an understanding of the communicative ecology dynamics within a specific community. In this case the analysis has only been used to successfully shed light on the technological and discursive aspects of the communicative ecology framework. The social aspects, which has been beyond the scope and focus of this paper could be explored through network analysis of the interactions between the various actors and role-players, and this remains an aspect of possible future research.

## Acknowledgements

Thanks goes to Making All Voices Count for their generous financial support for the MobiSAM project, and to National Research Foundation for their support of the primary author. Thanks also to Karthik Bhat for technical assistance in extracting Facebook data.

## References

- [1] B.-S. Gigler and S. Bailur, Eds., *Closing the Feedback Loop: Can Technology Bridge the Accountability Gap? Directions in Development*. The World Bank, 2014.
- [2] A. Sen, *Development as Freedom*. New York: Anchor Books, 1999.
- [3] C. Robson, *Real world research: A resource for social scientists and practitioner-researchers*. Oxford, UK: Blackwell, 1993.
- [4] F. Erickson D., "Qualitative methods in research on teaching," in *Handbook of research on teaching*, 3rd ed., New York, NY: MacMillan, 1986, pp. 119–161.
- [5] A. M. Pettigrew, "Longitudinal Field Research on Change: Theory and Practice," *Organ. Sci.*, vol. 1, no. 3, pp. 267–292, 1990.
- [6] G. B. Rossman and B. L. Wilson, "Numbers and Words Combining Quantitative and Qualitative Methods in a Single Large-Scale Evaluation Study," *Eval. Rev.*, vol. 9, no. 5, pp. 627–643, Oct. 1985.
- [7] R. K. Yin, *Case Study Research: Design and Methods*, 2nd Revised edition edition. Thousand Oaks: SAGE Publications, Inc, 1994.
- [8] N. K. Denzin, *Sociological Methods: A Sourcebook*. New Brunswick, N.J.: Transaction Publishers, 2006.
- [9] Makana Municipality, "IDP & Municipal Statistics," *Makana Municipality*, 2016. [Online]. Available: <http://www.makana.gov.za/statutory-documents/idp-municipal-statistics/>. [Accessed: 07-Dec-2016].
- [10] Statistics SA, "Poverty Trends in South Africa: An examination of absolute poverty between 2006 and 2011.," Statistics SA, Johannesburg, South Africa, 2012.
- [11] Statistics SA, "Census 2011," 2012.
- [12] M. T. Mogale, "Local governance and poverty reduction in South Africa II the role of micro-finance," *Prog. Dev. Stud.*, vol. 7, no. 4, pp. 345–355, Oct. 2007.

- [13] P. Bardhan and D. Mookherjee, "Corruption and Decentralization of Infrastructure Delivery in Developing Countries," Robinson Rojas Archive, Online, 2000.
- [14] D. L. Altheide, "An Ecology of Communication:," *Sociol. Q.*, vol. 35, no. 4, pp. 665–683, Nov. 1994.
- [15] J. Tacchi, "Studying Communicative Ecologies: An Ethnographic Approach to Information and Communication Technologies (ICTs)," presented at the 56th Annual Conference of the International Communication Association, Dresden, Germany, 2006.
- [16] N. Memarovic, M. Langheinrich, E. Rubegni, A. David, and I. Elhart, "Designing 'Interacting Places' for a Student Community Using a Communicative Ecology Approach," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, New York, NY, USA, 2012, p. 43:1–43:10.
- [17] G. M. Broad, S. J. Ball-Rokeach, K. Ognyanova, B. Stokes, T. Picasso, and G. Villanueva, "Understanding Communication Ecologies to Bridge Communication Research and Community Action," *J. Appl. Commun. Res.*, vol. 41, no. 4, pp. 325–345, Nov. 2013.
- [18] G. Hearn, N. Collie, P. Lyle, J. H.-J. Choi, and M. Foth, "Using communicative ecology theory to scope the emerging role of social media in the evolution of urban food systems," *Futures*, vol. 62, Part B, pp. 202–212, Oct. 2014.
- [19] D. Mercea, L. Iannelli, and B. D. Loader, "Protest communication ecologies," *Inf. Commun. Soc.*, vol. 19, no. 3, pp. 279–289, Mar. 2016.
- [20] M. T. Bastos, D. Mercea, and A. Charpentier, "Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media," *J. Commun.*, vol. 65, no. 2, pp. 320–350, Apr. 2015.
- [21] W. L. Bennett, A. Segerberg, and S. Walker, "Organization in the crowd: peer production in large-scale networked protests," *Inf. Commun. Soc.*, vol. 17, no. 2, pp. 232–260, Feb. 2014.
- [22] A. Ceron, L. Curini, and S. M. Iacus, "Using Sentiment Analysis to Monitor Electoral Campaigns Method Matters—Evidence From the United States and Italy," *Soc. Sci. Comput. Rev.*, vol. 33, no. 1, pp. 3–20, Feb. 2015.
- [23] - Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." *Foundations and trends in information retrieval* 2.1-2 (2008): 1-135.
- [24] - Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.
- [25] - Hofmann, Thomas. "Probabilistic latent semantic indexing." *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999.
- [26] - Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3.Jan (2003): 993-1022.