Women empowerment and gender based-violence in sub-Saharan Africa: a comparison between scientific research and social views from Twitter

Introduction

The latest Africa Gender equality index has revealed that in sub-Saharan Africa, there is a great barrier between women and men rights. As the backbone of the family unit, African 'women rights, freedom, and economic potential are still limited. Sub-Saharan Africa societies are ruled by a patriarchal system (Buzawa, 2017), male domination and gender stereotypes that have led women to be obedient to men; to be exposed to many injustices such as land ownership, female genital mutilation, rape, participation in the labour market and gender-based violence.

Rooted in gender inequalities between men and female, gender-based violence is a crucial, visible issue cutting across all ethnic, background and drastically affecting the community and the economy in this century. The 2017 United Report revealed that worldwide, 1 in 3 women have physically and or sexually been abused in her lifetime with 36% of the African population being affected by intimate partner violence (McCloskey *et al.*,2016). Many studies have revealed that women unemployment and financial dependence are among common risk factors of gender-based violence (Devries *et al.,* 2013; Jekwes *et al.*,2017). Congruently, economic empowerment is regarded as a key factor to reduce gender-based violence. Considered a basic human right, gender inequality is the 5th goal in the agenda of the 2030 Sustainable development Goals who have been looking for various ways in bringing positive changes to this plight.

From quantitative, qualitative, mixed methods to social networking, massive researches are done to understand, propose policies to curb gender-based violence. During these last decades, social media has become a powerful means to communicate and play a major rule in people lives in sub-Saharan Africa. The eMarketer in 2013 statistics revealed that 18% of the web users come from Twitter with more than 500 million tweets per day (Forbes, 2013). Moreover, from political, land repatriation or ownership, reproductive health to gender issues, trending topics are evolving in sub-Saharan. In a continent like Africa, where freedom of expression is a myth, from Facebook, Instagram, Snapchat, YouTube to Twitter, people are using various platforms to engage and share their opinion and as an agent of change in the community (Vardhan, 2016). Twitter can be used as a powerful channel to raise gender inequality issues and to mobilize national and international public attention. In that respect, this study used messages from a created profile on twitter to see how women's empowerment and gender-based violence are perceived by twitters users.

Though the literature on gender-based violence, there are few pieces of research that has both used scientific articles and twitter stream to that compare women' perception of gender-based violence in sub-Saharan Africa. This research attempts to address manly two goals. The first goal aim to determine the positive and negative sentiments from tweets about gender-based violence and women's empowerment. The second goal compare women perception of empowerment between tweets and scientific research.

Methodology

Data collection

This section provides an overview of the methodology used to produce the results of both the sentiment analysis and topic modelling. Firstly, a systematic search of articles related to gender-based violence and women's empowerment in sub-Saharan Africa were searched in English, scrutinized and extracted in

PubMed between 2010 and 2018. We used the key words such as : women, gender, empowerment, marriage, violence, abuse, sexual harassment, rape, forced marriage, Female genital mutilation, sub-Saharan Africa , etc. We systematically reviewed 29088 articles.

Secondly, we used Twitter search feature to collect the English tweets and retweets on feelings and concerns on women's empowerment and gender-based violence in 2018. Using the public Twitter REST search API, over 17980 tweets in English were gathered between January and December 2018. The keywords, as listed below, are passed as parameter arguments to the Twitter search API for the Tweets. These keywords are all variants of words used to describe gender-based violence and women's empowerment: women, gender, empowerment, violence, women status, physical violence, sexual violence, emotional violence, rape, partner.

Methods

To help reduce the influence of noisy data on results, all tweets were passed through a number of custom filtering and data transformation techniques to standardise the terms and symbols used across all tweets. After collecting the tweets, we first created a corpus and we made tweets cleaning and preprocessing techniques by using tm package in R programming language. We made all letters lower-case in the corpus and we removed the numbers, punctuation/symbols and extra white spaces. Afterwards, we have performed some preprocessing techniques such as Tokenization, Stopword removing, and Stemming.

To visualize the results, word clouds were created, based on the frequencies of the words used in all retrieved tweets and abstracts to determine sizes of words. The words in the cloud are located in accordance with their frequency. We've done the sentiment analysis by creating a dictionary of terms associated with women's empowerment and created the sentiment score. Finally, we ran topic models with Latent Dirichlet allocation (LDA). Latent Dirichlet allocation is a particularly popular method for fitting a topic model and to uncover hidden topics among different documents. It's a generative, unsupervised statistical model that allows a set of text observations to be clustered into groups and it treats each document as a mixture of topics, and each topic as a mixture of words (Silge and Robinson, 2019).

Preliminary results

Word clouds for tweets and abstracts from scientific research

Figures 1 and 2 show respectively the frequencies of the most words which appear in the tweets and scientific research. In the tweets word cloud, we can see that the word "women" has a large size. We can also see the words like femal, love, people, youth, woman, equal, economic, education, etc. This shows that the most tweets relating on women's empowerment talk about these words. Regarding the scientific research word cloud, we notice some words which have large size such as: health, prevalence, risk, patient, violence, etc.

Fig. 1. The most common words in tweets

Fig.2 The most common words in the scientific research
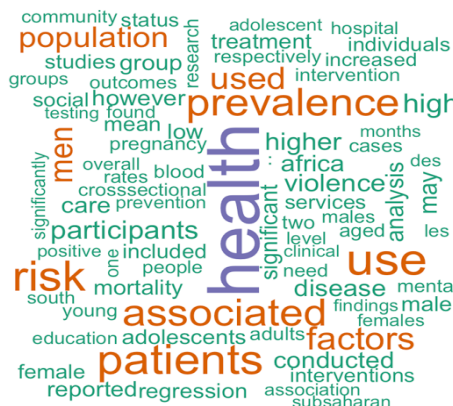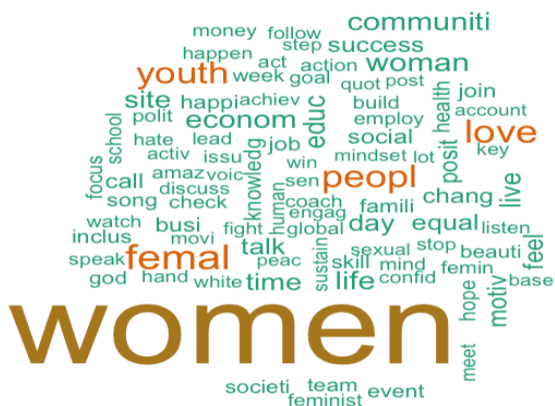
Sentiment analysis from tweets

Figures 3 and 4 present respectively a break-down of the percentages of feelings (positive and negative) and the most common words (top 10) that contribute to each of the feelings about women's autonomy and gender-based violence in tweets. Figure 1 clearly, shows that 57% of positive feelings were expressed in the tweets compared to 43% for negative feelings. Figure 3 shows the most expressed words for positive feeling are: love with the greatest contribution, then success, skill, free, strong, right, respect, lead, proud and freedom. For negative sentiment, there is no significant difference in the contribution of the first 10 most common words.

Fig 3: Sentiment distribution in tweets   Fig 4: Contribution of words to tweets sentiment
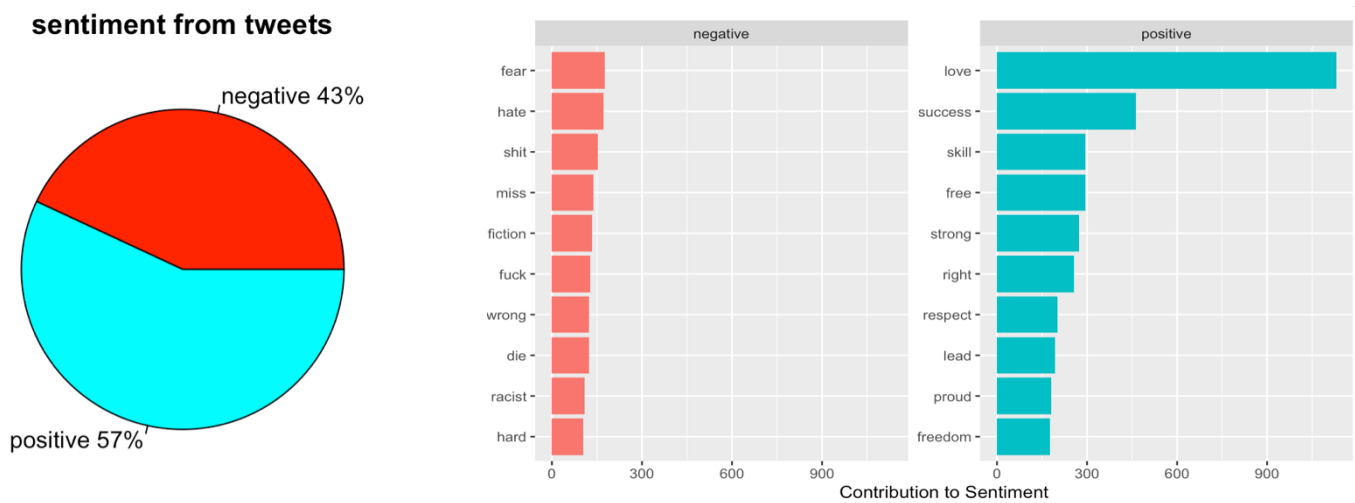
Topic Modeling from tweets

Figure 5 shows the top terms for 5 topics. The document-term matrix (dtm) has 17980 documents (rows) and 59563 terms (columns) if we do not remove stopwords and 52401 terms if we do remove stopwords. In this figure, we see some clear patterns for some of the topics, while for others we recognize some repetition. For each topic, we label them based on the top words per topic, as well as the top comments per topic. We notice that the most common words in topic 1 include "femal", "youth", "education" and "equal". Those most common in topic 2 include "love" "woman", "people", "feel" and "site".

Figure 5: LDA Top terms that are most common within each topic from tweets

3
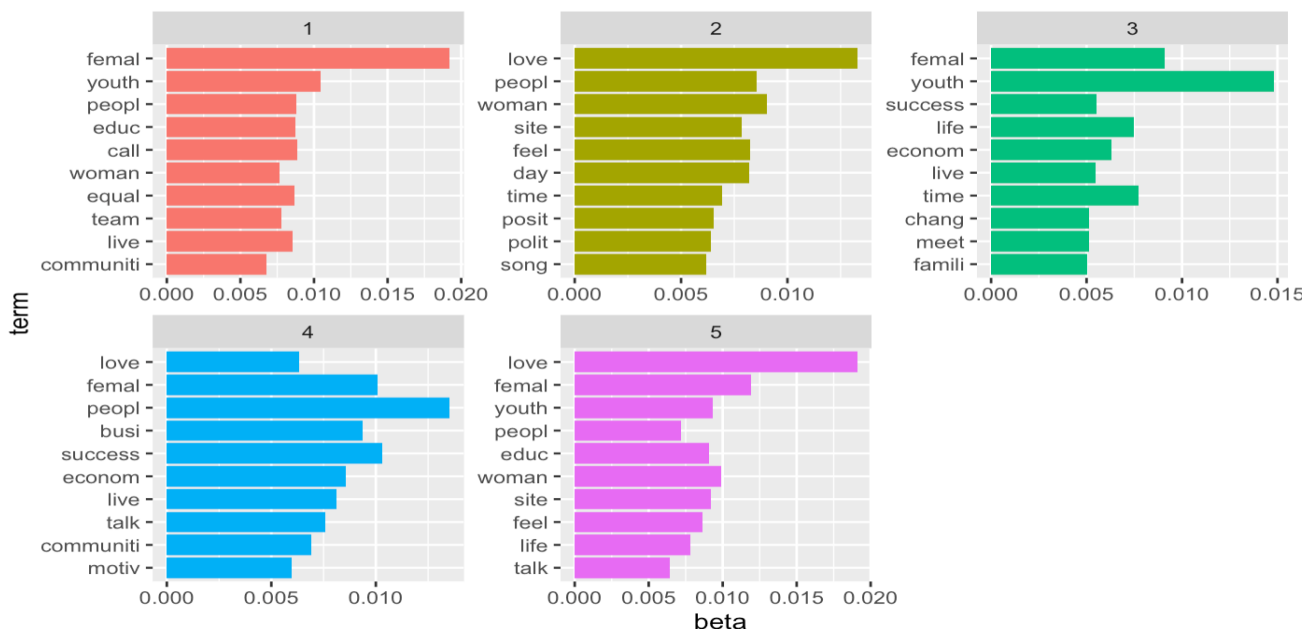
Conclusion and future work

Using sentiment analysis techniques and topic model the goal of this study was to explore different perception of tweet content relating to the women empowerment and the gender-based violence. The sentiment categories of negative emotion and positive emotion were analysed. This study show that Twitter can be used as a powerful channel to raise and to study online sentiment of issues of women's empowerment. Despite the preliminary results obtained in this study, there is much potential and several key next steps in terms of methodologies that we would like to consider for future work. First, we plan to do the same analysis with French key words. We also plan to calculate and generate the distribution of the average score of the feelings expressed in the tweets by month and by country or region of residence both in English and French. We would like to continue the analysis on topic model by increasing the number of topics. In addition to LDA, we also plan to use the Structural Topic Model (STM) to see if there are similar findings between the two techniques.

Regarding the articles from scientific research, we plan to get the distribution of articles by the year of publication, by the keywords, by the journal name, by the country and by the author name and see how the study of empowerment and gender-based violence have evolved over time. We would like to to get the correlation between some words from abstracts and "gender", "violence" and "empowerment" and compare women perception of empowerment between tweets and scientific research. We would like to consider all of these possibilities in future work and look forward to the expansion of the data, methodologies, and discussion.

References:

Buzawa, E.S. and Buzawa, C.G. (2017). Global Responses to Domestic Violence. Springer, p.318

Devries, K.M., Mak, J.Y., Garcia-Moreno, C., Petzold, M., Child, J.C., Falder, G., Lim, S., Bacchus, L.J., Engell, R.E., Rosenfeld, L. and Pallitto, C. (2013). The global prevalence of intimate partner violence against women. Science, 340(6140):1527-1528.

eMarketer. (2013). Twitter use rises across US age groups-eMarketer. Retrieved fromhttp://www.emarketer.com/Article/Twitter-Use-Rises-Across-US-Age- Groups/1010119

Forbes. (2013). Can Twitter save TV? (and can TV save Twitter?). Retrieved from http://www.forbes.comsitesjeffbercovicican-twitter-save-tv-and-can-tv-save-twitter

Jewkes, R., Fulu, E., Naved, R.T., Chirwa, E., Dunkle, K., Haardörfer, R. and Garcia-Moreno, C. (2017). Women's and men's reports of past-year prevalence of intimate partner violence and rape and women's risk factors for intimate partner violence: A multicountry cross-sectional study in Asia and the Pacific. PLoS medicine, 14(9), p.e1002381.

McCloskey, L.A., Boonzaier, F., Steinbrenner, S.Y. and Hunter, T. (2016). Determinants of intimate partner violence in sub-Saharan Africa: a review of prevention and intervention programs. Partner abuse, 7(3): 277-315.

Ranjay V. (2016). Social media and women empowerment: a sociological analysis. EPRA International Journal of Economic and Business Review, 5(8): 117-121

Silge J. and Robinson D. (2019). Text Mining with R, https://www.tidytextmining.com/