Towards an automated journalism framework for social data monitoring

A platform that assists journalists in newsrooms in real time and enables them to easily obtain and monitor their desired **newsworthy content** from the mass volume of unverified content from Twitter platform.

Bahareh Fatemi
University of Bergen
Bahareh.fatemi@uib.no

Introduction

- Twitter, the service for microblogging has grown to be an extremely popular medium for sharing ideas, reporting news, and communicating with friends.
- Several hundred microblogs (tweets), or real-time comments, are frequently made about events as they happen.
- Twitter is intriguing not only because to this real-time response, but also because it is sometimes ahead of the newswire.

Motivation

- **Automated journalism** according to [1] is "the auto generation of journalistic stories through software and algorithms, without any human input".
- Finding newsworthy news/events for journalists is a very challenging or even impossible task to do, since it requires inspecting millions of Tweets per day and eventually using a limited percentage of their findings in their story. Thus, automating news exploration in a way that increases the efficiency of their probing is a crucial task we are going to cover in this work.
- In journalism application, beyond newsworthiness, time aspect is an asset, thus finding the breaking news/ events before other news agencies is an important factor in an automated journalism system.
- In this research, in collaboration with MediaFuture SFI and Schibsted Media Group we are developing a platform that can assist journalists in newsrooms in real time and enables them to easily obtain and monitor their desired newsworthy content from the mass volume of unverified content from Twitter platform.

Prior works

- AI techniques have been been applied for analysing social media data but many of them do not function in real time.
- Secondly, most of prior works, either focus on collecting, filtering, and analysing tweets using predefined metrics [2] (such as number of replies, likes, etc.) or are only focused on analysing tweets' content [3][4].

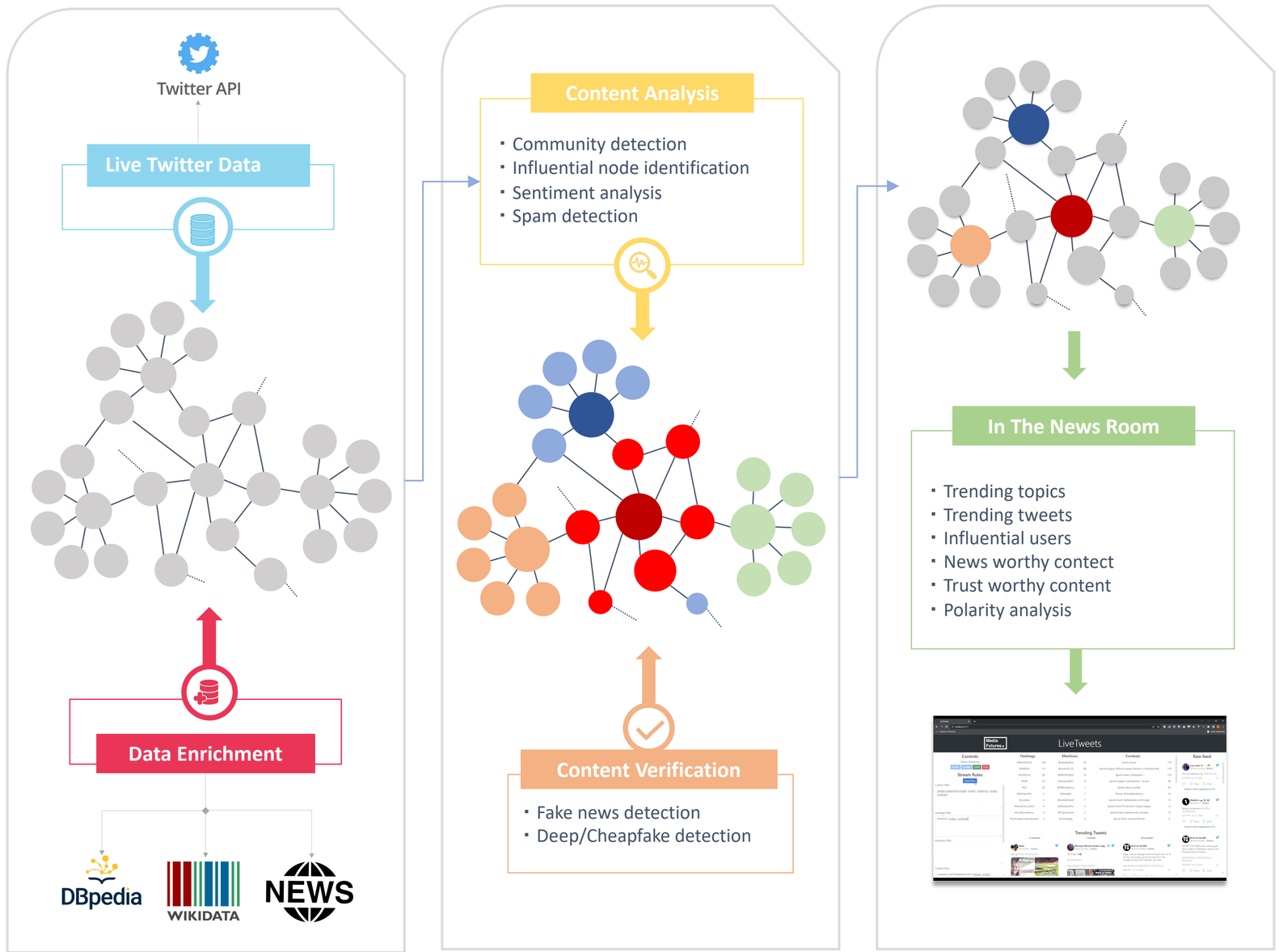
Method

We present our own visual analytical framework that is not only based on information retrieval from Twitter but also enriched by machine learning and network science. In this work, we intend to use state of the art techniques such as:

- Community detection
- Influential node identification
- Event detection
- First Story detection
- Semantic enrichment
- Fake news detection
- Engagement tracking

Demo

We currently are at the early stages of our research, but we have been developing a demo of our proposed system called "Livetweet" which you can see in figure 1. Initial prototype has been implemented by Daniel Rosnes, a research assistant at university of Bergen.



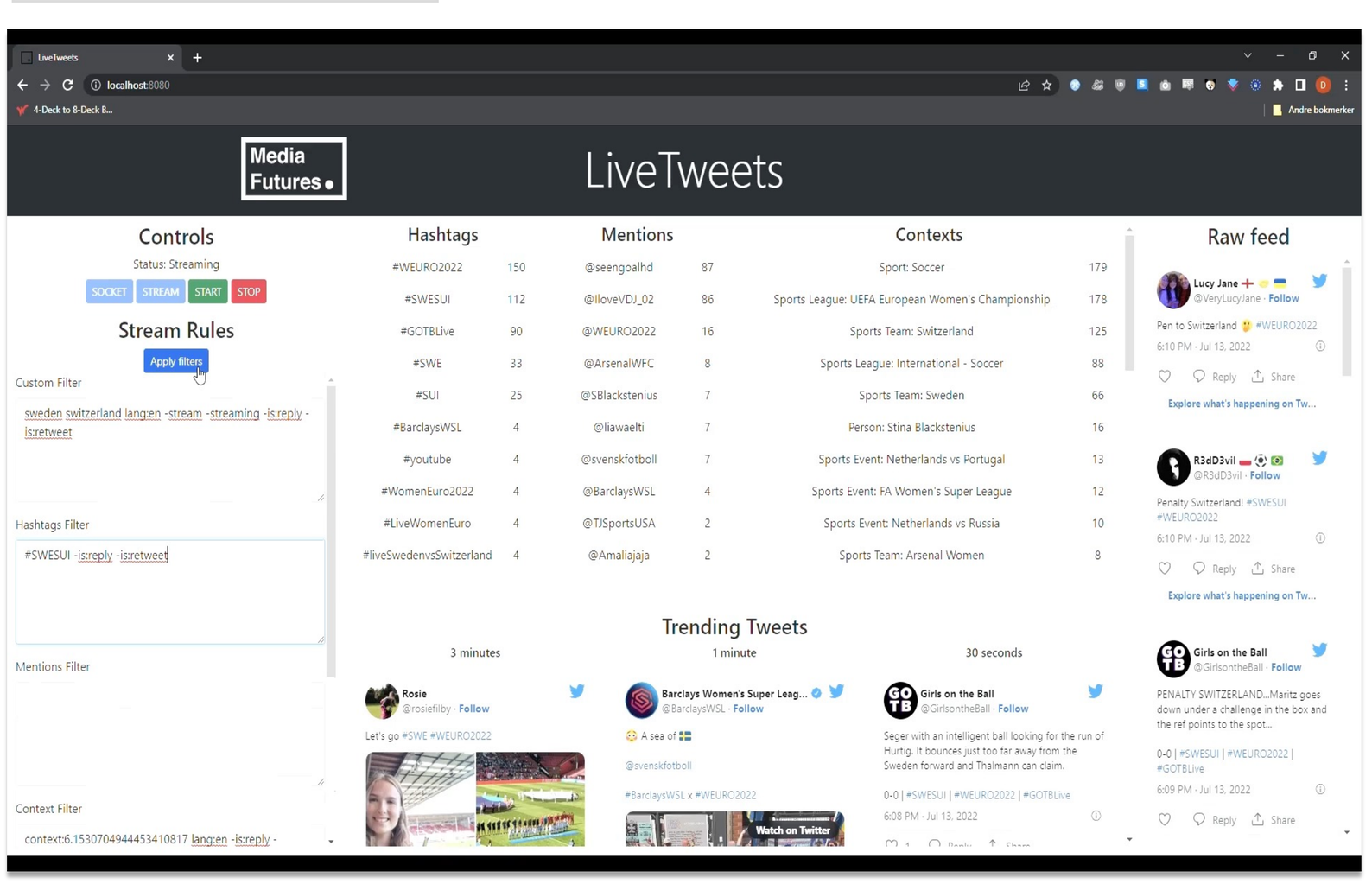


Fig 1. A screenshot of our prototype

REFERENCES

[1] S. Danzon-Chambaud,(2021)"A systematic review of automated journalism scholarship: guidelines and suggestions for future research," Open Res. Eur., vol. 1, p. 4.

[2] B. Yadranjiaghdam, S. Yasrobi, and N. Tabrizi,(2017), "Developing a Real-Time Data Analytics Framework for Twitter Streaming Data," Proc. - 2017 IEEE 6th Int. Congr. Big Data, BigData Congr. 2017, pp. 329–336, 2017.

[3] C. Brantner and J. Pfeffer,(2018), "Content analysis of twitter: Big data, big studies," Routledge Handb. Dev. Digit. Journal. Stud., pp. 79–92.

[4] S. Gaglio, G. Lo Re, and M. Morana,(2016), "A framework for real-time Twitter data analysis," Comput. Commun., vol. 73, pp. 236–242.

UNIVERSITY OF BERGEN

ACKNOWLEDGEMENTS

This work was supported by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the centers for Research-based Innovation scheme, project number 309339.