Title

Identify Twitter Communities Interested in African Affairs

Objective

There are three objectives of this week's project.

The first objective is to create github pages for your github 10 Academy folder and write your first Blog. There will be a tutorial on this object which will guide you through the process.

The second objective is to submit your week1 report as a Medium blog post.

The third objective is to build on top of your week1 project such that you use twitter search and stream API to collect users data, and identify major user communities that are interested in your country economical, social, cultural, and public health affairs.

The guide given in the expected outcome and methods sections as well as the papers listed in the reference section will give you a starting point along this third objective.

Expected Outcome

The following are the expected outcomes - all for your country and immediate neighbouring countries:

1. A list of top 1000 users that tweet, retweet, like, and comment on the following areas
    a. Economy
    b. Social values (sport, education, human rights, etc.)
    c. Cultural (entertainment, fashion, art, etc)
    d. Public health
2. Time series analysis of the topic and volume of these separate communities.
3. The interaction between these communities.

Method

You are required to compile active twitter users' handles for your country. You can start doing so by collecting first the twitter accounts of

- Top government officials including ministries and deputy ministers
- Active journalists and their media network
- Active ambassadors
- Active celebrities

You can then obtain the twitter accounts of the followers of these twitter users and create a large network of people who, we assume, represent the main communities interested in your countries issue.

By downloading ~500 tweets from all the twitter users you managed to collect, perform the following task

1. Label each tweet as **economy, social, cultural, health**
2. Each user is assigned a coordinate value in the (**economy, social, cultural, health)** space. For example a user with 200 economy tweets, 150 health tweets, 150 social tweets, and 100 cultural tweets will have a coordinate value of (200, 150, 100, 150).
3. Perform k-means clustering to find the cluster of communities in this 4-dimensional data.
4. Extract insights about each of the clusters of communities - seasonality, twitter relationship among and across communities, profession, and big global events (covid19 epidemic, holidays, seasons).
5. Write a three pages report on your findings.

References

1. Community Discovery in Twitter Based on User Interests
2. Identifying Topical Twitter Communities via User List Aggregation
3. Community Discovery in Twitter Based on User Interests
4. A social network analysis of Twitter: Mapping the digital humanities community
5. Improving Twitter Community Detection through Contextual Sentiment Analysis of Tweets