Social Network Analysis

By Albert Okiri Olweny

Use Case: World Cup

Introduction

The growth of social media platforms in conveying the sentiments of the market presents an opportunity for businesses and social initiatives to make better-informed and accurate decisions. Social networks unearth opinions from relevant communities, predominant notions and popular entities within them. This document details the use of graph analysis of tweets from the Twitter API endpoint. The dataset used in the analysis is retrieved based on keywords or communities of interest.

From contemporary trends and events, *The World Cup* is chosen as an ideal use case to illustrate the effectiveness and value of this social network analysis. The data was collected on 24^{th} October 2022 and features 12000 tweets within a span of the previous 7 days. The grouped (*world cup*) keywords search for a combination of the two words sequentially in a tweet.

Analysis

Graph Analysis

A table documents the centrality analysis for in-degree and out-degree measurements for tweet reply, quote, mention and retweet. These insights identify the most influential nodes that can be focused on depending on the interaction types. As pictured in Figure 1, a user can select the interaction type to analyze the centrality measures and network visualization. Figure 2 showcases the network visualization that permits a close examination of the network relationship between various nodes depending on interaction types.

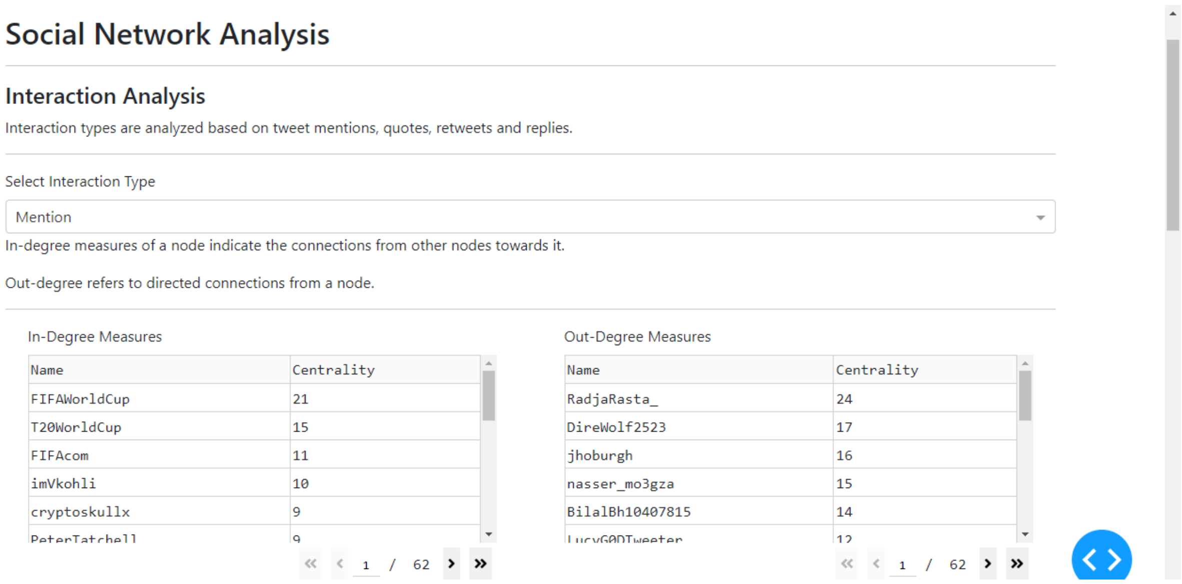

Figure 1The user interface that features interaction type selection and tables of centrality measures.

Figure 2The user interface of interactive network visualization and selection of layout of the graph.

Topic Modelling

Topic modelling from Latent Dirichlet Allocation (LDA) algorithm reveals the salient discussions and issues that are being discussed within a given conversational network. Consistency in topics creates an accurate perspective on the audience's sentiments. These topics are presented as a sorted list of 15 grouped topics.

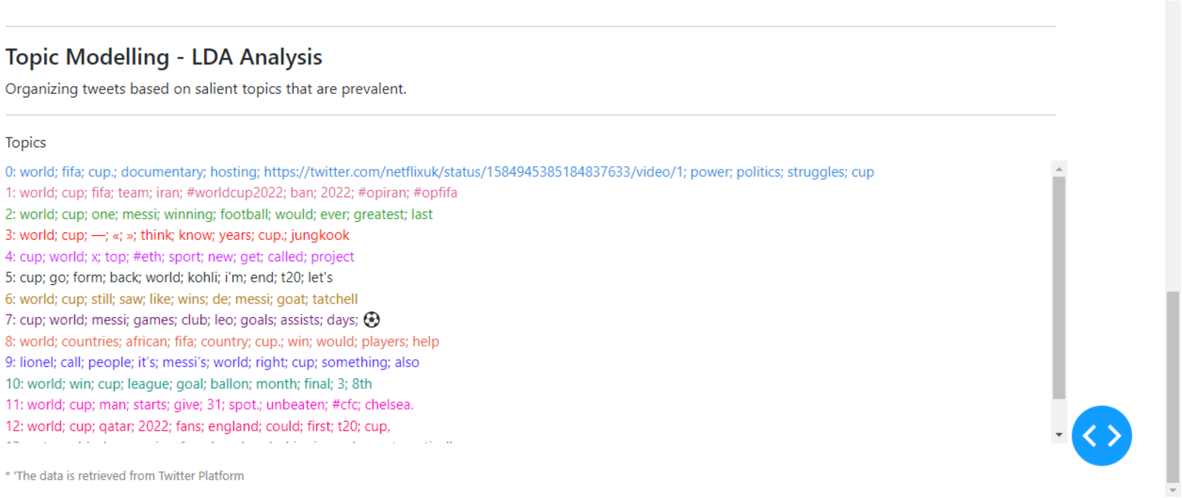

Figure 3The list documenting the salient topics from LDA analysis of retrieved tweets.

Synthesis of Reported Data

The centrality measures depicted in the tables when combined with the network visualization reveal the most influential nodes depending on the interaction types. These can inform the marketing approach to use by identifying specific users and how they are fundamental to the networks that are of interest.

For instance, *RadjaRasta_* has the highest out-degree value for the *mention* interaction type. This indicates the user's engagement by mentioning the keywords used in retrieving the tweets. A closer examination of the network visualization in Figure 4. shows a subset of the network that is not the giant component of the community. This can be relevant in tapping into the various sub-communities that are present. Focusing on users who are prevalent on the largest network, specific sub-networks, or those with the highest centrality in all sub-networks depending on the incentives can be realized.

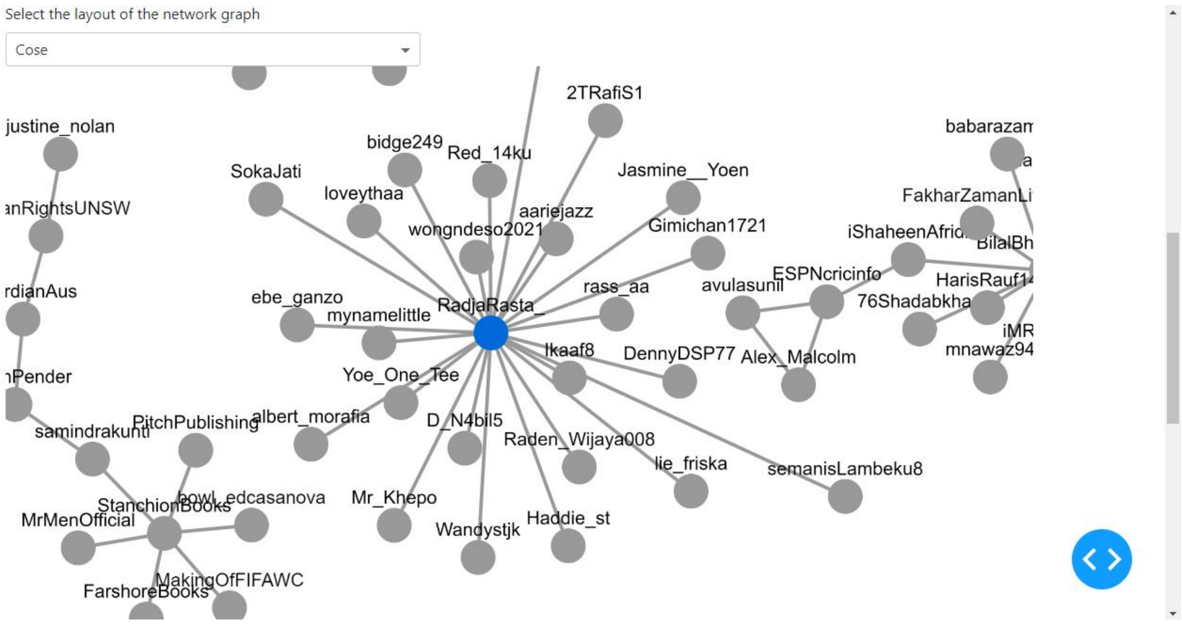

Figure 4Interactive sub-network visualization.