Twitter Trends Analysis

Omar Elnemr , Mohamed hesham , Mahmoud nady,
ziad morsy , Youssef abbas , kareem eltayeb
Egypt Japan university for science and technology
Computer science engineering department

Abstract - *the main idea of our research is to find the common patterns between twitter users depending on the latest trends / hashtags they contributed in , this includes scrapping twitter data and analyzing the results*

I. Introdcution

To get the common patterns between Twitter users we had to get this data first, so we scraped Twitter users, we got 20,000 samples of data to do our research on these sample are Collceted From Over 100 Trends/Hashtags Related To Egyption Culture in General.
The objective of this paper is mainly to analyze the data but this can't be done with the data itself, in analyzing the data to get the common patterns we had to use different kinds of analysis such as - Path analysis, Centrality analysis, Connected components analysis, Clustering coefficients., Density analysis, Network type, Connected components analysis, Community discovery, Dynamic community discovery.

II. The process

2.1 Data Collection: We collected Twitter data from Egyptian Trends and Hashtags and the People How Tweeted in these Trends/Hashtags.

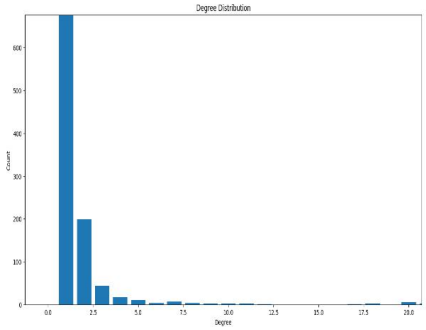
2.2 Data Preprocessing: Data preprocessing involved cleaning the data, removing duplicates, and ensuring the dataset's integrity. We extracted relevant features such as user IDs, hashtags, and interaction types.

2.3 Network Construction: For constructing the network, we treated each unique hashtag as a node. Users who interacted with these hashtags formed the connected nodes. We created edges between the hashtag nodes and the corresponding user nodes to represent the interactions.

III. Network Analysis and Results

3.1 Degree Distribution Analysis:

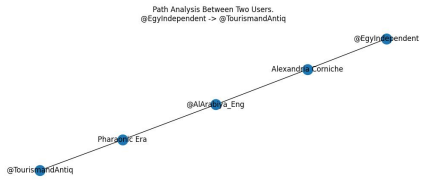
We analyzed the degree distribution of the hashtag network to understand the distribution of interactions among hashtags. We plotted the degree distribution histogram and examined the presence of highly connected hashtags or hubs.
The degree distribution analysis revealed a power-law distribution for the hashtag network, indicating the presence of highly connected hashtags or popular topics within the Egyptian Twitter verse. And this is the graph we got and we think that is normal because not too much people tweet in more than 5 hashtags.



3.2 Path Analysis:

Path analysis allowed us to measure the shortest paths between Two hashtags/Trends, providing insights into the flow of interactions and information diffusion within the network. Path analysis demonstrated the existence of efficient information flow and interaction pathways between different hashtags, facilitating the spread of ideas and discussions within the network.
in Our Netwrok Graph, Path Analysis Provide a Shortest Link Between Two People According to The Trends That They Both Tweeted in if Exist for Example, The Path Between the Two Nodes @EgyIndependent and @TourismandAntiq will be LIke:

@EgyIndependent -> Alexandria Corniche -> @AlArabiya_Eng -> Pharaonic Era -> @TourismandAntiq



3.3 Centrality Analysis:

Centrality analysis is a method used in network analysis to identify the most important or influential nodes in a network. In a network, centrality refers to the degree to which a node is central or important in relation to the other nodes in the network.

There are several different types of centrality measures that can be used to analyze a network, including:

1. Degree centrality: This measures the number of connections or edges that a node has in a network. Nodes with a high degree centrality are often seen as important because they are well-connected to other nodes.

2. Betweenness centrality: This measures the extent to which a node lies on the shortest paths between other pairs of nodes in the network. Nodes with high betweenness centrality are often seen as important because they act as bridges between different parts of the network.

3. Closeness centrality: This measures the degree to which a node is close to all other nodes in the network. Nodes with high closeness centrality are often seen as important because they can quickly communicate with other nodes in the network.

Centrality analysis can be used in a variety of fields, including social network analysis, transportation planning, and epidemiology, among others. By identifying the most central nodes in a network, researchers can gain insights into the structure and dynamics of the network, and how information or disease might spread through it.

Data Samples:

Node @Afrok_travel has degree centrality of 0.00010044698910150168, betweenness centrality of 2.0000152936721923e-05, closeness centrality of 0.20865862762169962, and eigenvector centrality of 0.0017097960937464674

-

Node @NEARProtocol has degree centrality of 0.0003515644618552559, betweenness centrality of 0.021191675522922122, closeness centrality of 0.22880027136944392, and eigenvector centrality of 0.009011023671302883

-

Node @Magy_wagdy12 has degree centrality of 5.022349455075084e-05, betweenness centrality of 0.0, closeness centrality of 0.20842622629337276, and eigenvector centrality of 0.001564095504841696

3.4 Connected Components Analysis:

The connected components analysis identified several distinct subnetworks within the hashtag network, representing different topical communities or discussions among Egyptian Twitter users. And this is an example of a subnetwork : {'@neokyklos', '@wdclibraries', '@NounoursProduc2', '@hqndmaidens', '@womptron', '@igototokyo', '@alexlibraryva', '@AlexandriaLib', '@lasacerdotessa', '@aurionvik', '@PoobisOK', 'Alexandria Library'}

3.5 Clustering Coefficients:

Clustering coefficients were calculated to measure the degree to which hashtags and users tended to cluster together, indicating the presence of cohesive communities within the network.

for our data The Resulted Cluster Coefficient is 2.432533860961312e-06

3.6 Density Analysis:

The density analysis revealed the extent of interconnectedness between hashtags and users, highlighting the overall engagement and interaction patterns within the Egyptian Twitterverse. And our graph density is : 0.00010218753210702746, the density value is extremely small, indicating that there are very few edges in the graph relative to the number of possible edges.

3.7 Network Type:

Based on the network characteristics and analysis results, we classified the network into specific types, such as scale-free, small-world, or random networks.Based on our analysis, the network appears to be a small-world network. In a small-world network, the average path length between nodes is short, and there is a high level of clustering, meaning nodes tend to have connections within their local neighborhood.

3.8 Community Discovery:

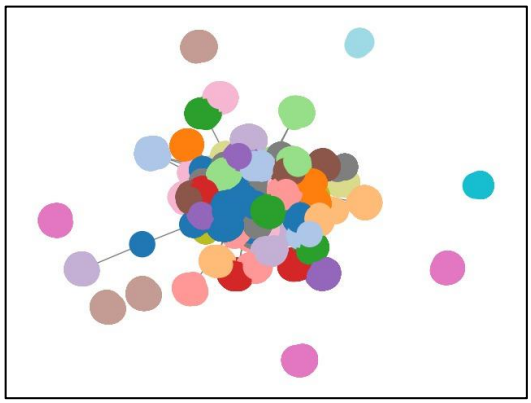
Community detection algorithms, such as the Louvain modularity optimization, were employed to identify cohesive groups of hashtags and users with high intra-group interaction.

for Exmaple the Community Dicovry for some nodes:

Community @OmarRamyFadi: 39

Community @IsmailEmad: 26

Community @HamzaAdham: 44



3.9 Dynamic Community Discovery:

To explore the dynamics of the network, we performed community discovery analysis over multiple time intervals, detecting changes in community structures and interactions.
Dynamic community detection has applications in a variety of fields, including social network analysis, neuroscience, and ecology, among others. By identifying communities that change over time, researchers can gain insights into the underlying processes that drive network evolution and help to understand how different nodes or groups of nodes interact with one another over time.
Data sample :["@essamalimahmoud", "@bbblanchett","@suckrfc","@cakeblanchett_","@blanchettcom","@tashajagger_","@S3leh29","@yousry_log24","@stillSHErises","@TanyaWahan","@soegypte","@MukMuk___","@Abelgiep","@hana khaled5","@claudinecassar","@TipsterFrosty","@ZSCOfficial_EN","@aswan8177","@UNICEFUSA","@Foot_Masry","@Samalentips","@aswan_jay","@abujastreets","@The_PharaohEG","@DamSpaces", "@WFP_MENA","@Psychaitrist69","@AswanSC_", "@ms_crean","@Rosa63456459", "@yemisizealtv1","@Fiery_Mike","@matigary","@abu_aliyah","@__Balsamk","@Hailunsa".... Etc.]

VI. Conclusion

This research paper presented a comprehensive network analysis of Egyptian users on Twitter, focusing on hashtag interactions. The analysis provided insights into the structural properties, dynamics, and community structures within the Egyptian Twitterverse. The findings contribute to our understanding of online social networks and their implications for information diffusion and community formation in the context of Egypt.

V. Our Dataset and Codes:

All our Dataset and Code Could Be Found in this Link ‘Github Repo’, Each Analysis is Done Seperatly in a Separate File named after the Analysis, all These Files is in The Repo

https://github.com/omarahmedelnemr/AlgorethmGraphAnalysis