

Social Network Analysis for Web-based Community

Ghaith Ekhaa Majeed, Dr.Abdulkareem M. Radhi

Al-Nahrain University, Baghdad, Iraq

ghaithit90@gmail.com, kareem_m_radhi@yahoo.com

Abstract : Social Network Analysis (SNA) is now a key technique in recent social analysis. This paper focuses on having a Facebook crawler system to extract data. It has a complete support for non-Latin languages. The proposed application is a platform independent and can be accessible from different devices. It is built using JavaScript, HTML5, PHP, CSS, SVG and MySQL. New crawler designed and implemented from scratch by using JavaScript to provide a powerful and suitable crawler for the project purpose and insert data directly to proposed SNA system database for analysing it later using SNA metrics to obtain the SNA report.

Keywords: Social Network Analysis, SNA, Crawler, Facebook analysis, SVG.

I. Introduction

SNA is being a necessary tool for media, inquiries and students which needs to display social relationships regarding to network theory consisting of individuals and relations in terms of Graph-based structures which is very complex, it's operate on multiple levels from individuals up to organizations and take an key A. technique in solving snags and accomplishing goals. Visual representations are very useful to realize network information and extract result of analysis. SNA play a dominant role as a significant procedure in latest sociology, communication, economics, information science and different studies. The importance of social network analysis came from its difference from traditional social studies, that assume it is the attributes of individuals whether they are friendly or not, etc. In addition, to B. test organizations interact with each other as well as connections between employees [1]. Social networks are growing on the web day after day by their size and number therefore social networking becoming the biggest inclinations on the web with millions of people on it, and hundreds of web-based social C. networks like Facebook, twitter, Instagram, etc. (10).

II. Proposed System

The proposed system have multi step implementation from web design to final output. The following sections presents system steps in more details with its block diagram in figure-1:

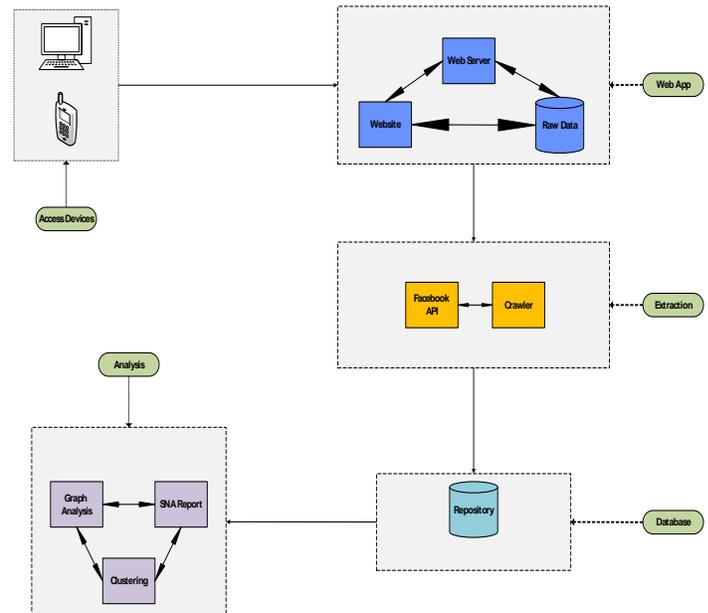


Figure-1 Proposed SNA system block diagram.

A. Website Design

First, build SNA application that meet the modern design requirements to work fast and smooth, by designing a web pages using Adobe Photoshop CS6. Then moving the output templates to Adobe Dreamweaver to get the complete HTML template files. At this point, CSS and the modern features of HTML5 used to complete the modern design of this application. Application was designed to work properly on different screen sizes of different aspect ratios.

B. Crawler design and implementation

Proposed Crawler extracted Facebook data then insert it directly to proposed system database for analyzing it later. The proposed crawler designed and implemented to diagnose extracted non-Latin languages so that it became more suitable with Arabic language to analyze Arabic Facebook data.

C. Database Design

SNA System requires a well-structured database to store the desired information to accomplish its jobs. This information includes information about the Facebook page posts, comments and like counts. The database type is MySQLi, and the language used to connect this database with SNA application is PHP. Database in SNA application implemented compatible with non-Latin languages (Arabic).

D. SNA Metrics

SNA Metrics is the core of SNA and according to these metrics, the analyzer can identify which node is the central node or in other words, which nodes are the most influential

nodes in the community. SNA metrics used in the proposed system are Degree, Betweenness and Closeness. Metrics results designed and implemented in the proposed system to display it in the final report of the resultant graph and shows the most popular Page, User, Post and Comment of the searched data. The proposed system contains added services to the SNA report, which is the execution time for each metric in order to view the time consumption for the whole analysis operation, the summation for each metric and the counter to calculate the total nodes in the graph.

E. Extract Data Periodically

Data set used in the system obtained through different time periods from different Facebook pages to build a database with wide variety range for different types of analysis.

System Architecture

Figure-2 shows the complete system architecture for the proposed web application. It shows all its webpages that are involved in the application. SNA application has two types of users; each type authorized to access specific pages and allowed to do particular jobs. Analyzers access the pages that allow them to analyze the data online which shows in analyzer GUI in figure-2 that have access to Facebook data analysis subsystem to generate analysis graph and report through the use of VivaGraph library and SNA metrics (Betweenness, Degree and Closeness). On the other hand, System admins access the whole system from user registration, data extraction to data analysis, therefore; it has access the database, edit its data and access the main pages, which control the application that forming the core module of the system.

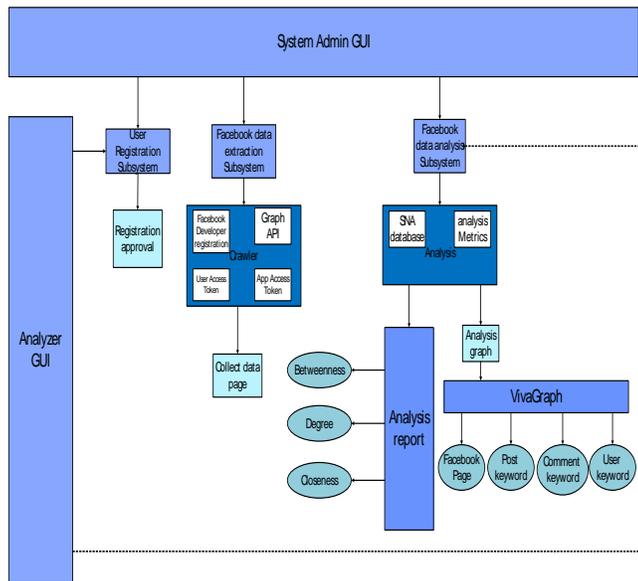


Figure-2 System architecture.

III. Results and Analysis

The challenge facing any social network analysis is the analysis of Arabian communities & Arabic language, therefore this give more contribution to the project & build a complete report to the whole Iraqi community rather than specific small community.

The challenges here is how to analyze broad Iraqi people so multiple steps would do this:

A- Choose the top ranking Iraqi pages, the verified pages and the more popular pages to be crawled through different times [Posts from 2012-11-28 22:25:43 to 2017-03-20 16:44:06].

B- Accumulation of Big Data by crawling the mentioned pages through many periods and gathering 118,011 Comments and 4,489 Posts from 157 Iraqi Pages to build a sample of Iraqi community.

C- Iraqi community classified to 17 different category about human needs (Company, Tourism and places, Music, Sport, Books & Education, Cars, Technology, TV News & Cinema, Health, Food, Religion, Political, Photographic, Government, Shopping/Retail & Game, Financial and Others).

D- Searching for 471 keywords for all categories that should build a huge variety database of human needs.

E- Get analysis report and graph from the proposed system for each search state then store the results in excel sheets to build the final big report. The results stored for each metric of Betweenness, Closeness and Degree then take the summation and maximum value of each metric from the graph of each keyword.

F- Draw a chart for each metric to represent the whole report of all categories, which shows the most interested trends for Iraqi community as shown in figures-3 through 8.

Sounds good from the following analysis charts that the Iraqi community cares about political news at the first grade and superior to all other interesting categories cause the current situation of our country. The other categories varies from Food, Sport to Photographic and Music at the lowest level. This shows the difference between SNA and traditional Social Statistic Analysis because “Kadim Al Sahir, Barbie” Pages represent the highest ranking pages in Iraq as pages like count equal to “12973401, 8349944” respectively but SNA shows the highest trends in Iraq is the Political news interactivity.

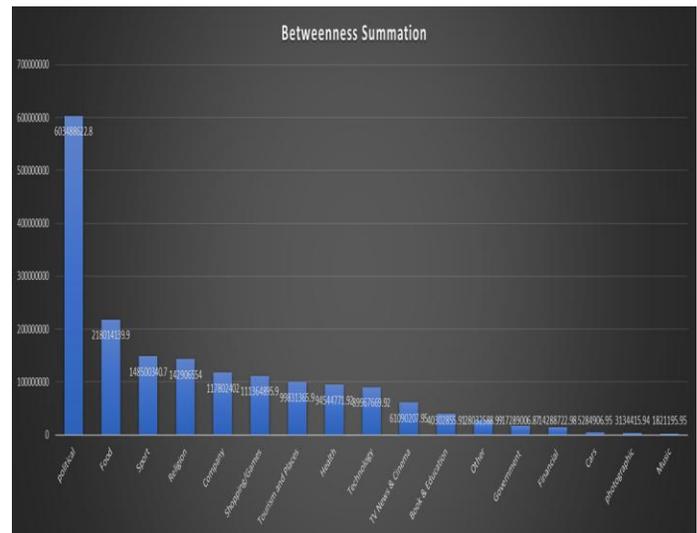


Figure-3 Betweenness summation of all categories.

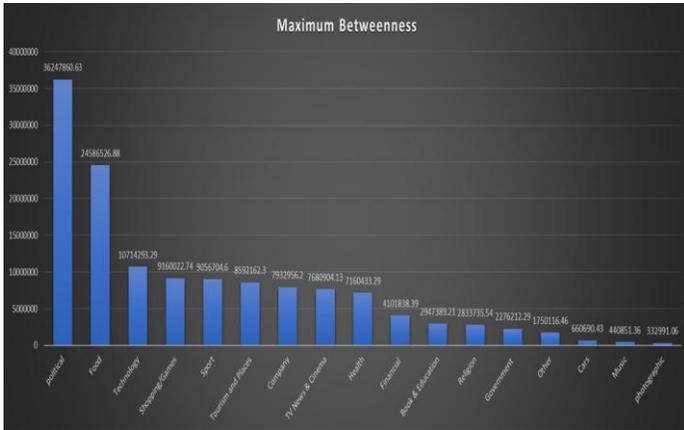


Figure-4 Maximum Betweenness of all categories.

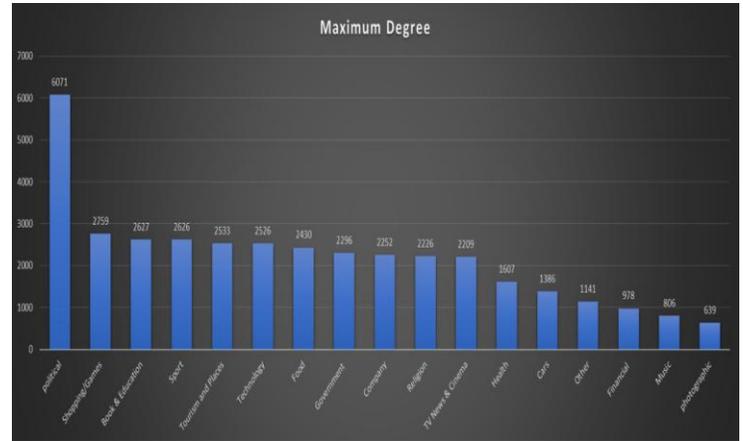


Figure-8 Maximum Degree of all categories.

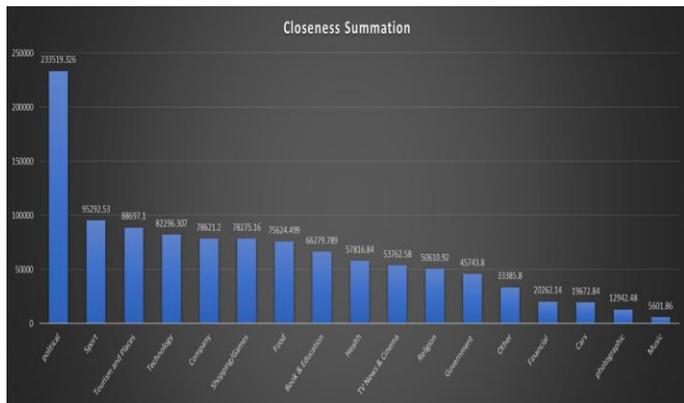


Figure-5 Closeness summation of all categories.

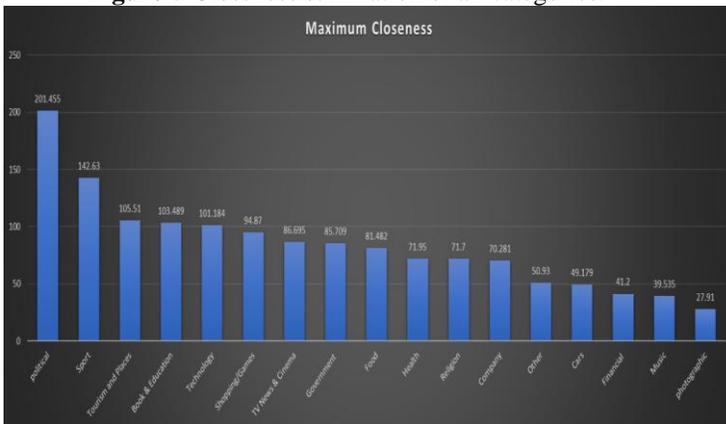


Figure-6 Maximum Closeness of all categories.

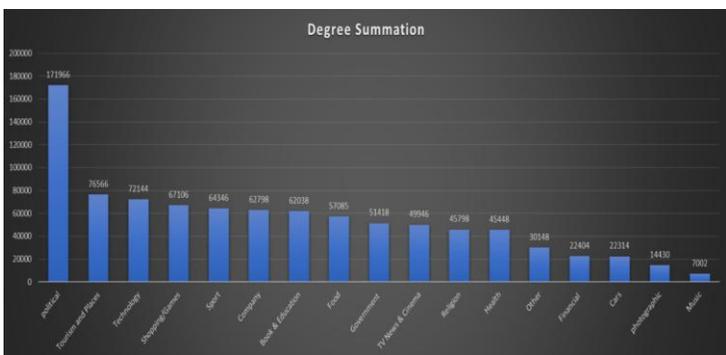


Figure-7 Degree summation of all categories.

Users Behavior Analysis
One of the methods for behavior analysis in social media is to search about individuals speak. SNA system support words search for specific person in both his posts and comments of Facebook pages to classify his behavior according to society categories. Figure-9 below shows samples of people behavior in social media (Facebook) one per each category:

User	Most Frequently Words	User Behavior Category
Majid	Iraq - champions - ISIS - Forces	political
Dhiaa	Kabab - taste	Food
Haitham Al Zuhairi	activate - balance - tech.	Technology
Alzubaidy Bareeq	price - expensive	Shopping/Games
Ali Rassul	result - round - game	Sport
Mohamad Hamoodly	scene - beautiful - heritage	Tourism and places
Prince Sad	Huawei - company	Company
Bohair	series - TV program	TV News & Cinema
Wahiba Yazid	salon - beauty - cures	Health
Miumi Omer Ali	finance - funds - loan - installment	Financial
Mohamed EL Ghouzi	students - research - postgraduate	Book & Education
Balquees Alkhaledy	Allah - paradise	Religion
Iraqi Aseel	president - state - council	Government
Samer Salah	car - buyer	Cars
Waleed Abo Baker	art - song	Music
Dounia Hassan	photo - shot	photographic

Figure-9 Users categories on social media.

Search
System users can search about any intended keyword written in the posts and comments of the Facebook pages as shown in figure-10 for searching on “companies” keyword.

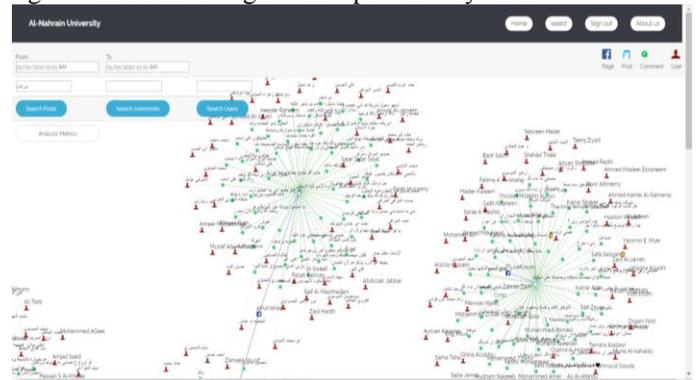


Figure-10 Searching for posts keyword.

SNA Reports

These reports are the core of SNA system and they implemented in the proposed system to be generated with any search criteria like searching in posts, comments or users. This reports use many graph analysis metrics to get its results like Degree and from these reports, analyzers can get knowledge about which entity is the centralized in the graph as shown in figure-11 in which the report shows the rows: execution time of each metric, total nodes of the graph and summation of metrics respectively.

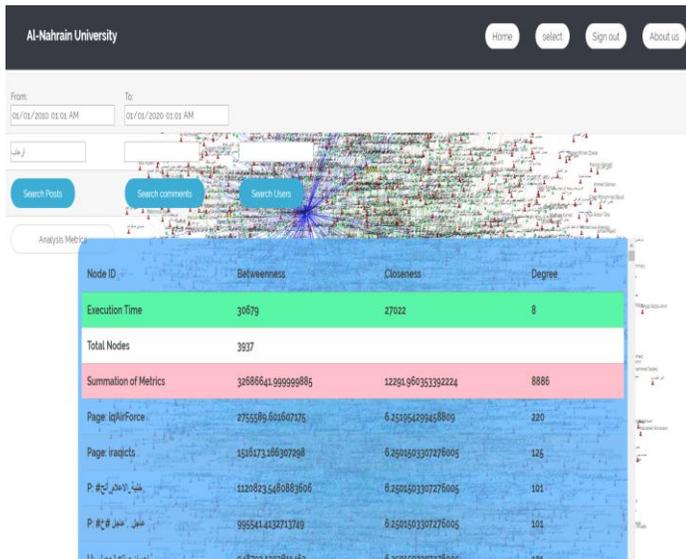


Figure-11 SNA report.

IV. Conclusion

This paper proposed a Facebook network analysis system to analyze datasets crawled from Facebook online database by proposed crawler subsystem. The proposed system is an integrated web system for crawling data from online Facebook database and store it offline in its well-designed and suitable database then analyze these data using SNA metrics rather than using separated different systems to do this job. The proposed crawler extract and insert data directly to a suitable database without intervention from end user, and it fully supporting non-Latin languages (Arabic). Using SNA metrics to analyze big data by its nodes and relations rather than traditional social statistics.

References

- i- N. Ghali, A. Abraham, "Computational social networks: Mining and visualization", *Comput. Soc. Networks Min. Vis.*, vol. 9781447140, pp. 1–385, 2012.
- ii- E. Ferrara, "Mining and Analysis of Online Social Networks", *Univ. Messina, PhD thesis*, p. 192, February 2012.
- iii- J. Golbeck, "Statistics.Com - Introduction to Social Network Analysis (SNA)", 2 Apr. 2017.
- iv- P. Singer, "Time Series Analysis of Online Social Network Data and Content", *Master's Thesis Graz University of Technology*, 26 September, 2011.
- v- P. Kazienko, R. Michalski, S. Palus, "Social Network Analysis as a Tool for Improving Enterprise Architecture", *International KES Symposium on Agents and Multi-agent Systems, Manchester UK*, 50-370 Wroclaw, Poland, 29 June – 1 July 2011.