2014 Fourth International Conference on Communication Systems and Network Technologies

Social Network Analysis Tools

Nadeem Akhtar
Department of Computer Engineering, Zakir Husain College of Engineering & Technology
Aligarh Muslim University
Aligarh, India
nadeemakhtar@zhcet.ac.in

***Abstract*. Social networks like Facebook, Twitter, and Google+ are most visited domains on the Internet. They contain huge data about the users and the relationships among them. To analyze and mine useful information from these huge social network data, special graph based mining tools are required that can easily model the structure of the social networks. A number of such analysis tools are available with their own features and benefits. Choosing an appropriate tool for a particular task is difficult to decide. This work present a comparative analysis of four social network analysis tools- Networkx, Gephi, Pajek, IGraph based on platform, execution time, Graph types, algorithms complexity, input file format and graph features.** *(Abstract)*

Keywords: social networks analysis; SNA tools; Gephi; Pajek; Networkx; IGraph (keywords)

I. INTRODUCTION

A Social network is defined as a network of relationships or interactions, where the nodes consist of people or actor, and the edges or archs consist of the relationships or interactions between these actors [1]. Social networks and the techniques to analyse them existed since decades [2]. There can be several type of social networks like email network, telephone network, collaboration network. But recently online social networks like Facebook, Twitter, LinkedIn, MySpace etc have been developed which gained popularity within very short amount of time and gathered large number of users. Facebook is said to have more than 500 million users in 2010 [3].

The field of social networks and their analysis has evolved from graph theory, statistics and sociology and it is used in several other fields like information science, business application, communication, economy etc. Analysing a social network is similar to the analysis of a graph because social networks form the topology of a graph. Graph analysis tools have been there for decades. But they are not designed for analysing a social network graph which has complex properties. An online social network graph may be very large. It may contain millions of nodes and edges. Social networks are dynamic i.e. there is continuous evolution and expansion. A node in social network usually has several attributes. There are small and large communities within the social graph. Old graph analysis tools are not designed to manage such large and complex social network graph.

In this paper, some graph analysis tools for the analysis of large online social networks are discussed and compared. We have considered four tools namely Gephi, Networkx, IGraph and Pajek. Such comparative study of social network analysis tools has already been done earlier [4][5]. We have added more comparative results concentrating on efficiency, visualization and graph features.

II. SOCIAL NETWORK ANALYSIS

Social network analysis (SNA) is the methodical analysis of social networks. Social network analysis views social relationships in terms of network theory, consisting of nodes (representing individual actors within the network) and ties (which represent relationships between the individuals, such as friendship, kinship, organizational position, sexual relationships, etc.) [6].

Analysis tasks of social networks includes following:

- Discovering the structure of social network
- Finding various attribute values for the network- Ex. radius, diameter, centrality, betweenness, shortest paths, density etc
- Finding communities in the social network
- Visualizing the whole or part of the social network

Several works has been done on various social networks to analyse and discover various kinds of relationships and information [7][8][9][10].

A. Kinds of Network Analysis

There are two basic kinds of social network analysis, ego network analysis, and complete network analysis.

Ego network analysis is concerned with analysis of individual nodes. A network can have as many egos as nodes in the graph. Egos can be persons, organizations or whole society. In ego network analysis, individual behaviour and its variation is mined and described.

Complete network analysis is concerned with the analysis of all the relationships among a set of nodes. Techniques such as subgroup analysis, equivalence analysis and measures like centrality (closeness, degree, and betweenness) all require complete networks [11].

978-1-4799-3070-8/14 $31.00 © 2014 IEEE
DOI 10.1109/CSNT.2014.83

IEEE computer society

B. Applications of Social Network Analysis

Following are some applications of social network analysis: [12][13]

- Identify new scientific trends becoming commercially viable, e.g. RFID, Genome sequencing, tissue engineering
- Analyse expert network, Co-authorship networks, co-citation networks, patent networks
- Measurement of success
- Ranking of trends, of authors, of companies commercializing trend
- Analysing page importance Page Rank (Related to recursive in-degree computation), Authorities/Hubs
- Discovering Communities: Finding near-cliques
- Analysing Trust: Propagating Trust
- Using propagated trust to fight spam: In Email, In Web page ranking

III. Social Network Analysis Tools

Social network analysis tools are used to identify, analyze, visualize or simulate nodes (organizations, or knowledge) and edges (relationship or interaction) from various types of input data including mathematical models of social networks. There are several tools available for analysis of social networks. The International Network for Social Network Analysis (INSNA) maintains a large list of software packages and libraries [14].

We have selected four analysis tools for comparison namely Gephi, Networkx, IGraph and Pajek. The selection is based on several facts. All four selected software are freely available for use and they can handle large graph size. Network analysis tools are either GUI based or packages/libraries which can be used in a programming language. Gephi and Pajek are GUI based network tools whereas Networkx and IGraph are packages based tools.

Following are the brief detail about each of these four tools-

Networkx is a Python language software package for the creation, manipulation and the study of structure and functions of the complex networks. With this tool you can load and store networks in standard data formats, can generate many types of random and classic networks, analyze network structure, build network models, draw networks, and much more. Networkx has many features like language data structures for graphs, dIGraphs, and multIGraphs. Nodes can be "anything" (e.g. text, images), Edges can hold arbitrary data (e.g. weights, time-series), Standard graph algorithms, Network structure and analysis measures etc [15].

Gephi is an interactive visualization and exploration platform for all kinds of networks, dynamic and hierarchical graphs. Runs on Windows, Linux and Mac OS X. Gephi are a tool for people that have to explore and understand graphs. Like Photoshop but for data, the user interacts with the representation; manipulate the structures, shapes and colours to reveal hidden properties [16].

Pajek: A widely used Software for drawing networks, Pajek also has analytical capabilities, and can be used to compute most centrality measures, identify structural holes, block-model, and so on[17].

IGraph is a free software package for creating and manipulating graphs. It includes implementations for classic graph theory problems like minimum spanning trees and network flow, and also implements algorithms like community structure search. The efficient implementation of IGraph allows it to handle graphs with millions of nodes and edges. IGraph can be installed as libraries for C, R, Python and Ruby [18]. Above four tools are compared on the following six criterion- platform, Graph types, algorithm time complexity, graph layout, graph input file format, graph features

Dataset for SNA tools comparisons: We have used Slashdot dataset [19]. It contains 77317 nodes and 982787 edges (Directed). Slashdot is a technology-related news website features user-submitted and evaluated news stories about science and technology related topics.

IV. Social Network Software Tools Comparisons

A. Comparison Based on Platform

Social network analysis tools Pajek and Gephi are stand alone software, Networkx and IGraph are libraries. Pajek/gephi runs on windows platforms and Networkx use python library and IGraph use python/r/c library for social network analysis. Networkx, IGraph or Pajek can handle more than one million nodes and Gephi can handle up to 150000 nodes.

TABLE I General Comparison of Tools

Software	NETWORKX	IGRAPH	GEPHI	PAJEK
TYPE	LIBRARY	LIBRARY	STAND ALONE	STAND ALONE
PLATFORM	PYTHON	PYTHON\R\C LIBRARY	WINDOWS	WINDOWS
COMPUTATIONAL TIME	FAST	FAST	FAST	MEDIUM
NO. OF NODES	1 MILLION	1 MILLION	0.15 MILLION	1 MILLION

B. Comparison Based on Network Types

In social network Analysis there are four types of network Graph [6].

In a one-mode network, each vertex can be related to each other vertex. In one mode network we have only one set of nodes and ties are connected to these nodes. In a two-mode network, vertices are divided into two sets and vertices can only be related to vertices in the other set.

Two mode network Graph are a particular type of networks with two sets of nodes and ties are only established between nodes belonging to different sets. Techniques for analysing one-mode networks cannot always be applied to two-mode networks without modification or change of meaning. Special techniques for two-mode networks are very complicated. We can create two one-mode networks from a two-mode network

In Multi relational network there will be multiple kinds of relations between nodes. Nodes may be closely-linked in one relational network, but distant in another.

In temporal networks (dynamic graphs) networks can change over time. The lines and vertices in a temporal network should satisfy the consistency condition: if a line is active in time t then also its end-vertices are active in time t.

For one mode or two mode network analysis we can use any of software tools but for multi relational network graph we have only Pajek software tools .for temporarily network graph we have Networkx and Pajek tools.

TABLE II NETWORK TYPES SUPPORTED BY TOOLS

Graph type	Networkx	IGraph	Gephi	Pajek
1-Mode network	Yes	Yes	Yes	Yes
2-Mode network Graph	Yes	Yes	Yes	Yes
Multirelational network Graph	No	No	No	Yes
Temporarily network Graph	Yes	No	No	Yes

C. Comparison Based on Graph Layout

In social network analysis we have many layout algorithms. IGraph or Pajek have most famous and recent layout algorithms Fruchterman Reingold or Kamanda Kawai. All of them software have circular or spring layout. Gephi provide user friendly layout capability for user. Gephi provide capability like Photoshop where users have many additional facilities. Force layout is another famous layout algorithm given by Gephi software.

For user friendly visualization we can use Gephi software but it cannot handle large or complex graphs. For large and complex network we can use either Networkx or IGraph. IGraph or Networkx provide support for many other tools for visualization. For small dataset we can use standalone software (Gephi, Pajek) because we can handle standalone software easily.

TABLE III GRAPH LAYOUT SUPPORTED BY TOOLS

Layout	Networkx	IGraph	Pajek	Gephi
Circular layout	Yes	Yes	Yes	Yes
Random layout	Yes	Yes	Yes	No
Spectral layout	Yes	No	No	No
Spring layout	Yes	Yes	Ye s	Yes
Graphviz layout	Yes	No	No	No
Kamanda kawai	No	Yes	Yes	No
Fruchterman reingold	No	Yes	Yes	No
Force Atlas layout	No	No	Yes	No

D. Comparison Based on Algorithm Time Complexity, Input File Formats and Graph Features

Social network analysis software has many algorithms for graph features. We compare these tools for each feature based on algorithms complexity.

IGraph and Networkx have algorithms for maximum number of features. Based on algorithms complexity we can say that IGraph is more useful software compare to other softwares. IGraph provide efficient algorithms for page rank, all types of centrality, density, MST and shortest path.

Comparisons among the four social networks on the basis of time complexity, input file formats and graph features are shown in Table IV, V and VI respectively.

TABLE IV TIME COMPLEXITIES

FEATURES	NETWORKX	IGRAPH	GEPHI	PAJEK
ISOMORPHISM	$O(n^2)$	EXP	NA	NA
CORE m=no. of lines	O(M)	O(M)	O(M)	O(M)
CLIQUES	O(|V|/(log)2)	O(3|V|/3)		O(N)
SHORTEST PATH	O(|V|.|E|)	O(|V|+|E|)	O(|V|+|E|)	O(|V|+|E|)
CLUSTRING	O(V)	NA	O(V)	NA
ALL SIMPLE PATH	O(|V|+|E|)	O(|V|+|E|)	NA	NA
CLOSENESS CENTRALITY	O(n.|E|)	O(n.|E|)	NA	NA
DENSITY	$O(n^3)$	O(1)	NA	NA
MST	NA	O(|V|+|E|)	NA	NA

CYCLES	O((|V|+|E|).c+1)	NA	NA	NA
PAGERANK	NA	O(|E|)	O(E)	NA
BETWEENNESS	NA	O(|V|.|E|)	NA	NA
EIGENVECTOR	NA	O(|V|+|E|)	NA	NA

TABLE V FILE FORMATS SUPPORTED BY TOOLS

INPUT FILE FORMAT	NETWORKX	IGRAPH	GEPHI	PAJEK
.NET FORMAT	YES	YES	YES	YES
.GML	YES	YES	YES	NO
.GRAPHGML	YES	YES	YES	NO
EDGESLIST(.TXT)	YES	YES	NO	NO
EDGESLIST(.CSV)	Yes	Yes	No	No
.DOT	YES	YES	YES	NO
.PAJEK	NO	YES	NO	YES
.DAT	NO	NO	NO	YES
ADJACENCY LIST	YES	YES	NO	NO
.GDF	No	No	Yes	No
GRAPH DB	No	Yes	Yes	No

TABLE VI VARIOUS GRAPH FEATURES SUPPORTED BY TOOLS

GRAPH FEATURE	NETWORKX	IGRAPH	PAJEK	GEPHI
APPROXIMATION	YES	NO	NO	NO
ASSORTIVITY	YES	YES	NO	NO
CENTRALITY	YES	YES	YES	YES
NETWORK DIAMETER	YES	YES	YES	YES
CLUSTRING	YES	YES	YES	YES
FLOW	YES	YES	YES	NO
COMMUNITIES	YES	YES	NO	YES
COHESION	NO	YES	NO	NO
BLOCK MODELING	YES	NO	YES	NO
DENDROGAM	YES	YES	YES	NO
CLIQUE	YES	YES	YES	NO
PAGE RANK	YES	YES	YES	YES
BFS	YES	YES	YES	NO
DFS	YES	YES	YES	NO
HITS	YES	YES	YES	YES
DENSITY	YES	YES	YES	YES
CORE	YES	YES	YES	NO
ISOMORPHISM	YES	YES	YES	NO
PARTITION	NO	NO	YES	NO
POWERLAW	YES	YES	NO	NO
MST	YES	YES	YES	NO
BI PARTILE	YES	YES	YES	NO
BRIDGE	YES	YES	YES	NO
DYAD	YES	YES	YES	NO
HITS	YES	NO	NO	YES

TABLE VII EXECUTION TIME FOR VARIOUS FEATURES

SNA Features	Networkx	IGraph	Gephi	Pajek
Load time	54.67 sec.	3.707 sec	29 sec	3 sec
Degree centrality	58.57 sec	6.199 sec	4 sec	2 sec
Graph degree	60.87 sec	6.22 sec	4 sec	2 sec
Page rank	120.78 sec	9.81 sec	10 sec	No

Hits	57.23 sec	15.43	8 sec	No
Cliques	66.98 sec	9.35 sec	Na	No
Density	58.94 sec	3.302 sec	4 sec	No
Modularity	81 .4 sec	9 sec	30 sec	6 sec
Network diameter	35 sec	3.51 sec	120 sec	No
Core	65.84 sec	6.532 sec	No	1 sec
Cohesion	No	8.943 sec	No	No
Clustering coefficient	3303.99 sec	1800 sec	1200 sec	108 sec
Hub	76.57	5.831 sec	3 sec	No
Authority	Array is to big	6.783 sec	3 sec	No

V. CONCLUSION

Stand alone software is very useful for graph Visualization (up to a maximum of few thousands of nodes), data format conversion. IGraph is fastest tools that provide most of graph features and handle large and complex network. Libraries (Networkx or IGraph) are more useful for tasks involving millions of nodes and for operations such as the union and the difference between sets of nodes or for the clustering. Stand alone software are easy to use and easy to learn so for beginner Pajek and Gephi is suitable software. For complex dataset and research purpose we can use Networkx and IGraph software.

For one mode or two mode network analysis we can use any of software tools but for multi-relational network graph, we have only Pajek software tools. For temporarily network graph we have Networkx and Pajek tools.

All of the software can handle .Net file format. But mostly data are present in .txt format. We can easily understand or handle .txt format or many websites provides data set in .txt format so we can use IGraph or Networkx software tools. We have many file format conversion software they can convert .txt file format into .Net file format but these software can handle only small size file. So for large size data set we use IGraph or Networkx software tools.

IGraph provides mostly graph features and it also handle large and complex network. All of them software can compute centrality, clustering coefficient, network diameter, page rank, density**.** But if we want to compute some specific feature we choose different software. If we want to compute Cohesion, we can Use Networkx or IGraph tools. If we want to compute Bridge and dyad, we can use Networkx, IGraph and Pajek software tools. Gephi does not provide the facilities for dyad or bridge computation.

IGraph and Pajek are faster software Tools compare to others. But Pajek does not provide all graph features. So if we want to analyse all graph features we can use IGraph software. IGraph gives fastest result to almost all graph features. Load time is minimum for Pajek software. Execution time for centrality, page rank, graph degree and cliques is minimum in IGraph software. So based on execution time IGraph is better software.

REFERENCES

[1] Measurement and Analysis of Online Social Networks by Alan Mislove, Massimiliano Marcon,Krishna P. Gummadi, Max Planck Institute for Software Systems

[2] For a historical overview of the development of social network analysis, see: Carrington, Peter J. & Scott, John (2011). "Introduction". The Sage Handbook of Social Network Analysis. SAGE. p. 1. ISBN 978-1-84787-395-8.

[3] http://www.facebook.com/notes/facebook/500millionstories/409753352130

[4] David Combe, Christine Largeron, El″od Egyed-Zsigmond and Mathias Géry, "A comparative study of social network analysis tools", International Workshop on Web Intelligence and Virtual Enterprises 2 (2010)

[5] M A huisman, MAJ van duijn, " Software for social network analysis", pages 270-316, 2004

[6] Graph and Network Analysis Dr. Derek Greene Clique Research Cluster, University College Dublin, Web Science Doctoral Summer School 2011

[7] Monclar, Rafael Studart, et al. "Using social networks analysis for collaboration and team formation identification." Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on. IEEE, 2011.

[8] Nadeem Akhtar, Hira Javed, Geetanjali Sengar, "Analysis of Facebook Social Network", IEEE International Conference on Computational Intelligence and Computer Networks (CICN), 27-29 September, 2013, Mathura, India

[9] Zelenkauskaite, Asta, et al. "Interconnectedness of complex systems of internet of things through social network analysis for disaster management." Intelligent Networking and Collaborative Systems (INCoS), 2012 4th International Conference on. IEEE, 2012.

[10] Li, Jianfeng, Yan Chen, and Yan Lin. "Research on traffic layout based on social network analysis." Education Technology and Computer (ICETC), 2010 2nd International Conference on. Vol. 1. IEEE, 2010.

[11] Online Social Networks: Measurement, Analysis, and Applications to Distributed Information Systems by Alan E. Mislove, Houston, Texas

[12] Social Networks Overview: Current Trends and Research Challenges" November 2010 Coordinated by the —NextMEDIA CSA.

[13] Business Application of Social Network Analysis BASNA-2013 www.basna.in

[14] International network of Social Network Analysis INSNA www.insna.org

[15] Networkx http://Networkx.lanl.gov/index.html

[16] Gephi https://gephi.org

[17] Pajek vlado.fmf.uni-lj.si/pub/networks/pajek

[18] IGraph IGraph.sourceforge.net

[19] snap.stanford.edu/data

View publication stats