

The review of social networks analysis tools

Nasrin JOKAR¹, Ali Reza HONARVAR², Khadijeh ESFANDIARI³, Shima AGHAMIRZADEH⁴

Department of Electrical and Computer Engineering, Safashahr Branch, Islamic Azad University, Safashahr, Iran

Department of Electrical and Computer Engineering, Safashahr Branch, Islamic Azad University, Safashahr, Iran

Department of Electrical and Computer Engineering, Safashahr Branch, Islamic Azad University, Safashahr, Iran

Department of Electrical and Computer Engineering, Safashahr Branch, Islamic Azad University, Safashahr, Iran

Abstract

Social networks can include anything ranging from family, friends, classes, objects and other similar cases, important and effective members, members of exception, the formation of such networks can be discovered by using the relationships between the members of the network which are important to business and research works, to achieve these cases, social networks should be analyzed using special tools. Social network analysis tools generally includes two packaged based on graphical user interfaces (GUIs) and packages made for programming / scripting. These tools are powerful and extensible and are able to analyze big data networks and visualize networks, isolated or central data and other important data can be simply discovered by data visualization. In this paper, some of the most important tools of social network analysis are presented and compared according to some their capabilities.

Keywords: social network analysis, Social network analysis tools, analysis, graphical user interface, visualization

1. Introduction

Social networks are everywhere which can be formed of anything like the family, the project team, classrooms, sports teams, membership in some sites like Facebook, Twitter etc.

¹ Adresse email: nasrin.jokar@gmail.com

² Adresse email: alireza_honarvar@shirazu.ac.ir

³ Adresse email: khadijaesfandiar@gmail.com

⁴ Adresse email: sh.ghamirzadeh2014@gmail.com

The nodes are created by entities and groups of its member and dependencies between these entities constitute the edges between nodes. The network is getting more complex by more nodes and edges between them [5].

Social networking is something that the most of people know about it, but they do not know about the analysis of social networks, which is important in many areas. Examples of social networks include: Biologists who study the formation of protein interactions of a network, Criminologists and law enforcement agencies analyze crime networks, Epidemiologists study the relationships between individuals, Zoologists study the animal behaviors in the network and researchers of telecommunications analyze contact networks. Social network analysis is an attempt to answer some questions such as: Which entities or people are leaders and which of them are followers? What are the influential elements? Are there any groups and how they are formed? Which elements are important in a group? Which elements are the outliers? Which relationships are important? When these networks are small, the manual analysis will be easy but impossible with large networks, so when the network is large and complex, the social network analysis software can be used.

2. Social network analysis software

Social network analysis (SNA) software includes two packages based on graphical user interfaces (GUIs) or packages made for programming / scripting. GUI packages are generally easier to learn while programming tools are powerful and extensible. GUI packages have been well documented and widely used as UCINet, Pajak and so on. Programming tools have been also well documented and used to analyze network including Net Miner with Python scripting engine, a set of statistical programming language R and other packages. However, learning is difficult, some of these open source packages are growing very fast in terms of performance and features of the software.

The features of SNA software are the calculation of network standards and statistics (eg, Betweenness, Centrality and other similar cases), creating adjacency matrix, list edges, or edges, visualization and more. Visual presentation or visualization of social networks which are understanding of network data and result of analyzing are very important. Visualization is often facilities of qualitative data network that includes graphs, tables, timelines, maps, and other isolated cases. It can also be used to isolate data.

3. Software introduction

In this paper, some software of SNA (both packages) have been studied in terms of language implementations, features (access to plug-ins and library functions , calculating statistical measures, expandable, etc.), platform, input and output formats, volume of received data, download (free or not), open sources , parallel implementation. Statistical measures include centrality, betweenness, the degree of nodes, the leading index, density, diameter and more. Layout algorithm includes cyclic, tree, weight of edges and so on. Modular allows developers simply add and extend the features which very little work is required to add new connections or external data source [8,9,5].

3.1 Software SocNet

It is a cross-platform, user-friendly software to analyze and visualize social networks. This software allows you to create social networks with a few clicks on a virtual canvas or load social network data in various formats. Calculating the statistical metrics, creating random network automatically, including a web crawler, easy user interface, support loading and editing of multiple relationships. Layout algorithm for graphs, clustering coefficient, clicks, that automatically renovates unknown SN dataset is extensible. The purpose of creating this software is for student, business and trading researches. The software was generated in 2004 [16].

3.2 Gephi

It is an application for people who are interested in discovering and understanding charts, for example, in Photoshop user reveals the hidden features with the manipulation of form and color. The aim is to help analyzing hard data for hypothesis, direct discovery of patterns, the separation of wonderful structures or faults in data sources. This is a complementary tool to traditional statistics which is recognized as visual thinking with interactive relation to facilitate argument. Some features of this software are: Statistical criteria, the state of the art algorithm, community recognition, random generator, clustering and hierarchical graphs, dynamic filtering, dynamic network analysis, drawing, user-centric, modular architecture, the HITS ranking, expandable, access to plug-ins. The purpose of

creating this software is for student and trading researches. The software was generated in 2008 and its stable version is in 2013 [16].

3.3 R

This software is a programming environment and programming language for statistical computing and data analysis. This application is presented as a very useful tool in the field of social network analysis. This application can use multiple libraries to analyze data which these services include: egram, degreenet, igraph, begra and so on. Some features of this software are included: A statistical programming environment in R, creating graphs, simplifying the graph, calculating the statistical metrics, visualization of graphs 3D, 2D, modular, it has conditional, loop and recursive functions, calculate the minimum spanning tree, scalable, access to its software packages. The purpose of creating this software is for business researches. The software was generated in 1993 and its stable version is in 2015 [19].

3.4 UCINet

The application is created to analyze and visualize the social networks. This software contains a package of software pajek, netdraw .Some features of this software are: Clustering coefficient, calculating the statistical measure, calculating clicks, calculating incongruity - many modes and interaction of graphs, creating symmetric matrix, visualization by using software Netdraw and E-Net, including Excel editor, matrix and text. The software format is h###. The purpose of creating this software is for student and business researches, and is introduced 2006 [18].

3.5 NodeXL

This application has a ready and flexible framework in Excel environment for Microsoft in 2007, 2010 and 2013 which makes discovering patterns easy and shows easily graph networks as the output. Some features of this software are: Creating new edges based on similarity, calculating statistical criteria, eigenvector of centrality, clustering coefficient, PageRank, flexible layout , zoom, dynamic filtering, group of powerful top, a series of repetitive tasks with a single click, communities detection, access to plug-ins, direct connection to SN such as Facebook, email, twitter and other social networks and expandable. The purpose of creating this software is for business researches, and is introduced in 2008 [6].

3.6 Graph-tool

This application is an efficient module for statistical analysis of graphs and visualizing social networks. The software has multi-core architecture. Some of its features are: A convenient interface for Graphviz packages, supports edge, nodes and properties of graphs arbitrarily, calculate the statistical criteria, implements the topological standard algorithms (minimum spanning tree, the implementation connected to Dominator tree), detects Modules and communities through statistical inference, many algorithms are implemented by using a library in parallel, uses graph ML, is able to filter nodes and edges and extended. The purpose of creating this software is for business researches, and its stable version was presented in 2015 [12].

3.7 Pajek

The software is designed to transform the network into smaller networks and analyze and visualize them out. Some features of this software are: Uses 6 structures to implement the algorithms : network, vector, clustering, partitioning, hierarchical, permutation , calculates statistical criteria, calculates click and core, extracts sub network, eliminates loops in the network, implements splc-kn-cpm methods , Shirking clustering in the network, scalable , 3D visualization , implements layout algorithm , identifies strong and weak connections and two connections, identifies network threshold , reorders network, supports different languages for data files. The purpose of creating this software is for businesses and almost was created in 1999 [20, 21,1].

3.8 Graphviz

It is a visualization software, the graph visualization is a way to show the structural information as summary chart of graph and network. This software is used for the field of bioinformatics, software engineering, database and web design, machine learning and other areas. Some features of this software are: The charts are described by using Graphviz layout algorithm in a simple language, shows directional edges as a hierarchical graph by default, creates an object-oriented relationship for graph and graph algorithms in the library, uses Bayesian networks, knowledge representation, database schema design, digital logic design, and decision chart and so on., supports libraries that have up to 45,000 lines, is genetic algorithm and extensible. The purpose of creating this software was for businesses and students, and introduced before 2000 [11].

3.9. NetMiner

It is a software to analyze and visualize the data networks, Python-based script. The software allows you to explore your network visually and interactively and recognize patterns and network infrastructures. Some features of this software are [22]: The recognition of p patterns and network infrastructure, Python-based script, compatible with plugins, can be shared with others, distributed, shows drawings as a 3D, includes editor, includes data mining module based on machine learning that has the ability to filter, cluster, categorize, and recommend and is extensible. Analysts can easily keep track of their working experience and use them. The software can be run on the server and multiple users simultaneously can use it (for institutions and companies are allowed). The purpose to create software is for businesses and its initial version has been published in 2001 and its stable version has been released in 2011 [22].

3.10 AutoMap

It is a software to extract text (text mining) that is designed to work with software ORA. The software is able to extract text from text using text analysis method and supports various types of data from unstructured documents, the type of information extracted can be : Analytical content data (words and repeat), semantic data network (a network of concepts), Meta-data networks (cross-classification of ontology concepts like the class of people, objects, and connections between them) ,sentiment data (attitudes and beliefs), uses speech tagging and analyzes them and can be developed , algorithm for analysis is based on Carley's method for coding the text. The purpose to create software is for businesses and students and has been published in 2008 [15].

3.11 Tulip

It is a software to analyze and visualize social networks. Its purpose is to provide the development of a complete library and support the visualization of interactive data for relational data. Some features of this software are: Calculating the statistical measure, graph, click, graphical user interface: Histogram, Geographical, parallel coordinate and adjacent matrix allows several loops and a few edge between two nodes for directed graphs. All elements are from C ++ objects, supports the access to the library and the reform of the graph, reverse edge, supports hierarchical graphs that have subset, it can be developed, can be valued

for each node and edge. The application does not have any warranties of merchantability or fitness for a particular purpose. Stable version of this software has been released in 2013 [14].

3.12 Cytoscape

It is a software to analyze and visualize social networks and includes domains such as bioinformatics, analysis of social networks and semantic web. The new version of this software is with modular architecture. Some features of this software are: It is an extensible software, its new version is with modular architecture, it is a JavaScript library to analyze network and is a web application. Implementation of various layout algorithms, it is compatible with plug-ins, has ability to filter out nodes and customize the home screen, supports multiple languages and searching target node is applicable by advanced search. The purpose to create software is for businesses and its initial version has been published in 2002 and its stable version has been released in 2015 [13].

3.13 GraphChi

It is a software to analyze and visualize social networks that can do large calculations on machine by using a new algorithm for processing the graph of Dyskra. It is top model-driven software and runs synchronous programs. This application is compared with Graph Apache that software Graph has implemented 5 billion edges in about 75 minutes while GraphChi does it at the same time with just one machine. Some features of this software are : supports updating flow graph and deletes edge of the graph , implements algorithm CF, linear scalability, the ability to run billions of edges with a modern laptop, it can be developed. The purpose to create software is for businesses and has been released in 2010 [23, 7].

3.14 Meerkat

It is a software to analyze the text, which automatically summarizes large volumes of text and figures out its social networking. It uses online conversations in media sites such as Facebook, YouTube, blogs and so on. It was first emerged to help doctors to analyze the data in the form of social networks. Some features of this software are: It can be developed and provides information about the most influential central node. It visualizes network, edits interactive networks, supports dynamic network, calculates the statistical metrics, identifies populations , shows dynamic society at time , supports libraries C ++. The purpose to create software is for businesses and has been released in 2015 [2].

3.15 NetWorkit

It is a software to analyze and visualize social networks with a high-performance and uses multi-core architecture, and is comparable to the software NetWorkX. Some features of this software are: it is a Python module written in C ++ language, calculates the statistical metrics, interactive workflow, easy integration with Python libraries, parallel shared-memory, productive model, visualize based on matplotlib and network, integration with other software, network traffic analysis, population detection, clustering coefficient, launches and controls a remote server for calculations and codes are written in Google test framework. The purpose to create software is for businesses and has been released in 2015 [10].

3.16 NetWorkX

It is a software to analyze and visualize social networks which some its features are: It is a software package in Python language to create, manipulate and study the structure of dynamic and complex network functions, ability to have multi-graphs, network analysis, Measuring productive for classic graphs - random graphs and artificial networks, the node can be anything open source - click discovery - K-Core, calculates the statistical metrics, 3D visualization, it can be developed. The purpose to create software is for businesses and has been released in 2004 [3].

3.17 Visone

It is a software to analyze and visualize social networks and help beginners and experts to apply innovative and advanced techniques easily and accurately. Some features of this software are: It has a graphical user interface tailored to the social network, visualization of innovative network, supporting relationship that has not been confirmed, calculates statistical measures, scalable, page rank. The purpose to create software is for businesses and its initial version has been published in 2002 and its stable version has been released in 2015 [4].

Capabilities Software	Language Implementation	open source	Platform	Input formats	Output Format	The volume of incoming data	downloa d	Paraell Executio n lon
SocNetV	C++	No	Windows •Linux •MacOS	Pajek(net,gml) UCInet(.dl), Graphviz(.dot), Adjacency, edge	GraphML Default, Pajek, Adjacency Matrix, PDF,PNG, JPG	Big Data	Free	No

				list(.csv), GraphML,				
Gephi	Java	Yes	Windows Linux MacOS x	GraphViz(.dot), Graphlet(.gml), GUESS(.gdf), LEDA(.gml), NetworkX(.graphml, .net), NodeXL(.graphml, .net), Pajek(.net, .gml), Sonivis(.graphml), Tulip(.tlp, .dot), UCINET(.dl), yEd(.gml), Gephi (.gexf), Edge list(.csv), databases	GUESS(.gdf), Gephi(.gexf), .svg, .png	Node10 , ⁴ Edge10 ⁶	Free	No
R	R	Yes	Windows Linux MAC	Almost all formats	Most popular formats Pajek, Graphml And other	Big Data	Free	No
UCINET	Java	Yes	Windows Linux	DL, Excel, VNA, Pajek, Text	DL, Excel, Pajek, Mage, Metis, VNA from netdraw.	Big Data	Free trial version for 90 days After paymen t	No
NodeXL	C#.net	Yes	Windows XP, Vista, ⁷	email, .csv (text), .txt, .xls (Excel), .xlslt (Excel 2007, 2010, 2013), .net (Pajek), .dl (UCINET), GraphML	.csv (text), .txt, .xls (Excel), .xlslt (Excel 2007), .dl (UCINET), GraphML	Big Data	Free	No
Graph tool	C++, python	No	Mac, Linux	GraphViz(.dot), Graph ML	GraphViz(.dot), GraphM L, .bmp, .canon, .cmap, .eps, .fig, .gd, .gd2, .gif, gtk, .ico, .imap, .cmapx, .ismap, .jpeg, .pdf, .plain, .png, .ps, .ps2, .svg, .svgz, .tif, .vml, .vmlz, .vrml, .wbmp, .xlib	10 ⁷ Node	Free	Yes
Pajek	Java	No	Windows Linux	Convert text file and excel The format pajek UCINET(dl) GED, Ore-graph, p- graph, Some molecular formats PDF,	Pajek (.net) UCINET(dl) VOS	Very big data	Free	No
Graphviz	C, C++, Java	Yes	Windows Mac Linux	GraphViz(.dot)	.bmp, .canon, .cmap, .eps, .fig, .gd, .gd2, .gif, gtk, .ico, .imap, .cmapx, .ismap, .jpeg, .pdf, .plain, .png, .ps, .ps2, .svg, .svgz, .tif, .vml, .vmlz, .vrml, .wbmp, .xlib	Size Medium volume 1000Node Big Data	Free	No
AtuoMap	Java	No	Windows	text documents	.csv and DyNetML	Big Data	Free	No
Tulip	C++	No	Windows Mac Linux	Tulip format (.tlp), GraphViz (.dot), GML, txt, Csv, pajek (net)	.tlp, .gml	Big Data	Free	No
Cytoscape	Java	Yes	Windows Linux	Xml, csv, jar, Cxf	PDF, BMP, IPEG, PNG, SVG, PS	10 ⁷ Nod e and Edge	Free	No
GraphChi	Java & C++	Yes	Linux MacOSX	Adjacency matrix format Edge list format CSC format	CSV SVD SGD	10 ⁹ Edg e Big Data	Free	Yes
Meerkat	Java	No	Ubuntu MacOSX	TXT	XML TXT MBZ	Big Data	Free	No

NetWorkit	C++	Yes	Linux , Mac , Unix	Dibap, GraphViz, METIS, SNAP, various edge list formats	GUESS, Graphlet,METIS, SNAP, VNA	Big Data	Free	Yes
NetworkX	Paython	Yes	Windows Linux	GML, Graph6/Sparse6, GraphML, GraphViz (.dot), NetworkX (.yaml, adjacency lists, and edge lists), Pajek (.net), LEDA	GML, Gnome Dia, Graph6/Sparse6, GraphML, GraphViz (.dot), NetworkX (.yaml, adjacency lists, and edge lists), Pajek (.net), and assorted image formats (.jpg, .png, .ps, .svg, et al.)	Big Data	Free	No
Vison	Java	No	Windows , Linux , MacOS	Most of the formats like PDF, SVG,JPEG ,GraphML...	Most of the formats like PDF, SVG,JPEG ,GraphML...	Big Data	Free	No

4. Conclusion

In this paper we have considered to social networks and their analysis, and introduced some tools that were able to analyze or analyze the major social networks. All these tools are able to calculate network metrics and statistics (of the center, the density, diameter, etc.), edit, detect populations and filter the network. Most of them were open and extensible, and most of them support some input formats like Pajek, UCINET and other popular formats. The implementation language of the most of them was C ++ Java and was applicable on some platforms such as Windows, Linux and MAC OS. Some of these tools also were able to visualize the network which is their advantage against other tools. Among these tools, only GraphChi and NetMiner have recommender algorithms. The only tool that was able to receive a high volume of data is GraphChi software with 10^9 nods. Most of these tools are free and do not need to pay.

Reference

- 1-http://ocw.mit.edu/courses/economics/14-15j-networks-fall-2009/assignments/MIT14_15JF09_pajek.pdf
- 2-<http://meerkatapp.co/guidelines>
- 3-<http://networkx.github.io/>
- 4-<http://visone.info/index.htm/>
- 5-https://en.wikipedia.org/wiki/Social_network_analysis_software
- 6-<http://nodexl.codeplex.com>
- 7-<https://github.com/GraphChi/graphchi-cpp>
- 8-<http://www.orgnet.com/sna.html>
- 9-<http://butleranalytics.com/10-free-social-network-analysis-software>
- 10-<https://networkit.iti.kit.edu>

- 11-<http://www.graphviz.org>
- 12-<http://graph-tool.skewed.de>
- 13-<http://www.cytoscape.org>
- 14-<http://tulip.labri.fr>
- 15-<http://www.casos.cs.cmu.edu/projects/automap>
- 16-<http://socnetv.sourceforge.net>
- 17-<http://gephi.github.io>
- 18-<https://sites.google.com/site/ucinetsoftware/home>
- 19-<http://www.r-project.org>
- 20-<http://vlado.fmf.uni-lj.si/pub/networks/pajek>
- 21-BatageljVladimir,1999.Pajek– Program for Large Network Analysis.
- 22-<http://www.netminer.com>
- 23-kyrola.aapo.etal.(2013). GraphChi: Large-Scale Graph Computation on Just a PC.
<http://www.cs.cmu.edu>