Description of GNP (Gephi, NodeXL, Pajek) Social Network Analysis Tools

M. Thangaraj[1], S. Amutha[2]

[1]Associate Professor, Department of Computer Science, Madurai Kamaraj University, Madurai, India

[2]Research Scholar, Department of Computer Science, Manomaniam Sudaranar University, Tirunelveli, India

Abstract: *This paper explains about important useful tools in Social Network Analysis and describe about the tools extract and comfortably for the WWW users. It explains elaborately the details about SNA tools like Gephi, NodeXL and Pajek. The SNA tools features, metrics, applications, architecture, input/output format, motivation of the selected tools.*

Keywords: Social Network Analysis, Gephi, NodeXL, Pajek

1. Introduction

Nowadays Billions of people create trillions of connections through social media each day every second, but few of us consider how each click -mouse and key button-press builds relationships that, in cumulative, form a vast Social Network. Ardent users of social media tools such as email, blogs, micro blogs, and wikis eagerly send personal or public messages, post strongly felt opinions, or put in to community knowledge to build up partnerships, encourage cultural heritage, and advance development. [1]

Devoted Social Networkers create and share digital media and pace or urge resources to pool their experiences provide help for neighbors and colleagues, and express their creativity research is faced with both over- the-top opportunities and challenges [2].

Society is willing to invest in research as the basis of changing acknowledges economy as long as research proves to be responsive to its needs [14]. One of the most visible trends on the web is the emergence of Social web sites, which help people make and gather knowledge by simplifying user contribution via blogs, marking and folksonomies, wikis, podcasts and Online Social Networks (ONS)[4].

The quickly growing up web based SNs applications provide abundant data for mining, at the same time bring big gainsays to the field. In this paper various types of SNA tools with different types of features, metrics, operations, architecture and applications are explained elaborately. SNA is an active area of study beyond sociology [5].

The total number of nodes and relations make drawing the entire network completely infeasible [7].It also gives result in a mismatch between the density of information and screen resolution available. Next, large SNs commonly have complex relations among their actors [9].

SN structures are created when people connect to one another through a range of ties. SNA tools are support to connect Researchers, Practitioners Collaborative, Employees, and Businessman etc [12]. SNA tools have documented the impact of social structure to perform a set of core operations that measure and map the resulting dataset [22].

SN is also characterized by a typical methodology about techniques for collecting data, Statistical analysis, visual representation, etc [17]. The following SNA tools are explained in detail.

Figure 1: SNA tools
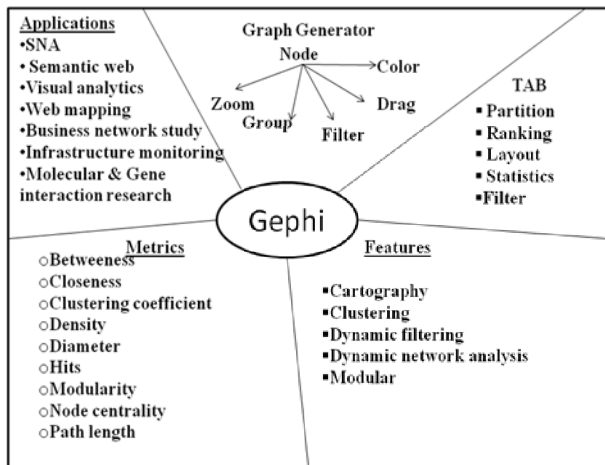
2. Gephi Tool

2.1 Goal of Gephi

Gephi tool is a highly scalable and built in Java based open source and multi-task model SNA tool. It is focused on interactivity by fast graphics, Multithreaded and visual updates [3]. It is to help in data analysts to make guess discover patterns, set a part of structure singularities or fault during data sourcing [15].

2.2 Graph Generator of Gephi

In graph generating to explore large number of nodes use interaction technique includes node coloring, node dragging, node filtering, node grouping and zooming the nodes like node painter and shortest path discovery. It offers a plug-in library for extending the functionality [10]. By using the random color coding detects the cluster communities and label it [8]. Gephi tool very easy and quickly deploy the graph to wider audience and also the figures can be saved in various format. It is run in windows and MAC.

2.3 Format of Gephi Tool

Large amount of data are processed in this tool. This is mainly best supporting format in Gephi are CSV (comma-separated values), DL(full matrix), UCINET, DOT (Document Template File), GraphViz (Graphic format), GDF(Game Definition File), GEXF(Graph Exchange XML Format),GML(Geography Markup Language), GraphML (Graph Markup Language), NET, Pajek, TLP (Project Timeline File), Tulip, VNA(Video Capture Album), Netdraw, Spreadsheets etc [30].In gephi tool MySQL, SQL Server, PostgreSQL, SQLite and Teredata are used for Databases. For Graph databases Ne04j, Orient DB and infinite Graph used in this tool.

Figure 2: Details about Gephi Tool

2.4 Tab usage in Gephi

There are five types of Tabs are used for changing nodes and edges. Partition tab is using for change node or edge color, Ranking Tab is use to change color or size of node or edge or label, [18]. Layout Tab is select and customize layout algorithms, Filters Tab is use to select specific node or edge, Statistics Tab is use to calculate network or node or edge.

2.5 Features of Gephi

In Network representation the partition data or ranking to make a meaning creation is called Cartography [21]. To find the large size of Social communities, Health Groups, Researcher groups, Collaborative groups to explore the hierarchical graphs and Clustering. To create the complex filter query without scripting and build new networks from the filtering result and save the favorite queries to filter the network to select nodes and/or edges based on the network structure or data is called Dynamic Filtering [23].

There are different type of file format also accept to import data in the ideal platform tool is Gephi to filter with the timeline component like dynamic structures in SNs, it is for leading the innovation about Dynamic Network Analysis. There is layout algorithms gives perfect shape to the graph. It is efficiency and quality in while running dramatically increase user feedback and experience.

2.6 Metrics in Gephi

There are various types of metrics in Gephi tool. Betweeness, which is explain the micro and macro between is meso [11]. Closeness is nothing but of a node is a compute of centrality in a network, considered as the computation of the length of the shortest paths between the node and all other nodes in the graph. Clustering coefficient is applied to a single node, is a measure of how complete the area of a node is applied to an entire network nodes.
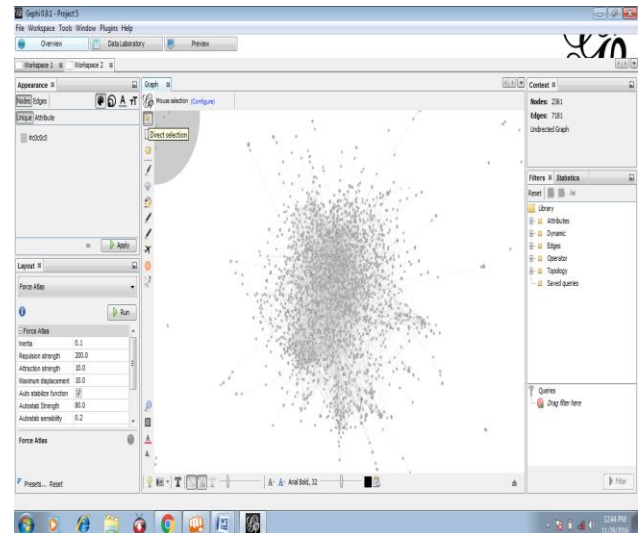
Figure 3: Screenshot for Gephi Tool

Density, which indicates what proportion of all possible relationships in a network, actually exists. Diameter is the maximal distance between all pairs of nodes [19].HITS is a Hyperlink-Induced Topic Search, which is used to assess the relationship between the nodes in a graph and also find the scores for each node in the graph [16].

Modularity score point out sophisticated inner structure. This structure, often called a society structure, depicts how the network is compartmentalized into sub-networks. Figure .3 shows the visualization graph about Force Atlas algorithm used for nodes 2361 and edges 7182 by undirected graph drawn. The attribute is Degree minimum size is 1 and maximum size is 4. The Force Atlas algorithm is used.

2.7 Algorithms

There are number of algorithms used in Gephi tool. Force Atlas algorithm, Fruchterman-Reingold, Force Atlas 2, Open Ord, Yrfan Hu Multilevel, GeoLayout, Radial Axis Algorithms are used in Gephi Tool.

3. NodeXL Tool

NodeXL is Network Overview for Discovery & Exploration in Excel. It is intended to allow Excel users to simply bring in, cleaning, analyze and visualizing network data [24]. It enlarges the existing graphing features of the worksheet with the extra chart type of network, thus lowering the road block for acceptance of network analysis.

NodeXL support for background images is simple. It can't scale a background images which is why the better on the graph gallery. It is integrated into the Excel 2007 spreadsheet to gain access to its rich set of data analysis and charting features [13]. In this tool always have a rule, sort, filter, or simply type the data into cells in the spreadsheet containing network data. It offers a variety of visual properties present advanced filtering capabilities, calculates frequency used individual and overall complex metrics and supports diverse visual networks layouts.
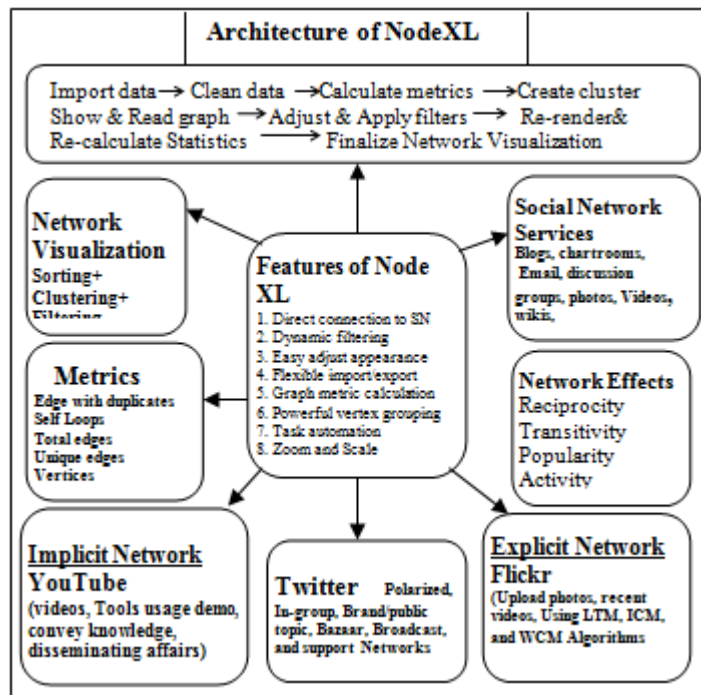

Figure 4: Details about the NodeXL tool

3.1 Architecture of NodeXL

First, import the data in excel spreadsheet and clean the data [31]. Second, calculate the metrics and creating the clusters, it shows and read the graph. It adjusts and applies the filters in spreadsheet process, re-renders and re-calculates the final network visualization.

3.2 Features of NodeXL

NodeXL may be one of the easiest ways to both manipulate network graphs and get graphs from a variety of Social Media sources. It also allows users to use a logarithmic mapping rather than the default linear mapping. In degree centrality is a count of the number of unique edges that are connected to it. It strength is the ease with which can filter and customize the network visualization [25]. It is used refine the graph visualization, only links between nodes from different content topics were retained and analyzed.

3.3 Format of NodeXL

This tool is easy to import and export graphs in GraphML, Pajek, UCINET and Matrix Formats. It is directly connect to SN directly from twitter, YouTube, Flickr, and email or use one of several available Plug-ins to get networks from Face book exchange and WWW hyperlinks [27].

In Network Visualization there are three major processes done by the NodeXL tool [32]. First, the nodes are sorted and it clustered into groups and finally filtered the nodes depends upon the categories of graph.

In VOSON system a web based software incorporating web mining, data visualization and traditional empirical social science methods. It collects messaging data from the Microsoft exchange server and displays the data as a graph. It zooms into areas of interest and scales the graphs vertices to reduce clutter [33]. It takes the flexible layout to use one of several force directed algorithm to layout the graph or drag vertices around with the mouse.
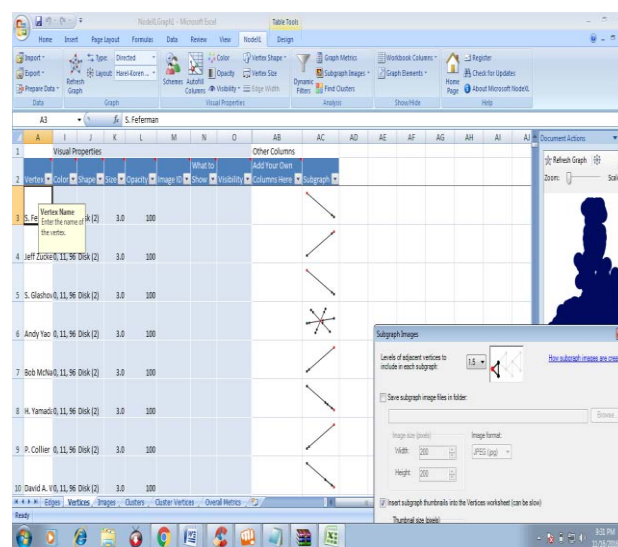

Figure 5: Screen shot of the NodeXL tool.

It had easily used to form set the color, shape, size, label and opacity of individual vertices by filling in worksheet cells or does it for based on vertex attributes such as degree, betweeness, centrality or page rank [34]. It does dynamic filtering directly hide vertices and edges using a set of sliders. It also done powerful vertex grouping and consider their connectedness and by design group them into clusters.

NodeXL was made to quick teach by social-media knowledge business professionals who already use Excel, as well as by English known graduate students who are learning SNA [43]. The tool provides simple import of data from important social media tools. It considerably expands the community of users who can carry out analyses that guide to actionable trade insights and research studies.

In NodeXL template shows the columns are vertex1 and vertex2, color, width, opacity, visibility in the visual properties. There is other option in it to add own columns in other columns sections. It's all in the edges path, in vertices additionally the image ID and what to show options are add and in images –image ID and image file path. In cluster – vertex color and vertex shape only, but in overall metrics – metric and value are calculated [39]. There are different types of layout algorithms and dynamic query filters allows users to change display to their needs and also centrality metrics for directed and undirected graphs, as well as a number of clustering algorithms, support exploration and discovery.

It is supported through the emerging Social Media Research Foundation [41]. Figure 5. Shows that NodeXL tool generate the color, shape, size, opacity and sub graph of the vertices. The type of graph is directed and layout is Harel-koern is used in the tool.

3.4 Social Network Services in NodeXL

The tool is mostly used in Social Network Services (SNS), Blogs which is used to ask their queries and post own comments about the different answers from different places of people [26]. Chat rooms are used have long conversations between which are known friends use through their E-mail way.

Discussion groups are used find a new solution for problem through senior persons in the particular field [20]. Photos and videos shared through the Facebook accounts and Fan Pages can be analyzed in a similar way to twitter accounts. Very popular fan pages generate lots of interactions between fans and most other organization's pages have fewer [46].

There are several types of connections identified within Facebook contributions and unimodal connections such as user-user. Now-a-days the popular people (politician) are says there wishes and update the news by their twitter page [35].It evaluates a set of basic network metrics, permitting users well-known with spreadsheet actions to apply this expertise to network data analysis and visualization [40]. The tool also contains algorithms for calculating network metrics like degree centrality, betweenness, centrality, clustering

coefficient, eigenvector centrality etc. In this tool there is implicit and explicit networks are available.

3.5 Pajek Tool

Pajek is a program, for Windows, for analysis and visualization of large networks having some thousands or even millions of vertices. In Slovenian language the word Pajek means spider [6].The main goals of tool are to provide the user with some potent visualization tool and also to implement a choice of efficient algorithms for analysis of large networks [51].

3.6 Motivation of Pajek

The main motivation for development of tool was the notice that there exist several sources of large networks that are already in machine-readable form [47]. It should provide tools for analysis and visualization of networks, association networks, organic molecule in chemistry, protein-receptor interaction networks, genealogies, Internet networks, citation networks, diffusion (AIDS, news, innovations) networks, data-mining (2-mode networks), etc [36].

The aim of tool is based on experiences added in development of graph data structure and algorithms libraries Graph and X-graph, collection of network analysis and visualization programs [37]. This tool generates transform, random network, Partitions, components, hierarchical, decompositions, numbering, citation weights, k-neighbors, path between 2 vertices, Critical path method (CPM), maximum flow, vectors and count in NET. There is a special edition of program Pajek-XXL. Its memory consumption is much lower. For the same network it needs at least 2-3 times less physical memory than Pajek previous version. [42].
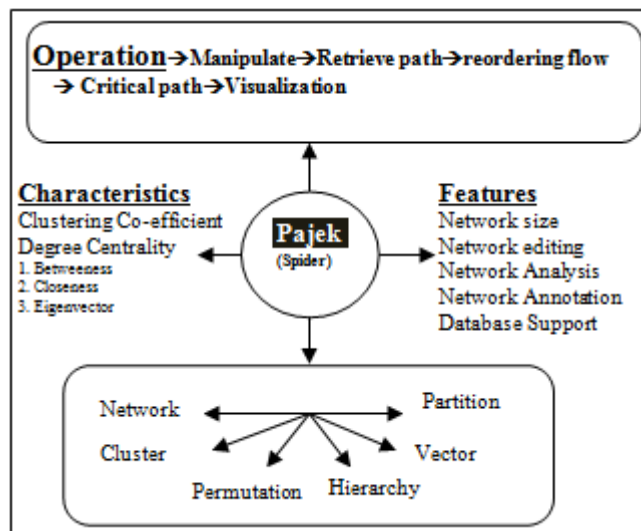

Figure 6: Details about the Pajek Tool.

There is special option tools in Pajek is the R and SPSS. It is easy to send and locate the edges and vertices directly through the Pajek tool.

3.7 Goal of Pajek

The main goals in the design of Pajek are to support notion by breakdown of a large network into several smaller networks that can be treated further using more sophisticated methods [28]. It provides the user with some powerful visualization tools. It implements a selection of efficient algorithms for analysis of large networks.

3.8 Architecture of Pajek

The Pajek tool manipulates the nodes and retrieves the path through reordering flow and also find the critical path of the node finally develop the visualization graph. The tool holds the major objects are networks, which is main object of the tool and connect vertices and lines. Cluster, which is subset of vertices which is used in visualization graph, belongs to same group.

Permutation is interchange or reform of vertices. Hierarchy is arranging nodes and edges ranking order. Vectors are used to store quantitative information of nodes and partition are used to divide nodes into none be related groups based on a nominal variable [29].
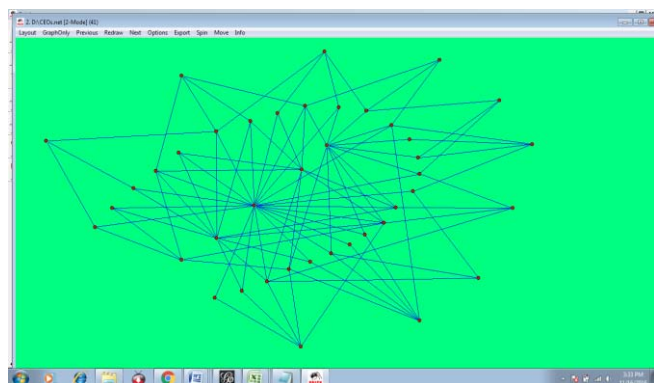

Figure 7: Screen shot for Random Graph generate by Pajek tool

In Partitions the degree, depth, core, p-cliques, center of the node are calculated. In Binary operations the union, intersection and difference are intended in this tool. The components are connected in strong, weak, bi-connected and symmetric etc [44]. Figure 7 show about the Random graph generate with nodes in dataset belongs to CEOs.net in the 2-mode model.

In this tool all paths between two vertices, find the shortest path(s), and Critical Path, there is maximum flow between two vertices and calculate the Citation weights, Neighborhood, extract sub-network and decrease clusters in network [48].

3.9 Algorithms

There are various types of algorithms used in pajek tool. The algorithms are mostly used, Kamada-Kawai optimization, Fruchterman -Reingold optimization, VOS mapping, Pivot MDS, Drawing in layers, FishEye transformation are used in Pajek Tool.

3.10 Features of Pajek

This tool is based on the Network Size, Network Editing, Network Analysis, Network Annotation and Database Support [50]. Network Size user can increase the nodes and edges the extension size of the Network, Network Editing the nodes and edges can update the paths whether it's critical or shortest path [38].

In Network analysis calculates the current flowing of edges and nodes. Network Annotation defines the explanation of the graph or diagram derives by the nodes and edges. By using database it support the millions of nodes and edges are stored, need for processing to the visualization graph [45].

4. Conclusion and Future enhancements

The most popular SNA tools Gephi, NodeXL and Pajek are elaborately described. The SNA tools have a various features to design different types of visualization graphs to derive in a single graph with millions of nodes and edges relationship [49].

In Gephi tool there is plug-in option to coordinate with other SNA tools like NodeXL and Pajek. The input file format are same it is easy to implement in forth coming paper.

References

[1] Abascal-Mena, R. ; Lema, R. ; Sedes, F. From tweet to graph: Social network analysis for semantic information extraction ,Research Challenges in Information Science (RCIS), 2014 IEEE Eighth International Conference on 2014 , Page(s): 1 - 10

[2] L.A. Adamic and E. Adar. Friends and neighbors on the web. Interanational journal of Social networks, 25(3): 211-230,2003.

[3] Adar, E-Guess: a language and interface for graph Exploration. In CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems,2006- pp791–800.

[4] Akhtar, N. Social Network Analysis Tools ,Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on IEEE: 2014 , Page(s): 388 - 392

[5] Ali Rahmani, Alan Chen, Abdullah Sarhan,Jamal Jida, Mohammad Rifaie, Reda Alhajj-Social media analysis and summarization for opinion mining:a business case study-Social Network Analysis and Mining. (2014) 4:171

[6] Andrej Mrvar1 and Vladimir Batagelj Analysis and visualization of large networks with program package Pajek- Complex Adaptive System Modeling (2016) 4:6

[7] Anuradha Goswami, Ajey Kumar-A survey of event detection techniques in online social networks-Social Network Analysis and Mining-December 2016, 6:107

[8] Bastian, M., Heymann, S., and Jacomy, M., Gephi: An open source software for exploring and manipulating networks," in [International AAAI Conference on Weblogs and Social Media], (2009).

[9] Bogdan Batrinca,Philip C. Treleaven-Social media analytics: a survey of techniques, tools and platforms- AI & SOCIETY-February 2015, Volume 30, Issue 1, pp 89–116

[10] Bostock, M. and Heer, J., Protovis: A graphical toolkit for visualization," Visualization and Computer Graphics, IEEE Transactions pages:1121to 1128 Nov. 2009.

[11] Brigitte Boden, Roman Haag, Thomas Seidl ,Detecting and exploring clusters in attributed graphs: a plugin for the gephi platform , CIKM '13: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, October 2013

[12] Bruna de Paula Fonseca e Fonseca, Ricardo Barros Sampaio, Marcus Vinicius de Araújo Fonseca and Fabio Zicker- Co-authorship network analysis in health research: method and potential use- Health Research Policy and Systems (2016) 14:34

[13] Changhyun Byun, Yanggon Kim, Hyeoncheol Lee, Kwangmi Ko Kim, Automated Twitter data collecting tool and case study with rule-based analysis. IIWAS '12: Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, December 2012

[14] Cherry Ahmed, Abeer ElKorany, Reem Bahgat- A supervised learning approach to link prediction in Twitter- Social Network Analysis and Mining. (2016) 6:24

[15] Cody Buntain, Jennifer Golbeck,Identifying social roles in reddit using network structure, WWW Companion '14: Proceedings of the companion publication of the 23rd international conference on World wide web companion, April 2014

[16] David Combe,Christine Largeron1, Elʺod Egyed-Zsigmond2 and Mathias Géry A comparative study of social network analysis tools, International Workshop on Web Intelligence and Virtual Enterprises 2 (2010)

[17] David F. Nettleton ,Survey: Data mining of social networks represented as graphs,February 2013 Computer Science Review , Volume 7

[18] Derek Hansen, Ben Shneiderman, Marc A. Smith ,Analyzing Social Media Networks with NodeXL: Insights from a Connected World ,September 2010

[19] Derek Hansen, Marc A. Smith, Ben Shneiderman, EventGraphs: Charting Collections of Conference Connections, January 2011 ,HICSS '11: Proceedings of the 2011 44th Hawaii International Conference on System Sciences, IEEE Computer Society

[20] de Nooy, W., Mrvar, A., & Batagelj, V. Exploratory social network analysis with Pajek. New York, NY: Cambridge University Press(2011).

[21] Elizabeth J. Sbrocco Visualizing a Mitochondrial Haplotype Network in Gephi-National Evolutionary Synthesis Center November 1, 2012

[22] Giorgio Uboldi, Giorgio Caviglia, Nicole Coleman, Sébastien Heymann, Glauco Mantegari, Paolo Ciuccarelli ,Knot: an interface for the study of social networks in the humanities ,September 2013

[23] A.A. Hagberg, D.A. Schult, and P.J. Swart. Exploring network structure, dynamics, and function using NetworkX. In Proc.7th Science Py Conf., pages 11–15, 2008.

[24] Heymann, S. ; Le Grand, B. Visual Analysis of Complex Networks for Business Intelligence with Gephi Information Visualisation (IV), 2013 17th International Conference : 2013 , Page(s): 307 - 312

[25] Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PLoS ONE 9(6): e98679.

[26] Jennifer Golbeck,Analyzing the Social Web, 1st edition,March 2013

[27] John R. Hargera,b and Patricia J. Crossnoa Comparison of Open Source Visual Analytics Toolkits- Jan 2010 page(s):204-213

[28] Kantert, Jan ; Bodelt, Sebastian ; Edenhofer, Sarah ; Tomforde, Sven ; Hahner, Jorg ; Muller-Schloer, Christian, Interactive Simulation of an Open Trusted Desktop Grid System with Visualisation in 3DSelf-Adaptive and Self-Organizing Systems (SASO), 2014 IEEE Eighth International Conference on 2014 , Page(s): 191 - 192

[29] KP Krishna Kumar and G Geethakumari- Detecting misinformation in online social networks using cognitive psychology- Human-centric Computing and Information Sciences 2014, 4:14

[30] S. G. Kobourov. Force-directed drawing algorithms.In Handbook of Graph Drawing and Visualization.CRC Press, 2012.

[31] Lin,F. Renner, R.Detecting Dynamic and Static Geo-social Communities ,Computing for Geospatial Research and Application (COM.Geo), 2013 Fourth International Conference on 2013 , Page(s): 134 - 135

[32] Mamta Madan, Meenu Chopra, Meenu Dave , Predictions and recommendations for the higher education institutions from Facebook social networks, 3rd International Conference on Computing for Sustainable Global Development (INDIACom) 2016

[33] Márcia Oliveira, João Gama, An overview of social network analysis,March 2012 ,

[34] Nagel, T, Duval, E., Heidmann, F. Exploring the Geospatial Network of Scientific Collaboration on a Multitouch Table, in Proc. 22nd ACM Conference on Hypertext and Hypermedia(2011)..

[35] Naz, S. Naeem, M. Afzal, M.T. Qayyum, A. Expertise identification and visualization Computing and Networking Technology (ICCNT), 2012 8th International Conference on 2012 , Page(s): 16 - 20

[36] Nguyen NP, Yan G, Thai MT (2013) Analysis of misinformation spread containment in online social networks.Comput Netw 57(10):2133–2146

[37] Przemyslaw A Grabowicz, Luca Maria Aiello3 and Filippo Menczer- Fast filtering and animation of large dynamic Networks- EPJ Data Science 2014, 3:27

[38] Puneet Sharma, Udayan Khurana, Ben Shneiderman, Max Scharrenbroich, John Locke, Speeding up network layout and centrality measures for social computing goals, SBP'11: Proceedings of the 4th international conference on Social computing, behavioral-cultural modeling and prediction, Springer-Verlag, March 2011

[39] Sameer Kumar, Jariah Mohd. Jan, Research collaboration networks of two OIC nations: comparative study between Turkey and Malaysia in the field of

`Energy Fuels', 2009---2011, Springer-Verlag New York, Inc.January 2014

[40] Salotti, J. ; Plantevit, M. ; Robardet, C. ; Boulicaut, J. Supporting the Discovery of Relevant Topological Patterns in Attributed Graphs ,Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on 2012 , Page(s): 898 – 901

[41] Sonia Stephens ; J. D. Applen- Rhetorical dimensions of social network analysis visualization for public health- Professional Communication Conference (IPCC), 2016 IEEE International- 14 November 2016

[42] Tugrul U. Daim , Monticha Khammuang, Edwin Garces - The Application of Social Network Analysis: Case of Smart Roofing- Anticipating Future Innovation Pathways Through Large Data Analysis Part of the series Innovation, Technology, and Knowledge Management 24 July 2016-pp 273-302

[43] Vafopoulos, M. 2010. Web Science Subject Categorization (WSSC), in Proc. Of the WebSci 2010.

[44] Vladimir Batagelj, Andrej Mrvar,cAnalysis of Kinship Relations With Pajek Social Science Computer Review , Volume 26 Issue 2 , May 2008

[45] Vladimir Batagelj, Selena Praprotnik- An algebraic approach to temporal network analysis based on temporal quantities- Social Network Analysis and Mining-December 2016, 6:28

[46] Wenger, E., Trayner, B., & De Laat, M.. Telling stories about the value of communities and networks: A toolkit. Ruud de Moor Centrum, Open Universiteit. 2011

[47] Wenhui Li ; Xuzhi Wang ; Qiuyu Zhu Topology analysis and clustering for localized network in SINA Weibo Smart and Sustainable City 2013 (ICSSC 2013), IET International Conference on 2013 , P: 321 - 324

[48] Wylie, B. and Baumes, J., A unified toolkit for information and scienti_c visualizations," in Visualization and Data Analysis , 7243(1), 72430H, SPIE (2009).

[49] Xu B, Huang Y, Kwak H, Contractor N Structures of broken ties: exploring unfollow behavior on twitter. In: Proceedings of the 2013 conference on computer supported cooperative work. ACM, pp 871–876.

[50] Yee, K.-P. Fisher, D., Dhamija, R., and Hearst, M., Animated exploration of dynamic graphs with radiallayout," Information Visualization, IEEE Symposium on 0, 43 (2001).

[51] Zhong-Yi Wang, Gang Li, Chun-Ya Li, Ang Li (2012). Research on the semantic-based co-word analysis.Scientometrics, 90, 855-875.