Journal of
Information Technology & Software Engineering

Agrawal et al., J Inform Tech Softw Eng 2015, 5:3
DOI: 10.4172/2165-7866.1000157

Research Article Open Access

A Comparative Analysis of Social Networking Analysis Tools

Hemant Agrawal, Ajay Thakur, Rajan Slathia, Sumangali K*

School of Information Technology and Engineering, VIT University, Vellore, Tamil Nadu, 632014, India

Abstract

Informal communities have known a vital improvement since the presence of web 2.0 stages. This prompts a developing requirement for informal community mining and Social Network Analysis (SNA) systems and instruments with a specific end goal to give more profound examination of the system additionally to distinguish groups in perspective of different applications. Thus, a ton of works have concentrated on chart portrayal or bunching and a few new SNA apparatuses have been created over these last years. The motivation behind this paper is to think about some of these instruments which actualize calculations devoted to informal community examination.

Keywords: Betweenness; Nearness; Social network analysis; Visualization; Vertices

Introduction

The blast of Web 2.0 (online journals, wikis, substance offering destinations, informal communities, and so forth.) opens up new viewpoints for imparting and overseeing data. In this setting, among a few developing exploration fields concerning "Web Intelligence", a standout amongst the most energizing is the improvement of utilizations spent significant time in the treatment of the social measurement of the Web. Especially, building and overseeing virtual groups for Virtual Enterprises require the improvement of another era of apparatuses incorporating interpersonal organization displaying and examination. A very long while back, the first deals with Social Networks Analysis (SNA) was done via specialists in Social Sciences who needed to comprehend the conduct and advancement of human systems. A few pointers were proposed to portray the on-screen characters and additionally the system itself. One of these indicators, for case, was the centrality that can be utilized as a part of showcasing to find the early adopters or the individuals whose action is prone to spread data to numerous individuals in a briefest manner. These days, the wide utilization of Internet around the globe permits to join quite a few people. As pointed in the Gartner study, this critical advancement of the systems offers ascend to a developing requirement for informal organization mining and informal community investigation strategies so as to give more profound perception of the system and to distinguish groups and study their development for applications in zones, for example, group advertising, social shopping, suggestion components and personalization sifting or graduated class management. For this reason, while numerous new advancements (wikis, social bookmarks and social labeling, and so on) and administrations (GData, Google Friend Connect, Open Social Face book Beacon) were proposed on web, a few new SNA devices have been created. These instruments are exceptionally valuable to dissect hypothetically an interpersonal organization additionally to speak to it graphically. They figure diverse markers which describe the system's structure, the connections between the performing artists and additionally the position of a specific performer. They additionally permit the examination of a few systems.

The motivation behind this paper is to present some genuine significant instruments and to depict some of their functionalities. A comparable examination has as of now been done in, however with a more measurable vision. Our relative review on the condition of- the-craftsmanship apparatuses for system visualization and investigation is centered on three fundamental focuses:

– Graph visualization;

– Computation of different pointers giving a neighborhood (i.e. at the hub level) or a worldwide portrayal (i.e. in general diagram);

– Community discovery

The hypothetical system for informal community investigation was presented in the 1960s. Taking after the fundamental thought of Moreno who proposed to speak to operators by focuses associated by lines. Therefore, they are considered as the organizers of the cutting edge diagram hypothesis for informal organization investigation. Two sorts of diagrams can be characterized to speak to an interpersonal organization: one-mode and two-mode charts.

One-mode graph

At the point when the connections between performing artists are viewed as, the interpersonal organization can be spoken to by a graph $G=(V,E)$ where V is the situated of hubs (or vertices) related to the performers, and $E \in V \times V$ is the situated of edges which compare to their connections. This is the case, for occurrence in a traditional dataset identified with a karate club where the hubs compare to the individuals from the club and where the edges are utilized to portray their fellowships. At the point when the connections are coordinated, edges are supplanted by bends. Hubs and also edges can have traits. All things considered, we can speak then about named diagrams.

Two-mode diagram

At the point when the connections between two sorts of components are considered, for instance the individuals and the rivalries in the karate club, a two-mode diagram is most suited to speak to two sorts of vertices. The edges are permitted just between hubs of distinctive sorts. The most widely recognized approach to store two-mode information is a rectangular information network with the two hub sorts separately in columns and segments. Case in point, a 2 dimensional grid with the performing artists in lines and the occasions in sections can speak to a two-mode chart for the karate club. This representation is exceptionally regular in SNA. Two-modes diagrams can be changed in one-mode

***Corresponding author:** Sumangali K, School of Information Technology and Engineering, VIT University, Vellore-632014, Tamil Nadu, India, Tel: 0416 224 3091; E-mail: sumivenkata@gmail.com

Received July 23, 2015; **Accepted** September 12, 2015; **Published** September 30, 2015

Citation: Agrawal H, Thakur A, Slathia R, Sumangali K (2015) A Comparative Analysis of Social Networking Analysis Tools. J Inform Tech Softw Eng 5: 157. doi:10.4172/2165-7866.1000157

charts utilizing a projection on one hub sort and making edges between these hubs utilizing distinctive conglomeration capacities. The idea of diagram can be summed up by a hypergraph, in which two arrangements of vertices can be associated by an edge.

Visualization

Visualization is a standout amongst the most needed functionalities in chart taking care of projects, and this stays valid for system examination programming. Numerous calculations comprise in pushing confined vertices toward unfilled spaces and in gathering contiguous hubs. These calculations are specifically roused by physical phenomena. Case in point, edges can be seen as springs and hubs can be taken care of as electrically charged particles. The area of every component is recalculated regulated. These techniques require a few cycles so as to give a decent result on extensive charts. Power based formats are easy to grow however are liable to poor neighborhood least results. Among these calculations, we can specify, Fruchterman and Reingold [1], which is an all-around utilized power based calculation for diagram visualization. An option is the calculation of Kamada and Kawai [2], which has a speedier union than Fruchterman and Reingold, yet, which regularly does not give so great results than this last one. It can be conceived to utilize Kamada and Kawai so as to ascertain a first position of the vertices. These two systems are among those called "spring calculations". Some different formats are distinctive in the way they give a perspective of the area for a hub (i.e. outspread design, hyperbolic format). 3d diagram visualization is the legitimate augmentation of planar representations. The majority of the systems proposed are versatile to 3d. Neighborhood zoom based, alleged fish-eye usefulness can be additionally intriguing to outwardly investigate extensive diagrams.

Pointer based system depiction

Numerous quantitative pointers have been characterized on systems. The descriptors at the system level are utilized to look at the extent of hubs versus edges, or to assess properties of the chart like the arbitrariness or little world appropriations. Then again, the descriptors at the hub level are helpful for recognizing the hubs deliberately put in the system or highlighting those that take a critical part in correspondence, for example, extensions or center points.

Vertex and edge scoring: The spot of a given performer in the system can be portrayed utilizing measures in view of vertex scoring. Regular sorts of vertex scoring are the centrality measures. For instance, to gauge how critical an individual is inside an interpersonal organization, Freeman [3] has recognized three principle centralities:

a) Degree centrality: The first and easiest measure is the degree centrality. It underlines hubs with the high degrees.

b) Closeness centrality: For associated charts, closeness centrality is the reverse of the normal separation to every other hub. This marker can be helpful for some applications in this present reality. Case in point, if edges were lanes, the intersection (vertex) with the most elevated closeness centrality would be the best place for crisis administrations.

c) Betweenness centrality: Betweenness centrality is another centrality measure of a vertex inside a chart. Vertices that happen on numerous briefest ways between different vertices have higher betweenness than those that don't.

d) Page rank: The score registered by Page Rank is higher for hubs that are exceedingly joined and associated with hubs that are very joined themselves.

e) HITS calculation: Hyperlink-induced topic search (HITS, otherwise called center points and powers) figures two scores: center point and power score. The more a vertex has active circular segments, the higher is its center point score. The more a vertex has approaching connections, the higher is its power score. Toward the starting each hub are considered as center and power scores are settled to a steady. At that point the scores are upgraded and they focalize after couple of cycles.

System scoring: System thickness is the rate of edges in the system over the quantity of edges that could exist in the system. This measure shows if the fundamental diagram is meager or thick. These markers have following been made an interpretation of in renditions relevant to coordinated charts, valuable in data dispersal hypothesis. This asymmetry prompts the idea of notoriety.

a) Dyad census: A dyad is a term obtained from social science used to portray a gathering of two individuals, i.e. the littlest conceivable social gathering. By expansion, it is utilized as a part of informal community investigation for outlining two associating hubs. Every dyad is arranged into one of the common, hilter kilter or invalid classes and the extent of each of these cases is given. These checks help to know whether the connections take after an irregular or a small world conveyance.

b) Triad census: with a specific end goal to augment the dyad check, Davis and Leinhardt [4] have proposed the triad number, with 16 particular cases (coordinated diagrams). Triadic examination performs the include of the triads every setup. Data gave is again valuable to contrasting a system and the irregular model.

Diagram and vertices closeness: In interpersonal organization investigation instruments, one can hope to discover capacities communicating likeness of hubs in a chart furthermore capacities to gauge the similitude between diagrams themselves. A few samples of similitude measures accessible in virtual products are the Jaccard, Dice or Tanimoto likeness.

Grouping or group location

The point of grouping is to identify gatherings of hubs with thick associations inside the gatherings and sparser associations between the gatherings. These gatherings are called groups by analysts and information mining experts while sociologists like to utilize the word groups. An extremely finish study on diagram grouping can be found in.

Fundamental methodologies of group discovery: Among the diverse systems proposed to distinguish groups, two primary methodologies can be recognized: from one perspective there is the progressive approach in which the hubs are totaled in a chain of importance of bunches from the discrete part to the entire system. This methodology assesses the vicinity between two hubs through a similitude measure and fabricates the gatherings utilizing an agglomerative procedure, similar to the single linkage calculation or the complete linkage calculation. Then again, there is the partition bunching which comprises in specifically isolating the system into a predefined number of gatherings. The base cut strategy is a case of this methodology in which the gatherings are characterized so as the quantity of edges between them is minimized. The virtual products considered in this benchmarking incorporate three grouping strategies. The first is the Newman [5] and Givan system. This is a progressive system, in view of the betweenness of the edges, which comprises in evacuating the edge with most astounding betweenness, and rehashing this procedure until no edge remains. The second strategy, called Walk

trap is a partition calculation that uses an arbitrary stroll in the diagram keeping in mind the end goal to recognize the segments in which the walker has a tendency to remain. A progressive bunching is then performed to get the bunches. The last calculation is called spinglass. With various leveled techniques, a dendograms the best representation for picking the quantity of groups to hold. Another approach to focus the quantity of gatherings that must be held comprises in expanding a specific criteria, for example, measured quality.

Benchmarking

Numerous apparatuses have been made for system investigation and visualization purposes. An extensive rundown of devices is accessible on Wikipedia1, with altogether different methodologies. Numerous are absolutely scholarly programming. Some are situated toward visualization, other comprise in APIs permitting chart and hypergraph displaying with once in a while the likelihood of movement on vertices, for example, JUNG. A few apparatuses are enhanced for expansive information control. Others propose low level usage of particular calculations. Five instruments are considered: Pajek, Gephi, Netlytic, GraphViz and Social Network Visualizer. The decision of them is taking into account:

– A harmony between settled devices and more up to date ones, in view of late advancement gauges (regarding ergonomics, measured quality and information convenience),

– A SNA perspective. The devices must give fundamental measurements to systems,

– The systems size can achieve countless hubs.

Pajek is a legacy programming, with its own chart situated methodology. Gephi speaks to a present day answer for diagram study with GUI (graphical client interface), open source logic and plugin introduction. Networkx and igraph are two key libraries for proficient huge chart taking care of. The accompanying segments portray the dataset and the criteria utilized as a part of the benchmark (Figures 1-5).

Dataset

The dataset considered in this overview is broadly utilized

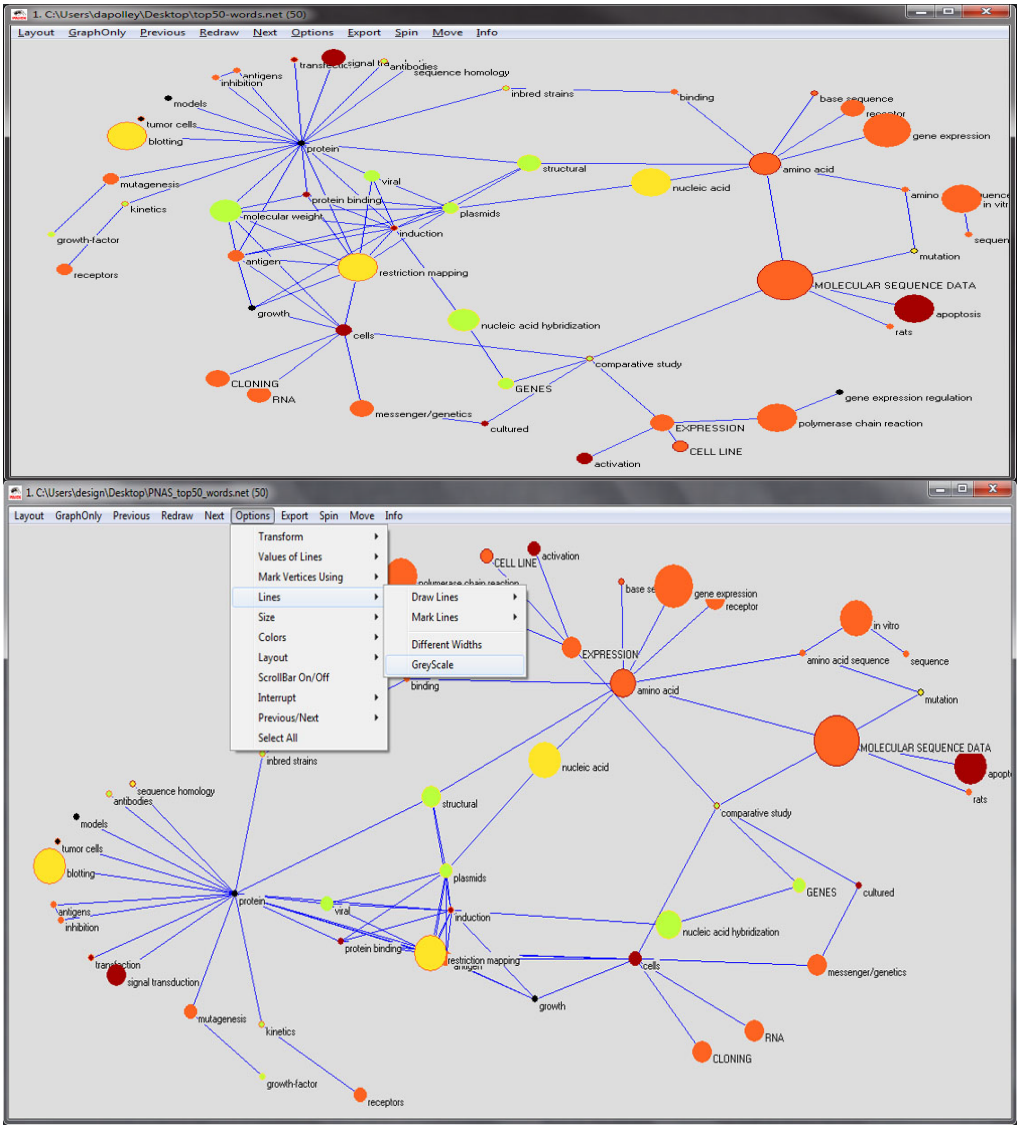

Figure 1: Pajekscreenshots.

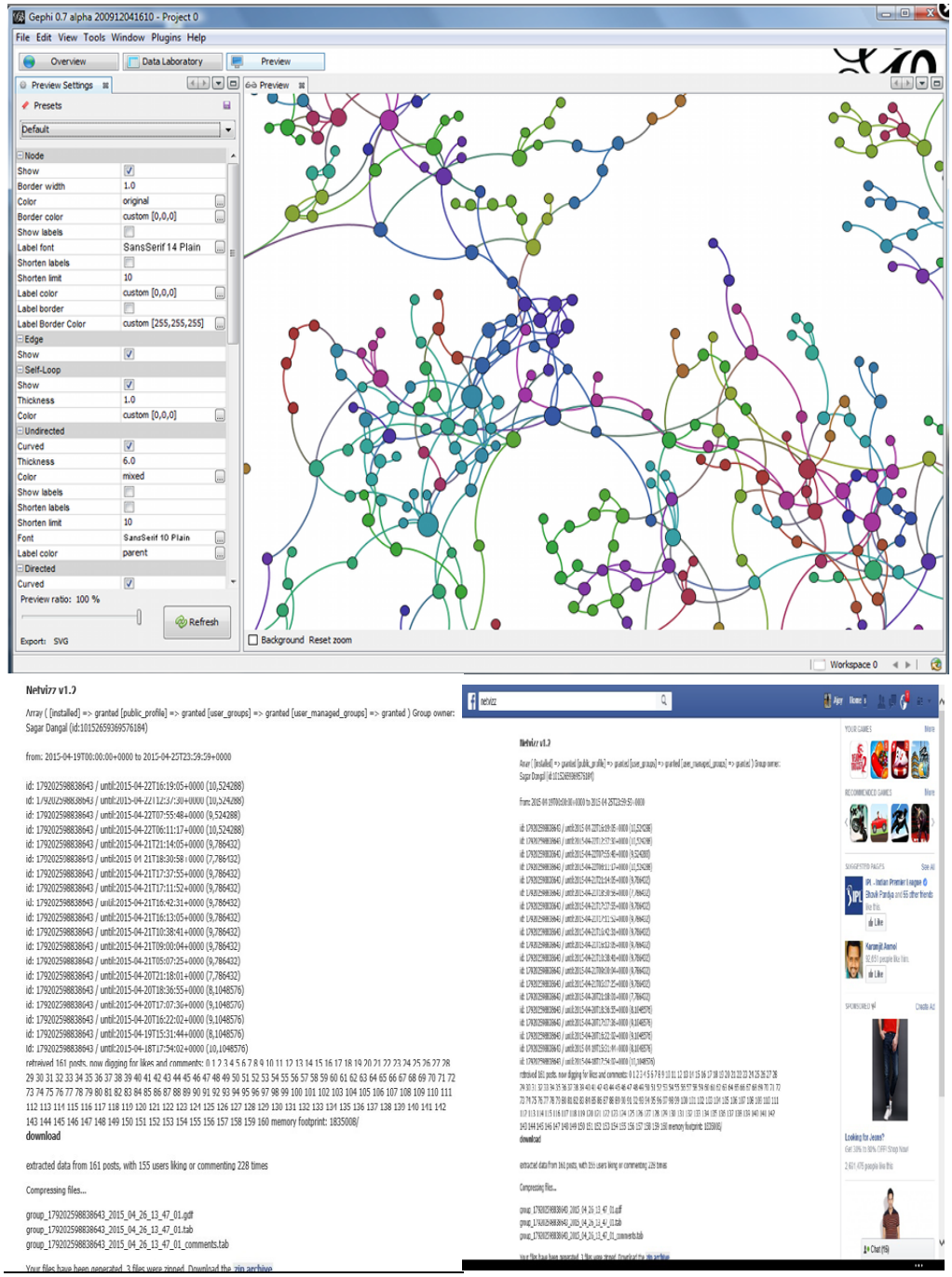

Figure 2: Gephi and Netvizscreenshots.

information set as a part of SNA writing. This dataset presents the connection diagram between 34 individuals from the karate club of a US college in 1970. Zachary's Karate Club2 has 34 vertex and 78 edges. Every vertex is numbered. An edge is available between two hubs when the two relating people "reliably connected.

Assessed criteria

In our benchmark, we have chosen an arrangement of assessment criteria. These criteria are the permit of the instrument, the information arrangement took care of, the diagram sorts upheld, the measure of hubs that can be loaded in a sensible time, the accessible markers, the bunching calculations included and the visualization formats accessible. Every basis is itemized in the accompanying areas.

Document forms: There are primarily three approaches to express in a serial way the structure of a system:

– Nearness framework (square for coordinated charts, triangular for undirected ones)

– Nearness records (for coordinated diagrams), where the source hub is trailed by the rundown of the hubs that are the focuses of each

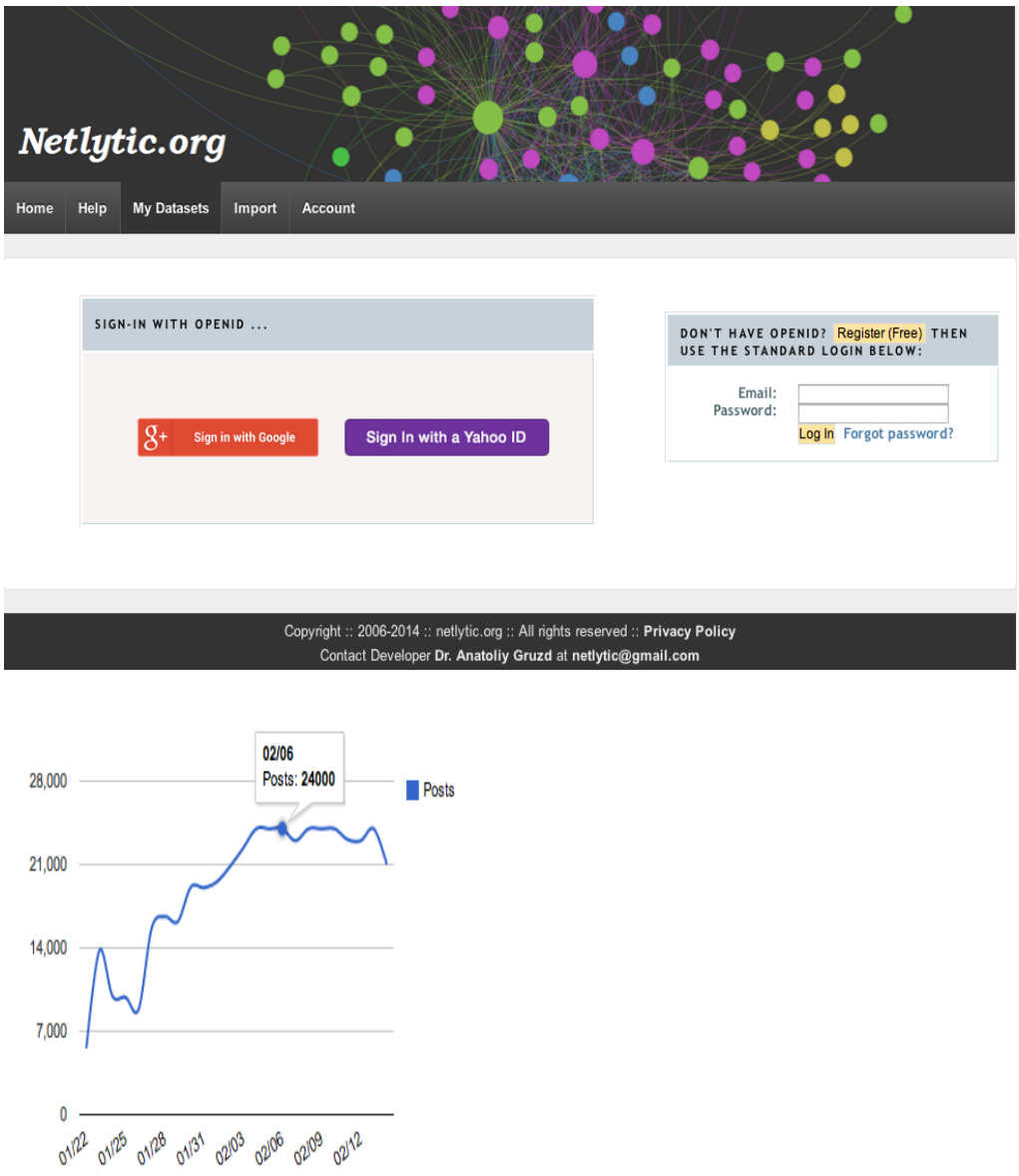

Figure 3: Netlyticscreenshots.

curves beginning from the hub

– Vertices sets. A few document configurations have been made keeping in mind the end goal to give diagram representations.

Here are the primary ones:

a. Pajek chart record configuration (.net augmentation), while not exceptionally very much archived, is extremely famous among interpersonal organization examination devices. It speaks to in a content document, first the vertices (one every line) and afterward the edges. This organization is not regularly taken care of in alternate usage aside from the Pajek program, which permits edge representation with a network or an edge rundown or curve list (for coordinated charts). Weighted systems are permitted. Weights in the discretionary third section are for the bends.

b. GML (Graph Modeling Language) is likewise a structures content document, where hubs and edges start with "node" and "edge" watchwords and their substance is between"["and"]". It permits annotations as content, such as directions for vertices.

GML bolsters:

– Coordinated and undirected diagrams

– Hub and edge names

– Graphical situation of hubs (directions)

– Different annotations

c. GraphML is a XML-based chart portrayal dialect

– coordinated, undirected, and blended charts,

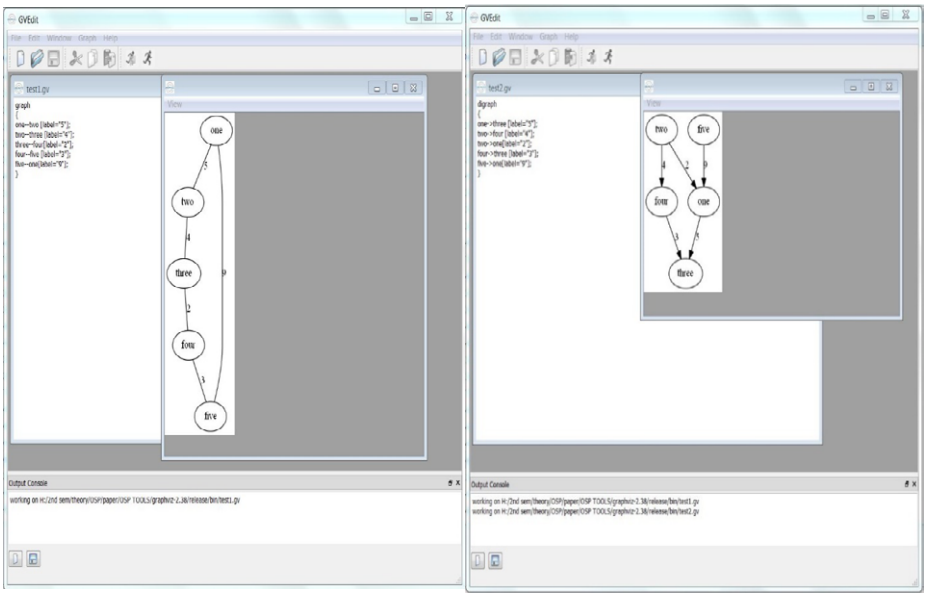

Figure 4: Graphviz screenshots.

Software	Pajek	Gephi	Social Network Visualizer	Netlytic	Graphviz
Version	1.26	0.7 alpha	1.56 Beta	Tier 1,2,3	2.38.0
Type	Stand-alone software	Stand-alone software	Stand-alone software	Stand-alone software	Stand-alone software
Platform	Windows	Java	Windows	Windows	Windows
License	Free	GNU GPL	Free	Tier 1,2 (Free) Tier 3 (CS)	Free
Expectable Computing Time	Fast(C)	Medium(JAVA)	Fast(C)	Medium(JAVA)	Fast(C)
Tractable number of nodes	500000 nodes	150000 nodes	100000 nodes	300000 nodes	1400000 nodes
Time to load 10^5 nodes and 10^6 edges	24 seconds	40 seconds	46 seconds	50 seconds	34 seconds
File formats					
GML	No	Yes	Yes	Yes	No
Pajek(.net)	No	Important Only	No	No	No
GraphML	Export only	Yes	Yes	Yes	No
DL	Yes	Yes	Yes	Yes	No
GEXF	No	Yes	Yes	Yes	No
Graph types					
Two-mode graphs	Yes	No	No	No	Yes
Multi-relational graphs	Yes	No	No	Yes	Yes
Temporality	Yes	No	No	Yes	Yes
Visualization layouts					
FruchtermanReingold	Yes	Yes	Yes	Yes	No
Kamada Kawai	Yes	Yes	No	No	Yes
Other spring layouts	No	Yes	Yes	No	Yes
Indicators					
Degree centrality	Yes	Yes	Yes	Yes	Yes
Betweenness centrality	Yes	Yes	Yes	Yes	Yes
Closeness centrality	Yes	Yes	Yes	Yes	Yes
Dyad census	No	No	No	No	No
Triad census	Yes	No	No	No	No
HITS	No	Yes	Yes	No	No
Page Rank	No	Yes	Yes	Yes	No
Clustering Algorithms					
Edge Betweenness	No	No	No	Yes	No
Walktrap	No	No	No	Yes	No
Spinglass	No	No	No	Yes	No
Dendogram Display	Yes	Yes	Yes	Yes	Yes

Table 1: Comparative analysis of Social Networking Analysis tools.

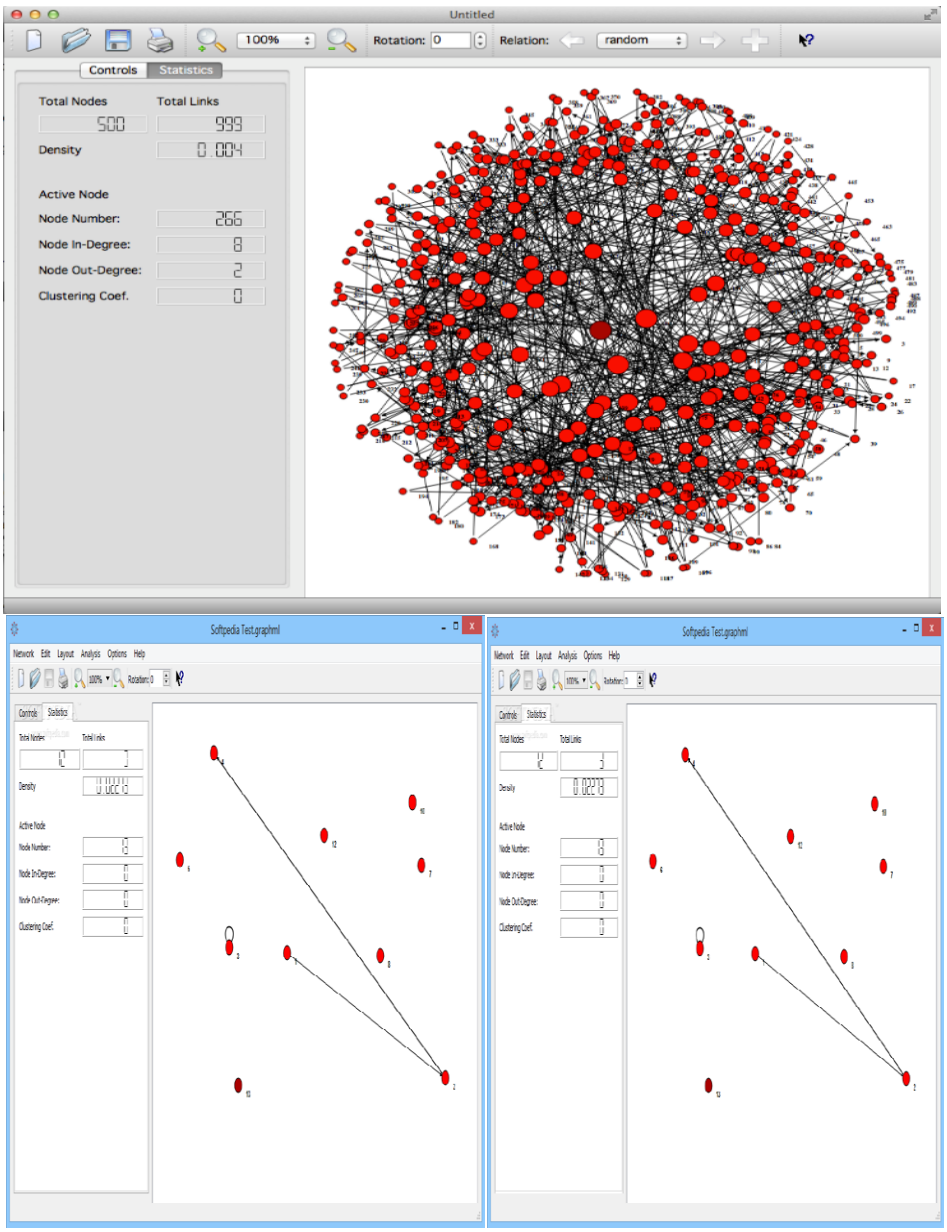

Figure 5: Social Network Visualizer screenshots.

– Hypergraphs,

– Various leveled charts,

– Graphical representations, and

– Application-particular characteristic information.

As all XML-based representation, it is very much a verbose one.

d. *DL* (Data Language) arrangement originates from the Ucinet program. The normal augmentation for this arrangement is .dat. A sample is given Figure

– Edge representation with a full grid, a halfmatrix, a bends rundown or an edges list,

– Record names,

– Rectangular networks for two-mode systems.

e. DOT is another famous chart depiction dialect, took care of mostly by Graphviz.

f. The Fruchterman-Reingold Algorithm is a force-directed layout algorithm. The idea of a force directed layout algorithm is to consider a force between any two nodes. In this algorithm, the nodes are represented by steel rings and the edges are springs between them.

g. A dendrogram (from Greekdendro "tree" and gramma "drawing") is a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering. Dendrograms are often used in computational biology to illustrate the clustering of genes or samples.

h. Kamada and Kawai is a force directed layout algorithm. The idea of using only spring forces between all pairs of vertices, with ideal spring lengths equal to the vertices' graph-theoretic distance.

i. GEXF4 is a XML-based organization, from the GEXF Working Group. It supports

– Element diagrams,

– Application-particular property information, through the utilization of clients XML namespaces,

– Progressive structure (hubs can contain hubs)

– Visualization and situating data, for example, 3D directions, hues, shapes

Assessed instruments

Five instruments have been thought about: Pajek [6], Gephi [7], Netlytic [8], Social Network Visualizer [9] and Graphviz [10].

The benchmarking results are compressed in Table 1. They are point by point in this area, taking after the assessment criteria presented already: the permit of the instrument, the information organization took care of, the diagram sorts upheld, the accessible markers, the bunching calculations included and the visualization designs accessible. The main point is authorizing. It creates the impression that Social Network Visualizer [9] has the most lenient permit, permitting joining in exclusive programming. Both GraphViz [10] and Gephi have picked GNU GPL which does not permit the reconciliation in restrictive programming. Pajek [6] source code is undisclosed and the utilization of the product for business utilization is not free. In matter of information organization, Gephi handles all the organizations said here. GEXF is not accessible somewhere else predominantly in light of the fact that this arrangement began in the Gephi venture. DL accompanies UCINET; this last one being a task connected to Pajek, it is one of the favored configurations for this device. GML and GraphML are not upheld in Pajek, so you can lean toward the .net arrangement, which is widespread in our board. Concerning the bipartite charts study and their control, most devices propose a couple of primitives, for example, projection (change of a bipartite chart into a one-mode diagram), however we would not suggest Gephi [7] for that as two-mode diagrams is not entirely two-mode diagram empowered. Pajek can deal with connections from various types. The transience begins being considered in diverse undertakings. For the present, the information can be sifted in capacity of a year related to the hubs for instance, if the information organization is adjusted. The instrument showing up as the less proficient in matter of permitted vertices in memory is Gephi. After 200,000 hubs on our reference PC (Intel Core 2 Duo 2.5 GHz, 2 Go RAM, Windows), a few mistakes or messages welcome to expand the devoted memory for the virtual machine appear. The visualization sheet is an essential piece of Gephi, while alternate instruments can handle pointers autonomously of drawing the Graph. Such a structural engineering could punish the application for this standard. Pajek does not languish over this point and can stack 500,000 in 52 minutes. The igraph is quick for information stacking (22 seconds for 2.9 a huge number of hubs, however the dataset was sans characteristic (no name for hubs gave, as .net import is very confined for this apparatus). Gephi and Social Network Visualizer seem, by all accounts, to be restricted in their ability by the RAM utilization. Informal organization Visualizer is moderate for stacking 100,000 hubs, yet the stacking is sensible past. A few highlights, for example, administration of multi-diagrams can be the reason for debased execution. The five virtual products are suitable for figuring basic pointers, for example, chart measurements, degree centrality, closeness centrality and betweenness centrality (igraph and NetworkX executions of betweenness centrality are taking into account the calculation from Brandes [11]) Dyad and triad enumeration are accessible in igraph and Pajek (for triad statistics). For HITS and files you cannot depend on Pajek which is not up and coming. In the event that you have to make your own pointers, the two libraries and Gephi are valuable. Group identification is exploratory in Gephi with a beta rendition of Markov bunch calculation (MCL) while couple of calculations is accessible in igraph. Pajek offers progressive grouping abilities. It can give a dendogram representation of a progressive bunching, as an EPS (PostScript) picture. The igraph offers the dendogram plotting abilities of R. On the off chance that you need propelled visualization, you need to change your information to another stage. The three different apparatuses perform the Fruchterman Reingold and Kamada and Kawai well known power based calculations. Be that as it may, the bunching calculations are missing. Hubs and edges can be any sort of articles (the main condition is to give a hash capacity to it). Utilizing programming dialects it makes simple to rethink protests, for example, hubs keeping in mind the end goal to handle them as subjective items. It has likewise some intriguing capacities on the off chance that you utilize bipartite diagrams. The igraph offers numerous calculations among which some bunching focused ones. It is accessible for both Python and R situations, and C libraries are accessible also. With R, it is anything but difficult to incorporate igraph schedules in a measurable methodology. Numerous functionalities are as of now upheld, however a few calculations are missing [12].

Other intriguing programming for interpersonal organization investigation

There are numerous other SNA apparatuses accessible, in this paper some of them were tried, for example,

– Tulip can deal with more than 1 million vertices and 4 millions edges. It has visualization, grouping and augmentation by modules capacities.

– UCInetis not free. It utilizes Pajek and Netdraw for visualization. It is had practical experience in factual and material investigation. It computes markers, (for example, triad evaluation, Freeman betweenness) and performs various leveled bunching.

– JUNG, for Java Universal Network/Graph Framework, is basically created for making intuitive diagrams in Java client interfaces, JUNG has been stretched out with some SNA measurements.

– GUESS is committed to visualization purposes. It is distributed under the GPL permit. The reasons why different apparatuses haven't been definite above are:

– Their tight and specific functionalities centered on a solitary viewpoint, i.e. Figure on visualization,

– Truthfully supplanted by different devices with the same target highlights and gathering of people (Tulip with Gephi),

– Are not centered around a software engineering vision,

– Are not freely available.

Conclusion

The way that Social Network Analysis is arranged between a few areas (social science, software engineering, math and material science) has prompted various methodological methodologies and to a ton of apparatuses. That is the reason such a variety of projects has been made with a specific end goal to control. While a standalone programming is extremely helpful for diagram visualization (up to a greatest of couple of a great many hubs), information design change or pointers processing, libraries are more adjusted for undertakings including a

huge number of hubs and for operations, for example, the union and the contrast between sets of hubs or for the grouping. A reasonable partition of the calculations, the client interface and the visualization sheet is critical. Gephi received this methodology with the late arrival of the Gephi tool compartment, a library made from the Gephi rationale and calculations. We can likewise say that today the uninhibitedly accessible apparatuses have the capacity to give an exceptionally rich arrangement of functionalities, however in the event that one needs particular investigation, a business programming or correlative code improvements may be required. At long last right now, the fundamental difficulties concerning the diagram investigation are arranged toward abnormal state visualization (i.e. progressive diagrams), while amongst the conceivable improvements of informal organization investigation devices, we can specify firstly the fleeting examination which ought to permit to study the advancement of systems over the long run, and furthermore social mining which at the same time misuses the qualities of hubs and the chart structure.

References

1. Fruchterman TMJ, Reingold EM (1991) Graph Drawing by Force-directed Placement. Software: Practice and Experience 21: 1129-1164.
2. Kamada T, Kawai S (1989) An algorithm for drawing general undirected graphs. Information processing letters 31: 7-15.
3. Freeman LC (1979) Centrality in social networks conceptual clarification. Social networks 1: 215-239.
4. Davis JA, Samuel L (1967) The Structure of Positive Interpersonal Relations in Small Groups. Sociological Theories in Progress 2: 1-52.
5. Newman MEJ, Girvan M (2004) Finding and evaluating communitystructure in networks. Physical review E 69.
6. http://vlado.fmf.uni-lj.si/pub/networks/pajek
7. http://gephi.github.io
8. https://netlytic.org/home
9. http://socnetv.sourceforge.net
10. http://www.graphviz.org
11. Brandes U (2001) A Faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25: 163-177.
12. Combe D, Largeron C, Zsigmond EE, Gery M (2010) A comparative study of socialnetwork analysis tools. International Workshop on Web Intelligence and Virtual Enterprises 2.