INTERNATIONAL REVIEW of SOCIAL RESEARCH

International Review of Social Research

An overview of Software Applications for <u>Social Network Analysis</u>

Ioana-Alexandra APOSTOLATO[·]

Department of Sociology
University of Bucharest

Abstract: There is a great variety of software tools that has been developed within the last 20 years, as to facilitate and support the qualitative and quantitative analysis of social networks. This paper gives a brief overview of some of the most popular software packages for social network analysis: Pajek, UCINET 6, NetDraw, Gephi, E-Net, KeyPlayer 1, StOCNET and Automap. Pajek has efficient algorithms for the analysis of large networks, while UCINET 6 includes multiple analytical tools highly efficient for exploring and measuring social network structures. NetDraw, nested in UCINET 6, and Gephi allow network visualization. E-Net and KeyPlayer 1 satisfy rather specific and well-oriented purposes: ego-network analysis and network key-player operations (node removal or utilization). StOCNET provides a platform for statistical methods focusing on probabilistic models, while Automap is a text mining tool for analyzing text relational data.

Keywords: *Pajek, social network analysis, software packages, UCINET, visualization.*

There is a great variety of software tools that has been developed within the last 20 years, in order to facilitate the qualitative and quantitative analysis of social networks. An extended presentation of social network analysis (SNA) software packages is available on a special web page constructed by Mark Huisman and Marijtje A. J. van Duijn[1], a website accompanying their chapter published in Scott and Carrington (2011).

Basically, software packages can be classified in several categories. Some packages are used for academic purposes (e.g. applied graph and network analysis, exploratory analysis, dynamic network analysis), others for relational data visualization (e.g. graph visualization, visualizing virtual communities, visualizing molecular interaction networks, tools for topology display etc.). Furthermore, there are several software packages especially designed for investigating other structures than social networks

[·]e-mail: ioana.apostolato@yahoo.com. Ioana Alexandra Apostolato is MA student of Research in Sociology at University of Bucharest, Department of Sociology and holds a BA degree in Sociology. She currently works as a program officer at the Lab for Empirical Operations and Network Analysis (LEONA). Her main research stream is focused on analyzing the effects of informal intra-organizational structures.

(e.g. network text analysis, ecological network analysis, simulation of evolutionary environments etc.).

In the following section, I introduce some of the most popular software packages within the SNA community: Pajek, UCINET 6, NetDraw, Gephi, E-Net, KeyPlayer, StOCNET, Automap. This overview, yet limited to only 8 software packages, is intended to bring forth the power that software packages have in supporting the extensive study of social networks or similar relational data sets. Beginner's tutorials and advanced courses on SNA using software packages are yearly organized during the Sunbelt Conferences of the International Network for Social Network Analysis. In Romania, intensive introductory courses into SNA using either UCINET or Pajek are yearly organized by University of Bucharest, Department of Sociology, as special sessions of the Romanian Social Network Analysis Research Program for Undergraduates (SoNAR)[2].

Pajek

Pajek (the Slovene word for *spider*), a free software application for analysis and visualization of large networks, has been developed by Vladimir Batagelj and Andrej Mrvar. According to Borgatti (1998), the main goal in the design of this software is to provide the user powerful visualization tools, efficient algorithms for analyzing large networks and abstraction by factorization of large network into smaller networks. Pajek has six data structures to implement the algorithms: network, permutation, vector, cluster, partition and hierarchy, and it is based on transformations that support transitions between this data (Batagelj and Mrvar, 1998, 2011).

By using the data structures above mentioned, the software has a set of algorithms as it follows:

- Partitions: *degree, depth, core, p-cliques, centers;*
- Binary operations: *union, intersection, difference;*
- Components: *strong, weak, bi-connected, symmetric;*
- Decomposition: *symmetric-acyclic;*
- Paths: *all paths between two vertices, shortest path(s);*
- Flows: *maximum flow between two vertices;*
- Citation weights;
- Neighborhood: *k-neighbors;*
- Critical Path Method;
- Extracting sub-network;
- Shrinking clusters in network;
- Reordering: *topological ordering, Richard's numbering, depth/breadth first search;*
- Reduction: *hierarchy, subdivision, degree;*
- Simplifications and transformations: *multiple lines, deleting loops, transforming arcs to edges.*

The software also has a draw window tool, where users can edit a network by hand, select a part of the picture, edit lines that belong to the selected vertex or spin the picture. The layout of the network can be generated by itself, and the picture can be in 2D or 3D format. Pajek provides tools for analysis and visualization of large networks like: collaboration networks,

genealogies, Internet networks, citation networks, diffusion networks (news, innovations), data-mining (2-mode networks). Updated information and details on the development of Pajek are available on the Pajek Wiki[3].

UCINET 6

UCINET 6, a Windows software package used for the analysis of social network data, has been developed by Steve Borgatti, Martin Everett and. Lin Freeman (2002). This package provides the tools to analyze 1-mode or 2-mode data, and it can handle a maximum of two millions nodes, but in practice most of the procedures are too slow to run networks larger than 5000 nodes.

UCINET contains dozen of network analytical tools, such as: centrality measure, subgroup identification, role analysis, elementary graph theory and permutation-based statistical analysis. The software can perform a wide variety of data transformation such as: sub-graphs and sub-matrices, merging datasets, permutation and sorts, transposing and reshaping, recodes, linear transformations, symmetrizing, geodesic distances and reachability, aggregation, normalizing and standardizing, mode changes.

In addition, the package has matrix analysis routines, such as matrix algebra and multivariate statistics. Updated information and details on the development of UCINET are available on the UCINET Software page[4].

NetDraw

NetDraw is a program developed by Steve Borgatti (2002) for visualizing 1-mode and 2-mode social network data. The program reads UCINET files, UCINET DL files, Pajek files and its own VNA format. The VNA data format allows the user to store not only network data but also attributes of the nodes, along with information about how to display them (color, size, etc.). A key feature of the VNA format is that it admits textual data instead of numeric codes. The DL protocol is a flexible language for describing data and itself encompasses a number of different formats such as: nodelist, edgelist and fullmatrix. DL data format is usually the most efficient format. The user can just create a text file using any word processor. The software has some analytical capabilities that partially overlap with UCINET.

NetDraw allows the user to save network and attribute data together along with layout information like colors or spatial coordinates and it can also handle multiple relations at the same time. By using the attribute data, you can set the color, the shape and the size of the nodes. The picture can be saved in metafile, jpg, gif and bitmap formats. Some of the features of this software include: multiple relations, valued relations, node attributes, social network analysis, 2-mode data, appearance options, layout. Documentation, FAQs and other useful information on NetDraw are available on the software's webpage[5].

Gephi

Gephi is an open interactive visualization and exploration platform, for all kinds of networks, dynamic and hierarchical graphs. The software runs on Windows, Linux and Mac OS X. The goals of this application are to help data analysts to make hypothesis, discover patterns, isolate structure singularities or faults during data sourcing. Gephi can handle networks up to 50.000 nodes and 1.000.000 edges. Layout algorithms give the shape to the graph and allows user to change layout settings while running. Gephi provides two types of algorithms: force-based algorithms and multi-level algorithms.

The statistics and metrics framework offer metrics for social network analysis and scale- free networks, as following: betweenness, closeness, diameter, clustering coefficient, average shortest path, pagerank, HITS, community detection (modularity), random generators. The user can customize the color, the size and the labels for a better network representation.

Gephi provides features in order to explore large, hierarchically structured graphs, social communities, biochemical pathways or network traffic graphs. The software can also filter dynamic structures, such as social networks with the timeline component. Documentation, FAQs and useful information Gephi's Wiki, Forum or Blog are available on the software's webpage[6].

E-Net

E-Net, a Windows software for analyzing ego network data, was written by Steve Borgatti (2006). The program accepts data pertaining to the egos and to the alters, such as age or sex and it also accepts ties among the alters. E-Net can calculate a number of indices in order to examine the composition of ego networks, the heterogeneity of ego networks, homophily and structural hole measures. In other terms, it calculates how diverse each person's network is and the tendency of egos to seek similar alters. The numerical analyses are automatically performed on all egos.

The network data can be imported in two formats: row-wise and column-wise, but it also accepts full network data stored in the UCINET data file format. E-Net includes multiple analysis options for ego-network data: compositional measures, structural characteristics (*density, structural holes*), data filtering, longitudinal analysis.

The software also has the ability to create crosstabs of ego variables against alter variables. The key feature of E-Net is its filtering capabilities, allowing users to select which egos, which alters and which relations should be active in any given analysis. Documentation, FAQs and other useful information on E-Net are available on the software's webpage[7].

KeyPlayer

Keyplayer is a program developed by Steve Borgatti (2003), for identifying an optimal set of nodes for one of two purposes: crippling the network by removing key nodes or electing which nodes to either keep under surveillance or to try to influence via some kind of

intervention. In the KeyPlayer menu, this basic distinction is reflected in the two menu choices under Analyze, namely *Remove* and *Observe*. There are two criteria for crippling a network: one is fragmenting the network into disconnected pieces and the other to lengthening distances between all pairs of nodes.

Through the instrumentality of the *Fragment* procedure the program will ask how many nodes you would like to remove, how many 'starts' you want, and the maximum number of iterations. The measure of fragmentation optimized by the program is based on the heterogeneity coefficient used in sta-tistics. The objective of the *Distance* feature is lengthening the average distance between pairs of nodes by logically deleting key nodes. KeyPlayer only has one option under *Observe*, which is called *Reach*. The idea of reach is to find a set of nodes who are linked to as many distinct others as possible. The software in incorporated into the UCINET package. Documentation, FAQs and other useful information on KeyPlayer 1 are available on the software's webpage[8].

StOCNET

StOCNET is an advanced software system for statistical network analysis that provides a platform to make a number of statistical methods, focusing on probabilistic models. StOCNET's main features consist of: execution of different statistical (stochastic) methods; calculation of some common descriptive network statistics; data transformation and/or selection; simulation possibilities.

StOCNET includes interfaces for the following five programs for statistical modeling of social networks:

SIENA - for the statistical estimation of models for the evolution of social networks;

BLOCKS - designed for stochastic block modeling of relational data;

PACNET - for the estimation of models for testing effects of actor variables and dyadic variables on the ties in a network, controlling for reciprocity and for the dispersion of the in- and of the out-degrees;

ULTRAS - for the analysis of undirected network data using ultra metric (i.e. hierarchical clustering) measurement models;

ZO: for the analysis of directed and undirected graphs with given degrees.

A more detailed view on StOCNET project is available on software package's web page[9]. Furthermore, there is also a users' group[10] for information exchange and technical support on using StOCTNET.

Automap

AutoMap is a text mining tool develo-ped by Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon for extract analyze and represent relational data from texts using Network Text Analysis methods.

AutoMap supports the extraction of several types of data such as: content analytic data (words and frequencies), semantic network data (the network of concepts), meta-network data (the cross classification of concepts into their ontological category such as people, places and things and the connections

among these classified concepts), and sentiment data (attitudes, beliefs). Automap is designed: a) to extract, analyze and compare mental models of individuals and groups; b) to reveal structure of social and organizational systems from texts.

AutoMap also offers a variety of techniques for pre-processing Natural Language: Named-Entity Recognition; Stemming (Porter, KStem); Collocation (Bigram) Detection; Extraction routines for dates, events, parts of speech; Deletion; Thesaurus development and application; Flexible ontology usage; Parts of Speech Tagging. AutoMap uses parts of speech tagging and proximity analysis to do computer-assisted Network Text Analysis. Network Text Analysis encodes the links among words in a text and constructs a network of the linked words.

Documentation, FAQs and other useful information on Automap are available on the software's webpage[11].

Notes

[1] See http://www.gmw.rug.nl/~huisman/ sna/software.html Retrieved: October 20, 2013.

[2] See http://gabrielhancean.wordpress. com/sonar-2013/ Retrieved: October 20, 2013.

[3] See http://pajek.imfm.si/doku. php. Retrieved: October 20, 2013.

[4] See https://sites.google.com/site/ ucinetsoftware/home Retrieved: October 20, 2013.

[5] See https://sites.google.com/site/ netdrawsoftware/home Retrieved: October 20, 2013.

[6] See https://gephi.org/ Retrieved: October 20, 2013.

[7] See https://sites.google.com/site/ enetsoftware1/ Retrieved: October 20, 2013.

[8] com/keyplayer/keyplayer.htm Retrieved: October 20, 2013.

[9] See http://www.gmw.rug.nl/~stocnet/ StOCNET.htm Retrieved: October 20, 2013.

[10] See http://groups.yahoo.com/groups/ stocnet/ Retrieved: October 20, 2013.

[11] See http://www.casos.cs.cmu.edu/ projects/automap/ Retrieved: October 23, 2013.

References

Bastian M., S. Heymann, M. Jacomy (2009) 'Gephi: an open source software for exploring and manipulating networks', http://gephi.org/features/ Retrieved: October 20, 2013.

Batagelj, V. and A. Mrvar (1998) 'Pajek: A Program for large Network Analysis'. *Connections*, 21(2): 47-57.

Batagelj, V. and A. Mrvar (2011) 'Pajek: A Program for large Network Analysis and Visualization of Large Networks. Reference Manual', http://vlado.fmf.unilj.si/ pub/networks/pajek/doc/pajekman.pdf Retrieved: October 20, 2013.

Boer, P., M. Huisman, T.A.B. Snijders, C. Steglich, L. H. Y. Wichers and E.P.H. Zeggelink (2006) 'StOCNET: an open software system for the advanced statistical analysis of social networks', http://www.gmw.rug.nl/~stocnet/StOCNET.htm

Retrieved: October 20, 2013.

Borgatti, S.P. (2002). 'NetDraw Software for Network Visualization: Documentation&FAQs', https://sites.google.com/site/netdrawsoftware/ documentation-faqs Retrieved: October 20, 2013.

Borgatti, S.P. (2006) 'Identifying sets of key players in a network'. *Computational, Mathematical and Organizational Theory*, 12(1): 21-34.

Borgatti, S.P. (2003) 'The Key Player Problem'. In Breiger, R. , K. Carley, and P. Pattison (eds.) *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pp. 241-252. National Academy of Sciences Press.

Borgatti, S.P., M.G. Everett, and L.C. Freeman (2002). 'UCINET Software'*,* https://sites.google.com/site/ucinetsoftware/home. Retrieved: October 20, 2013.

Borgatti, SP. (2006). 'E-Network Software for Ego-Network Analysis'*,*   https:// sites.google.com/site/enetsoftware1/documentation. Retrieved: October 20, 2013.

Carley, K., J. Diesner and M. De Reno (2006) 'AutoMap User's Guide', http:// www.casos.cs.cmu.edu/publications/papers/CMU-ISRI-06-114.pdf. Retrieved: October 20, 2013.

Halgin D.S., S.P. Borgatti (2012) 'An Introduction to Personal Network Analysis With E-Net'. *Conections,* 31(1): 37 – 48.