SOCIAL NETWORK ANALYSIS

Autor: Lozovanu Mihaela
Conducător ştiinţific: lect. superior Ciorba Dumitru
Universitatea Tehnică a Moldovei
E-mail:mihaela.lozovanu@gmail.com

Abstract: *Nowadays dealing with social network analysis becomes more and more attractive. It is especially true for the field of agent-based modeling, in which represented actors are connected with other, with which they communicate. First, the paper introduces the definition of a social network and social network analysis and its tendencies. Then it introduces the JUNG(Java Universal Network) Framework, which is an open-source library that provides a common and extendible language for manipulation, analysis and visualization of social network data that can be represented as a graph.*

Cuvinte cheie: *social network, social network analysis, graphs, JUNG, cluster, ranking algorithm.*

1. Introduction

A social network is a map of the relations between individuals which indicates the ways they are connected. This interdependency can be based on friendships, common interests, financial exchange or relationships of beliefs, knowledge or prestige.

From the point of view of data mining, a social network is a heterogeneous and multirelational data set represented by a graph[1]. The graph is typically very large, with nodes corresponding to individuals(actors) and edges corresponding to links representing relationships or interactions between individuals. Both nodes and links have attributes.

Network analysis has been booming and developing for several decades. Network analysis is popular in every kind of academic social science, applied social science areas like marketing, braches of mathematics, computer science. Social network analysis provides its own theoretical statements, methods, social network analysis software and researchers.

Social network analysis aims to measure relationships and flows of information between people, groups, computers and other entities that processes information and knowledge. It offers a mathematical and visual(graphs) analysis of interpersonal relationships. There are several tendencies which characterize and distinguish a social network analysis[2]:

- The groups are the building blocks of society: the approach is open to studying less-bounded social systems, from nonlocal communities to links among websites.
- Rather than treating individuals as discrete units of analysis, it focuses on how the structure of ties affects individuals and their relationships.
- Network analysis looks to see the extent to which the structure and composition of ties affect norms.

This approach is useful for explaining many real-world phenomena, ex. the ability for individuals to influence their success.

2. Using graphs to represent social networks

Social network analysts use two kinds of tools from mathematics to represent information about patterns of ties among social actors: graphs and matrices. In this paper, I will speak about how to use graphs to represent social network data.

There are lots of different kinds of graphs. Bar charts, pie charts, line charts and many other things are called graphs and/or graphics. Network analysis uses one kind of graphic display which consists of points(nodes) to represent actors and lines(edges) to represent ties or relations.[3] Mathematicians know the kind of graphic displays by the names of *directed graphs, signed graphs* or simply *graphs.* The advantages of using graphs is that you can display data in the simplest way possible without having too many numbers to confuse people and also you can record comparison between two or more things, or objects. The graph that is neatly constructed is more attractive than simple figures. Thus, making comparison is made easy and it will save time of the user to make quick comparison of large data.

3. Social network analysis software

Social network analysis software is used for quantitative or qualitative analysis of social networks, by describing features of a network, either through numerical or visual representation.

Network analysis software generally consists of either packages based on graphical user interfaces (GUIs), or packages built for scripting/programming languages. GUI packages are easier to learn, while scripting tools are more powerful and extensible. Widely used and well-documented GUI packages include UCINet, Pajek (freeware), GUESS, ORA, and Cytoscape.[4]

Visual representations of social networks are important to understand network data and convey the result of the analysis. Visualization often also facilitates qualitative interpretation of network data. With respect to visualization, network analysis tools are used to change the layout, colors, size and other properties of the network representation.

4. Visualizing social network data with JUNG

JUNG – the Java Universal Network/Graph Framework – is a software library that provides a common and extendible language for the modeling, analysis and visualization of data that can be represented as a graph. It is written in Java, which allows JUNG – based applications to make use of the extensive built-in capabilities of the Java API, as well as those of other existing third-party Java libraries.

The major features of JUNG in social network analysis include the following[5]:

- Support for a variety of representations of entities and their relations, including directed and undirected graphs, multi-modal graphs (graphs which contain more than one type of vertex or edge), graphs with parallel edges (also known as multigraphs), and hypergraphs (which contain hyperedges, each of which may connect any number of vertices).
- Implementations of a number of algorithms from graph theory, exploratory data analysis, social network analysis. These include routines for clustering, random graph generation, statistical analysis, and calculation of network distances, flows, and ranking measures.
- A visualization framework that makes it easy to construct tools for the interactive exploration of network data. Users can choose among the provided layout and rendering algorithms, or use the framework to create their own custom algorithms.

5. JUNG algorithms

JUNG provides several different categories of network algorithms. One of them are ranking algorithms. These algorithms assign values to each vertex (or edge) according to a set of criteria that reflect structural properties of the network. These criteria are generally intended to measure the “influence”, “authority”, or “centrality” of a given vertex/edge.[6]

Ranking algorithms that JUNG provides include *BetweennessCentrality*(), which labels each vertex and edge in a graph with a value that is derived from the number of shortest paths that pass through it; *PageRank()*, which ranks each vertex in a modified Markov network according to its stationary probability; *PageRankWithPriors()*, a generalization of PageRank() whose ranks are calculated relative to a specified set of root vertices. Fig.2 shows a visualization of a network whose vertices have been ranked by PageRankWithPriors().

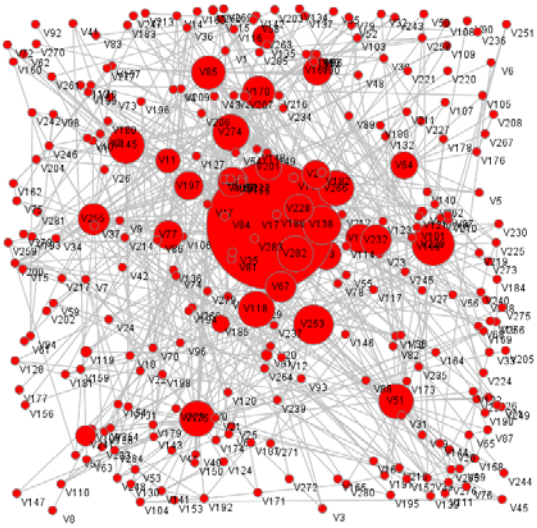

Fig.2

Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. In a network, similarity is often based on topological properties such as connectivity, but can also be based on the properties of vertices or edges in the network. Clustering algorithms provided by JUNG include *EdgeBetweennessClusterer()*, which computes clusters for a graph based on the betweenness property of the edges; *WeakComponentClusterer()*, which finds all weak components in a given graph, where a weak component is defined as a (maximal) subgraph in which each pair of vertices is connected by at least one undirected path; and *VoltageClusterer(),* which assigns vertices to clusters based on their tendency to have similar voltages in the network, treated as an electrical circuit.

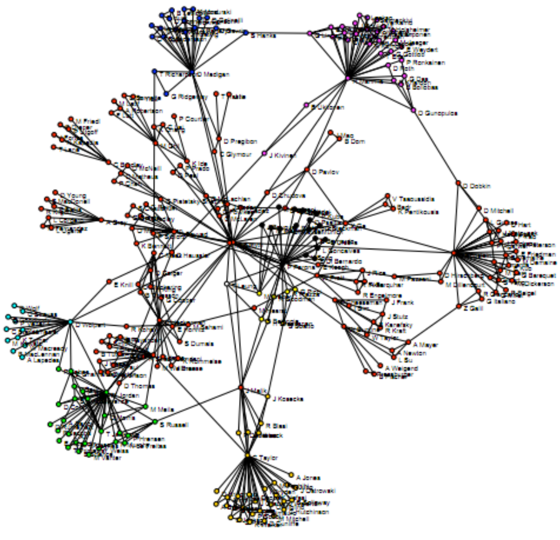

Fig.3

Fig.3 shows a subgraph of the CiteSeer(digital library for scientific and academic papers) graph whose vertices have been clustered by VoltageClusterer(). Each color represents a distinct cluster.

6. JUNG in comparison with other tools

There exist numerous other packages and tools for visualizing and manipulating networks. UCINET and Pajek are stand-alone applications that each provide a number of tools for visualizing and analyzing networks. So, they cannot be conveniently addressed programmatically by other applications. In comparison with JUNG they are applications rather than libraries, so users cannot write their own routines. This is particularly problematic for analysts that wish to use methods that are not provided as part of the application. Also, comparing to JUNG they are closed-projects and executables are available only for Windows environment.

R is a specialized programming language designed primarily for statistical computing and graphics. The sna package extends R in somewhat the same way that JUNG extends Java. Like JUNG, sna provides a number of routines for social network analysis and visualization. R is well-suited for the rapid development of scripts and for on-the-fly network analysis, especially for analysis which requires sophisticated statistical tools. In addition, while it is not technically platform-independent, R is available on several popular platforms (so code written in R is reasonably portable), and is open-source. However, R and sna do have some limitations, particularly for large sparse networks:

- R passes copies of a routine's arguments to it; this requires considerably more space and time than passing references to these arguments (as Java does), especially for large objects.
- R does not have a native sparse matrix format, which can greatly increase the space and time required.

These limitations can be addressed by implementing some of the critical portions of the code in C/C++, but at the cost of additional complexity of code and reduced portability. As a result, R and sna are generally not the most efficient way of manipulating or analyzing networks of more than a few thousand entities. By contrast, while Java is not as well-suited to rapid script development as R, it does provide a convenient platform for the development of more complex tools (for example, those with graphic user interfaces, database connectivity, and/or Web support), and since JUNG's representations and algorithms are both space- and time-efficient, Java and JUNG are more natively scaleable.

References

1. Han J. and Kamber M., *Data mining: Concepts and Techniques,* Elsevier, San Francisco, 2006, p.556-561.
2. Freeman, Linton, *The Development of Social Network Analysis,* Empirical Pres, Vancouver, 2006.
3. Mitrutiu M., *Analiza retelelor sociale,* Timisoara, 2005.
4. http://en.wikipedia.org/wiki/Social_network_analysis_software
5. http://jung.sourceforge.net/index.html
6. O'Madadhain J., Fisher D., Smyth P., *Analysis and Visualization Data using JUNG,* Journal of Statistical Software, 2005, p.1-25.