

Implementation of a Complex Networks Analysis System

Li XiangChao

School of Computer Science and Technology, Soochow
University
Suzhou, Jiangsu, China
1427402043@stu.suda.edu.cn

Lu Xun*

School of Computer Science and Technology, Soochow
University
Suzhou, Jiangsu, China
luxun@suda.edu.cn

Abstract—In recent years, the complex network plays a very important role in many fields of study, a highly efficient software for analysis and visualization of large networks, a large number of vertices, integrates a complex network of fast and efficient simulation of complex network analysis, and provides a Visual user interface platform. In our work, we simulate complex network diagram is generated using Pajek. As the tools and algorithms, it is possible to create user-defined network of vertices and lines have their own different attributes, such as arc/property represents the width of the line weight. And visualization, we are able to better understand and analyze complex network simulation information, before we get.

Keywords—Pajek; NetDraw; Cytoscape; Gephi; complex network

I. INTRODUCTION

Some researcher gave a more stringent definition of complex networks: self-organizing, self-similarity, attract, small world, part or all of the properties in the scale-free network known as complex networks.

In short appeared highly complex networks of complex networks. Its complexity is mainly manifested in the following aspects: 1) complex structure and performance in a large number of nodes, the network structure takes on a variety of different features. 2) network evolution: nodes or connections and disappear. For example, world-wide network, Web pages or links can appear at any time, or broken, resulting in changing network structure. 3) connecting diversity: the difference in weight between the nodes and direction that may exist. 4) dynamic complexity: nonlinear dynamical system may belong to a node set, for example, node state changes over time. 5) nodes: the complex nodes can represent any object, for example, interpersonal relationships constitute a complex network of nodes representing individual, complex network composed of nodes can represent different world wide web pages. 6) multiple integration complexity: complexity over multiple interactions, leading to more unexpected results. For example, the design of a power supply network needs to consider the evolution of this network, its evolutionary process of the topology of the network. When energy transfer between two nodes, they will increase the connection between weight, through continuous improvement of learning and memory performance.

Currently, the complex networks of content includes: the geometric properties of the network, network

formation mechanism of statistical laws of network evolution, model properties on the network, as well as the structural stability of the network, network problems, such as the dynamics of evolution. In the field of natural science, basic measure of network research include: degree (degree) and its distribution characteristics, relevance, degree and distribution characteristics, the shortest distance and distribution, betweenness (betweenness) and its distribution, connecting the size distribution of the group.

Complex networks generally have the following characteristics:

First, small world. Although it describes in simple terms most networks on a large scale but any two sections (top) has a fairly short route between points of facts. In everyday language, it is a reflection of the number of relationships can be small but can connect to the world the fact that, for example, in social networks, the relationship between people and people know each other little, but can find no far from other people. As McLuhan said, Earth is becoming smaller, into a global village, that is, into a small world.

Second, the cluster degree (clustering coefficient) concept. For example, there is always a circle of acquaintances or friends in a social network, in which each Member meet other members. Degree of significance is the degree of networking groups; this is a cohesive tendency of networks. Connected group concept is a reflection of a large network of clusters of small network and interconnected. For example, it can reflect the relationship between friends and another circle of friends.

Third, the power law (power law) concept of distribution. Refers to the network a (section) (equivalent to one individual) relationship with other vertices (expressed in the network) number of contact close to the correlation between the vertices; betweenness is an important part of the global geometry. Betweenness for vertex u meaning among all the shortest paths in the network, after the number of u . It reflects the vertex u (that is, networks that are associated with the individual) influence. Scale-free networks (Scale-free network) mainly reflects the concentration of concentration.

In fact, the complex network concept has been submerged in a variety of disciplines, including social networks. Although there is a wide range of complex network, all complex networks can be described using a common model-figure. Pajek can be used to analyze complex network structures.

It is well known that Visual Tools is one of the key design goals for analysis. NetDraw visualization tools, such as Cytoscape, Gephi, and UCINET, are able to demonstrate certain aspects of complex network characteristics and trends. However, unlike these tools Pajek. Pajek provides six data structures: networks, partitions, arranged in clusters, hierarchies and carriers, which provides users with a set of analysis of large and complex networks of fast and efficient algorithms.

Using Pajek, users can draw a complex network diagram, manual or automatic control of the visualization of complex networks, adjust the layout of complex network diagram, Center, and so on. With this Visual tool, users can visually learn some useful information, from the point of view of rational knowledge and perceptual knowledge, to study the relevant characteristics of the complex network.

II. RELATED WORK

A. Cytoscape

Cytoscape is a general platform for complex network analysis and visualization. It is also an open source tool for visualizing networks and integrating these with any type of attribute data. Cytoscape supports a lot of standard network and annotation file formats including: SIF (Simple Interaction Format), GML, XGMML, BioPAX, PSI-MI, SBML, OBO, and Gene Association.

We first give some examples of the use of Cytoscape's literature, so that we know what the significance of the. Jie Wang¹ et al HOA and Hualian HmiOA chip using this product, published an article, to find out the impact of disease research whether can be integrated from the mRNA and miRNA gene expression regulation network relationship (with BSS symptoms in patients with angina pectoris (biomarker) may be the key factor); the experimental results showed that the 23 miRNA is positive regulation (up-regulated) and 408 genes were negative regulation, the author will all these facts into the online analysis website miRTrail (Laczny² et al in 2012: the integration of mRNA and miRNA chip performance analysis tool) analysis, prediction of miRNA corresponds to the standard of using the microCosm gene (target genes) and cross comparison and 408 gene the intersection of the gene, found a total of 115; and on the basis of prediction of target genes and 408 genes cross comparison and in accordance with the negative regulation degree MiRNA (top five percent), a high genetic group, were selected from 23 miRNA (miR-199a-3p, miR-199a-5p, miR-326, and, miR-423-3p, miR-484, miR-146b-5p,). Finally found the 115 genes and six miRNA painted into the network (Figure 3), we can clearly observed the six miRNA and passing between the gene chip experiment verifies that the regulation of the relationship, identify high interaction frequency (size larger circles) genes. MiR-146b-5p, miR-199a-5p and TP53, CALR gene were selected, and a group of patients (including 15 patients with symptoms, 30 patients with BBS symptoms and 30 patients with non BBS symptoms of angina pectoris) were selected for RT-PCR verification.

In addition, the regulatory role of proteins is the Cytoscape's analysis of the strengths, so in order to observe the interaction between the 115 negative

regulatory genes corresponding to the interaction (PPI, Interaction Protein-Protein), the authors will be FI Reactome plug-in in Cytoscape, the corresponding protein interaction, in order to graphics rendering (Figure four), but also provide a reference for the follow-up study.

B. NetDraw

NetDraw is a free program for visualizing both 1-mode and 2-mode social network data. It can handle multiple relations at the same time, and can use node attributes to set colors, shapes, and sizes of nodes. Pictures can be saved in metafile, jpg, gif and bitmap formats. The program reads UCINET system files, UCINET DL files, Pajek files, and its own VNA format (which allows saving network and attributing data together, along with layout information like spatial coordinates, colors, etc.).

NetDraw supports two kinds of external data import, one is the introduction of the information of the node in the notepad file, the other is to import the current use of the more popular social network analysis software output format of the file (such as Ucinet, Pajek). Of course, if there is no relevant to organize a good description of the node information of the file, you need to study to collect, collate and describe the data information. NetDraw supports a very simple data entry and collation - notepad. The researcher can describe the node information in the notepad file in accordance with the requirements of the NetDraw data description format. Generally speaking, the content of the description is divided into three parts: data node (node represents the property data of the network), properties node (node attribute data) and data tie (node data), but it is not that each file must have these three parts, the research can be described as a selective. Data node, which is used to describe the attributes of the research object represented by the network, properties node and data node are very similar, which is used to describe the node's coordinates, size, color and shape, and data tie, which is used to describe the relationship between nodes and nodes.

C. Gephi

Gephi is an interactive visualization platform for many kinds of networks and complex systems, dynamic and hierarchical graphs. It can be used to interact with the representation, manipulate the structures, shapes and colors to reveal hidden properties. Gephi provides state-of-the-art algorithms layout algorithms, both for efficiency and quality. The Layout palette allows changing layout settings while running, and therefore dramatically increase feedback and experience.

Gephi is an open source free cross platform based on JVM's complex network analysis software, which is mainly used for various network and complex systems, dynamic and hierarchical graph of interactive visualization and exploration of open source tools. Can be used as: exploratory data analysis, link analysis, social network analysis, biological network analysis, etc.. Gephi is a visual tool for information data.

D. Ucinet

Ucinet is a comprehensive package for the analysis of social network data as well as other 1-mode and 2-mode data. It can read and write a multitude of differently formatted text files. Ucinet can handle a maximum of

32,767 nodes although practically speaking many procedures get too slow around 5,000 - 10,000 nodes. Integrated with UCINET is the NetDraw program for drawing diagrams of social networks. In addition, the program can export data to Mage and Pajek.

UCINET software is by California University Erwin (Irvine), a group of network analysts write. The team is now the expansion of the software by Stephen Bogati (Borgatti Stephen), Martin Everett (Everett - Martin) and Linton Freeman (Freeman Linton) composition.

UCINET network analysis integration software, including one-dimensional and two-dimensional data analysis of NetDraw, and the development of the application of three-dimensional display analysis software Mage, etc., and integrated Free for large-scale network analysis of Pajek application software. UCINET software can read text files, KrackPlot, Pajek, VNA, Negopy and other formats of documents. It can handle 32767 network nodes. Of course, from the practical point of view, when the number of nodes between 5000 to 10000, some of the procedures will be very slow.

Social network analysis including central analysis, subgroup analysis, role analysis, based on permutation analysis, etc.. In addition, the software package has very strong matrix analysis functions, such as matrix algebra and multivariate statistical analysis. It is currently the most popular, but also the most easy to use, the most suitable for the new social network analysis software.

III. VISUALIZING WITH PAJEK

Pajek is a large and complex network analysis tools are used to study present a variety of powerful tools for complex non-linear network. Pajek running under the Windows environment, for large networks with thousands and even millions of node analysis and visualization. In Slovenia of Pajek is a spider in the language.

Pajek, six data types (network, partitions, arranged in clusters, vector)-based, providing a set of complex network analysis algorithms, including compute nodes, clustering, calculate the distance between two points, complex network graph traversal finds key to the complex network of paths, many types of complex network diagram, the resulting complex networks, and so on. Pajek integrates a complicated network of simulation algorithm for fast, efficient analysis of complex network also provides a Visual and user-friendly platform, at the same time.

For the small size of the network (the number of nodes is not more than 100), can according to the pajek network. Net file data structure to create: create a text document, first line of the input *Vertices n in which n as the total number of nodes in the network created; in each row of the next input nodes and labels (here can also be what do not enter directly into *Edges *Arcs input and then pajek default node labels for V1, V2,..Vn), should pay attention to is each line can only enter a node and its label, a node on line. Then next in the new line of input *Edges (undirected graph using *Edges, directed graph is using *Arcs), and the other from the connection between the input node. After all the completion of the save, this time to modify the text file to modify the name, directly to the.Net save.

A. Graph Layout

Using Pajek layout/circular order, all nodes can be automatically arranged in a circle, these nodes are equal. In fact, the nodes can be arranged in a loop in 2 different commands, the original order of the files in the order or in an order. In our network, where the node (this is to determine the role node in the network play) defined in the network file. In order to adjust the layout, Pajek also provides a number of methods, for us, the network map automatically expand, enlarge or reduce network complexity, for three dimensional complex network has a circular shape, such as a network diagram as Figure 1 rotation, and element names to the network as shown in Figure 2.

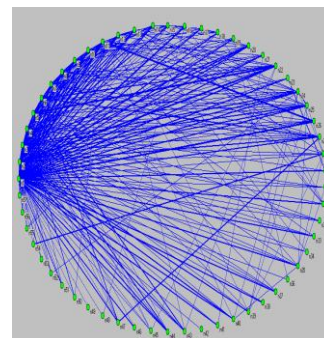


Figure 1. Network with a cycle shape.

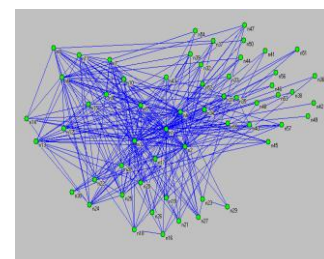


Figure 2. Network with element name.

B. Vertices manipulation

Pajek provides users with a variety of methods to determine each node. We can find, identify nodes by drawing window option/mark vertices using the menu, including the use of node number (number), name (tag) tag or don't do any tag (no label), and sometimes we need some special nodes in the network, we need to sign. However, when we are more concerned with overall marks become redundant. Pajek we can freely on a number of occasions, let us choose. Network node with a different color is displayed as in Figure 3.

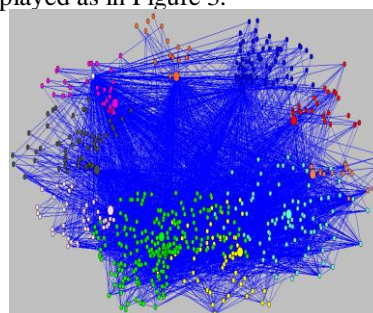


Figure 3. Elements with different colors.

We can also draw different sizes in complex networks using Pajek knots. Large node is more important in a complex network. Therefore, from the diagram, we can visually see the importance level of each node.

C. Lines manipulation

Edge weights for weighted complex network, Pajek also has its own methods and operations. /Line drawing window options menu offers several options, we can choose different ways to express different sides of weight. If you select a complex selection of different widths and edge weights are not the same, width of the line is different, and the greater the weight, wide lines. Hub focused network as shown in Figure 4.

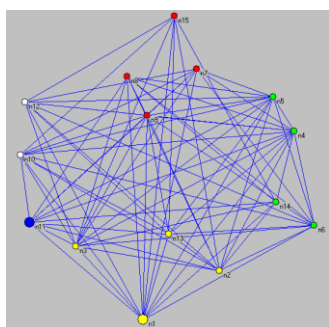


Figure 4. Network constructed after hub centralizing.

D. Case study

Figure you want in the vertices and edges of the network type of information is stored in the file. In fact, we use c # with Visual Studio to generate a network file. Here is part of the source code, here is a network file. Figure 5 is the complex network with thousands of nodes after hub and centrality processing.

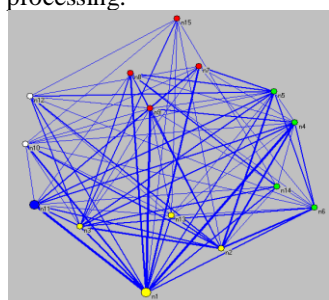


Figure 5. Network with thousands of nodes after hub and centrality.

IV. CONCLUSION

Through our work, we combine the complex network analysis with Pajek-the efficient visual platform very well. We use Pajek-supplied data types and algorithms, import our data with our program and achieve a complex network mapping. And finally we complete some important data analysis and processing.

The software that might satisfy your needs is Gephi, and the display effect should be considered as fancy, and the .Net file format in Ucinet can be imported directly into Gephi. Ucinet can do the effect of these second images, the color of the line thickness, the size of the arrow, the node color, the size of the label font size, distribution mode to

adjust. Ucinet's computational analysis is very powerful, but the display of the effect of the general, Gephi in this regard to do better. The main purpose is to look good and convenient. Personal understanding of the visual effect is better than the main three elements: 1. The same method, the same setting, the increase of the amount of data will greatly increase the amount of information and display results. 2 visual information: node shape color and size of the color and size of the connecting lines and color of the arrow label font color and size. Each part of this is to be adjusted and differentiated according to the purpose of expression, and the overall effect is the result of their combination, so this part of the adjustment is time-consuming. 3 post adjustment. Layout, the general layout of the software have a set of algorithms, the choice can be. After the selection, and then to fine tune. In addition, if you have additional information to display, you can be PS, such as the top of the first picture of the small picture should be the result of post adjustment. For convenient and quick, the personal feel that more than two of the software can also be set to the basic. Of course, to do the network diagram of the small software has a lot of special foreign, the paper should also be found that the way to write small software, personal use experience is also very limited to provide a possible way. Finally, it is important to say, but it is a very important step to find out what information you want to convey.

ACKNOWLEDGMENT

XiangChao Li(Student No. 1427402043) is a undergraduate student of School of Computer Science and Technology, Soochow University. This work is supervised by Xu Lu. Corresponding author: Xun Lu, luxun@suda.edu.cn.

REFERENCES

- [1] Pagani G A, Aiello M. The Power Grid as a Complex Network: a Survey[J]. *Physica A Statistical Mechanics & Its Applications*, 2011, 392(11):2688-2700.
- [2] Yang L X, Yang X, Liu J, et al. Epidemics of computer viruses: A complex-network approach[J]. *Applied Mathematics & Computation*, 2013, 219(16):8705-8717.
- [3] Hearnshaw E J S, Wilson M M J. A complex network approach to supply chain network theory[J]. *International Journal of Operations & Production Management*, 2013, volume 33(4):442-469(28).
- [4] Radin C, Sadun L. Phase transitions in a complex network[J]. *Journal of Physics A Mathematical & Theoretical*, 2013, volume 46(30):305002-305013(12).
- [5] Jin D, Liu D, Yang B, et al. Fast Complex Network Clustering Algorithm Using Agents[C]// Dependable, Autonomic and Secure Computing, IEEE International Symposium on. IEEE, 2013:615-619.
- [6] Paparo G D, Müller M, Comellas F, et al. Quantum Google in a Complex Network.[J]. *Scientific Reports*, 2013, 3(6154):127-132.
- [7] David P, Buldú J M, Stefano B, et al. Complex network theory and the brain.[J]. *Philosophical Transactions of the Royal Society of London*, 2014, 369(1653).
- [8] Ban J L Y. Complex Network Topology of Transportation Systems[J]. *Transport Reviews*, 2013, 33(6):658-685.
- [9] Wu J, Tse C K, Lau F C M, et al. Analysis of Communication Network Performance From a Complex Network Perspective[J]. *Circuits & Systems I Regular Papers IEEE Transactions on*, 2013, 60(12):3303-3316.
- [10] Toshiyuki S. Community Extraction and Automatic layout of Pajek[J]. *Faculty Journal of Komazawa Womens University*, 2014, 21:241-253.