# Comprehension and Thinking of Complex Network Construction and Clustering Based on Neural Network Algorithm

Bing Shen<sup>1,\*</sup>

<sup>1</sup>School of Economics and Management of Beijing University of Posts and Telecommunications, Beijing 100876, China

**Abstract.** With the development of computer technology and the enhancement of people's cognition of the world, more and more scholars have been focusing on the research of complex networks. At the same time, the emerging machine learning neural network algorithm has become a powerful tool for various researchers. This paper mainly discusses the construction and clustering of complex networks based on neural network algorithm. Firstly, the development history and main application fields of neural network are introduced. Then, several common methods of complex network clustering are summarized, and then the limitations of these clustering methods are discussed. At last, it proposes to improve the construction of neural network through the concept of small world in complex network and enhance the effect of complex network clustering by the characteristics of neural network algorithm, including the accuracy, reliability, stability, speed, etc.

#### 1 Introduction

The network fills the whole nature and society, and many systems exist in the form of networks. In the same dimension, each individual and unit can be regarded as nodes in the network, and the connection between individual and unit is the connection line in the network. The system formed by nodes and connection lines is the network. Specifically, computer networks are made up of routers; the World Wide Web is made up of web pages; the power network is made up of power stations; and the social network is made up of everyone. Different networks have their own different structures and characteristics, these networks have a very high complexity, and to understand the characteristics of these networks is a complex network of disciplines. Complex networks have become one of the most important interdisciplinary research fields [1-3], which is in parallel with other basic statistical features such as small-world [1] and scale-free [2-4]. As the research of complex network clustering has important theoretical significance and application value, it not only becomes one of the most challenging fundamental research topics in computer science, but also attracts many researchers from physics, mathematics, biology, sociology and complexity science, and has set off a wave of research [5-6].

Research on neural networks is also a hot topic in academia over the years. For a long time, researchers have proposed a neural network algorithm based on brain science and biological nervous system, which is used in the field of computer and artificial intelligence. Rely on its own characteristics of neural networks, such as neurons linear independent of each other, neural network

two-way transmission, neural network algorithm has efficient computing power. The computational ability of neural network algorithm is mainly reflected in the following aspects. 1) Large-scale parallel processing and distributed information storage; 2) high adaptability and self-organization; 3) strong learning capabilities, imagination features and fault tolerance. Through the study of neural network algorithm, we can improve the computer system and the corresponding artificial intelligence technology, and contribute to the interdisciplinary field.

At the same time, the research on complex networks mainly focuses on the construction and clustering of complex networks. The construction of a complex network depends on the characteristics of each node and cable, the complexity of the problem, and the dimensions of the problem. The clustering of complex networks focuses on the study of clustering algorithms. This paper first summarizes the common clustering methods of complex networks, then proposes a new discussion based on the neural network algorithm in computer.

#### 2 Neural Networks

Neural network has been studied for a long time. In the early 1940s, McCulloch and Pitts [7] first used the method of organizing model to study the action of nerve cells from the perspective of information processing. In 1943, a binary neuron threshold model was proposed. Hebb [8], a psychologist in 1949, proposed a method called Learning Hebb, which is a method of learning by changing the binding strength between neurons. In 1962, Roseublatt [9] proposed Perceptron model. In the 70's, due to the development of integrated circuits and microelectronics

technology, neural network experienced nearly 10 years of low ebb period. After the 1980s, neural network set off a second wave of research. In 1982, American biophysicist Hopfield proposed a new neural network model with "energy" function. In 1984, he implemented the model through a concise analog electronic circuit. In the 90's, the appearance of LeNet in 1998 indicated that the neural network had formally entered the arena of computer algorithms. A large number of researchers then devoted their energies to computerized neural network algorithms, followed by many classical algorithms such as AlexNet, VGGNet, GoogLeNet, ResNet and SPPNet.

Neural networks are widely used in various fields, especially in the field of computer because of their strong computing power, stable compatibility and broad application scenarios. Common applications are, image recognition, semantic recognition, video recognition, natural language analysis.

# 3 Complex Network Clustering

The network cluster structure or network community structure is one of the most common and important topological attributes of a complex network. It has the characteristics that the same cluster nodes are connected to each other and the cluster nodes are sparsely connected with each other [10]. The method of network clustering aims at revealing the real network cluster structure in complex network.

### 3.1. Common clustering method

At present, there are many kinds of complex network clustering algorithms. According to the basic solving strategies adopted, most of them are classified into two categories: optimization based method and heuristic method. The former transforms the complex network clustering problem into an optimization problem, and calculates the cluster structure of the complex network by optimizing the predefined objective function. For example, the spectral method converts the network clustering problem into a quadratic optimization problem, the eigenvector of a particular matrix optimizes the predefined "cut" function; the latter transforms complex network clustering problems into the design of predefined heuristic rules, for instance, the heuristic rules of the Girvan-Newman algorithm [11] is that the edge betweenness of clusters should be greater than the number of edges of clusters.

Spectral methods and local search clustering methods are two main types of complex clustering methods based on optimization. The characteristic of this clustering method is that it converts the complex clustering method into similar problems and proposes an optimized solution to find the (local) optimal solution from many clustering methods.

Common heuristic complex network clustering algorithms are HITS (hyperlink induced topic search) algorithm, MFC (maximum flow community) algorithm and WH (Wu-Huberman) algorithm. A common feature of these algorithms is that heuristic algorithms are

designed based on certain intuitive assumptions, and for most networks they can quickly find the optimal solution or approximate optimal solution, yet can not theoretically strictly guarantee that they will find a satisfied solution for any input networks.

#### 3.1.1 Spectral method

Spectral methods use quadratic optimization techniques to minimize predefined "cut" functions. When a network is divided into two sub-networks, "cut" refers to the connection density between sub-networks. The partition with the smallest "cut" is considered to be the optimal network partition. The spectral method is essentially a dichotomy, in each dichotomy, the network is divided into two approximately balanced sub-networks. When the network contains multiple clusters, the spectral method recursively splits the existing sub-network until a predefined stop condition is satisfied.

#### 3.1.2 Local search clustering method

The Kernighan-Lin algorithm (KL algorithm) is a clustering algorithm proposed by Kernighan and Lin in 1970 for graph partition [12]. The optimization goal of KL algorithm is to minimize the difference between the number of inter-cluster connections and the number of intra-cluster connections. The candidate solution search strategy is to move nodes to other clusters or exchange nodes of different clusters. KL algorithm belongs to Monte Carlo algorithm. From the initial solution, it generates, evaluates and selects candidate solution in each iteration, until no better candidate solution can be found from the current solution. Throughout the search, the KL algorithm only accepts better candidate solutions and rejects all the poorer candidate solutions, so the solutions it finds tend to be local ones rather than global ones.

In 2004, Newman proposed a fast and complex network clustering algorithm based on local search, a fast Newman algorithm (as known as FN algorithm) [13]. The optimization goal is to maximize the modularity evaluation function proposed by Newman and Girvan in the same year (they called it Q function) [14]. In general, a good network cluster structure corresponds to a larger Q value. The local search strategy for candidate solutions is: Select and merge two existing network clusters, starting from the initial solution (each network cluster contains only one node), in each iteration, the FN algorithm performs a consolidation operation that maximizes the  $\Delta Q$ value until there is only one remaining network cluster in the network. Through the hierarchical clustering process from low to high, the FN algorithm outputs a dendrogram that depicts the hierarchical relationship of network clusters. Reichardt and Bornholdt proposed in 2004 network clustering algorithm based on multi-spin Potts model [15]. In this model, each network node is considered as a spin with multiple spin states, and nodes within the same cluster have the same spin state. They indicated that the optimal network cluster structure should correspond to the most stable system state, that is, the state with the lowest energy. Therefore, the problem of network

clustering is transformed into the problem of distribution of spin states to minimize the energy of the system.

#### 3.1.3 Heuristic complex network clustering method

For the connection-based WWW clustering problem, Kleinberg et al. proposed the famous HITS algorithm in 1999 [16]. The algorithm is essentially a heuristic one and is based on the basic assumption that there are two basic types of pages in the WWW: authority and hub, depending on the connection. Authoritative pages tend to be more than one central page reference, and center pages tend to reference multiple authoritative pages. Based on the interconnected relationship between authoritative and central pages, The HITS algorithm finds all the network cluster structures that are hidden in the WWW by authoritative-central pages through calculating the main eigenvectors of some special matrices ( $AA^T$  and  $A^TA$ , A indicates the adjacency matrix of Web map) corresponding to the WWW.

In 2002, Flake et al. proposed a complex network clustering algorithm MFC based on the maximum flow-minimum cut theorem of graph theory [17]. The basic assumption of this algorithm is that the maximum traffic in the network is determined by the capacity of the network "bottleneck", whereas in the network with cluster structure, the "bottleneck" of the network consists of intercluster connections. From the maximum flow - minimum cut theorem it can be known that the maximum flow in the network is equal to the minimum cut-off capacity. Therefore, the inter-cluster connection can be identified by computing the minimum cut-off. After repeated identification and deletion of inter-cluster connections, network clusters can be gradually separated.

In 2004, Wu and Huberma proposed the rapid heuristic algorithm WH [18]. The algorithm models a complex network as a circuit system. The network connection is considered as a circuit with resistance. Network nodes in different positions have different potentials. The heuristic rule of WH algorithm is that when the two nodes are selected as positive and negative poles respectively in different clusters, the resistance between the clusters is much greater than the resistance in the cluster, so the potential of the nodes in the same cluster should be approximately the same, while cluster node potential should have significant differences. The WH algorithm first calculates the potential of each node based on the Kirchhoff equation and then distinguishes different network clusters by finding the maximum potential difference.

## 3.2 Clustering limitations

First of all, the network cluster structure identified by the optimization method completely depends on the optimization goal. Therefore, the "biased" objective function can lead to "biased" solution and can not completely and accurately describe the optimal (or real) network cluster structure. Similar to the FN algorithm mentioned above, many optimization-based complex network clustering methods take the maximization

function Q as the optimization goal. However, in most cases, the Q function is biased. Therefore, for some networks, the Q value corresponding to the real network cluster structure is local maximum, not global maximum. For heuristic complex network clustering methods, the main limitation comes from the structural properties of complex networks itself, as well as the shortage of prior knowledge and a priori algorithms. For example, the WH algorithm needs to know the total number of network clusters and the number of clusters approximate size.

Secondly, another limitation of the current complex network clustering method is that it can not meet the basic requirements such as fast computing speed, high clustering accuracy and unsupervised (that is, independent of prior knowledge and parameter insensitivity). Through qualitative and quantitative analysis, comparing with the existing main methods, it can be found that the method of high clustering accuracy often has high time complexity (higher than  $O(n^2)$ ), and the fast clustering method often comes at the expense of accuracy, and requires more parameters and prior knowledge.

Thirdly, most clustering algorithms do not consider overlapping network clusters, although overlapping network clusters have broader practical implications than the average network cluster. Most of the practical problems are actually overlapping network cluster structure, and overlapping network cluster structure is more complex, there is no good method for cluster fitting. For example, in a semantic web, polysemous words are allowed to appear simultaneously in multiple network clusters that represent different semantics.

#### 4 Discussion

Based on the existing complex network problems, combined with the emerging neural network algorithm, this article will propose two points about the neural network and complex network clustering thinking.

# 4.1 Building a neural network based on a small-world model

In the same layer of neural network, each neuron is mutually linear and independent. The main function of neural networks is to express data or features by using their linear features by weighting the various elements in the network. If we build a small world model or use complex network clustering algorithm to cluster cells with the same linear features with the same weights, the computational complexity of neural networks can be greatly reduced. Specifically, taking the convolution neural network as the most common image recognition algorithm as an example, using clustering algorithm to classify each item in clustering layer immediately after the convolution layer to get cluster k, then forward the amount of computation will be reduced by k times. With the two-way propagation mechanism of neural network self-learning, the computation amount of the clustered layer after clustering will be reduced by  $k^2$  times, thereby improving the operation speed.

# 4.2 Neural network to improve the clustering calculation

On the other hand, neural networks can also provide opportunities for the development of existing clustering algorithms. Neural network provides deep learning mechanism, which can effectively increase the accuracy of clustering. Reverse propagation mechanism also ensures the reliability of clustering fitting conditions. In addition, the neural network algorithm can also increase the computing speed of clustering speed, and replace the parallel computing in the linear algorithm in the traditional clustering algorithm with the neural network algorithm.

#### 5 Conclusion

As an interdisciplinary, complex network has always been a hot topic in academic discussion, and artificial intelligence machine learning has received more and more attention with the innovation of computer technology. This paper introduces the development history of neural networks and complex networks, and summarizes the classical methods of complex network aggregation, and puts forward the limitations of the current methods. Afterwards, the improved methods of constructing and clustering complex networks are discussed from the perspective of neural network deep learning. Neural networks and complex networks have several common points, which are all operated by the properties of nodes and connections. They have similarities in structure and can be functionally complemented. Therefore, this paper proposes to construct a neural network based on a smallworld model, as well as the idea to improve clustering computing point of view with neural networks. At the same time, how to improve the accuracy, reliability, stability, speed and so on of clustering algorithm by computer calculation also become the focus of future research.

# References

- 1. Watts DJ, Strogatz SH. Collective dynamics of Small-World networks. Nature, 1998, 393(6638):440–442.
- 2. Barabási AL, Albert R. Emergence of scaling in random networks. Science, 1999, 286(5439):509–512.
- 3. Barabási AL, Albert R, Jeong H, Bianconi G. Power-Law distribution of the World Wide Web. Science, 2000, 287(5461):2115a.

- 4. Albert R, Barabási AL, Jeong H. The Internet's Achilles heel: Error and attack tolerance of complex networks. Nature, 2000, 406(2115):378–382
- 5. Strogatz S H. Exploring complex networks [J]. Nature, 2001, 410(6825):268.
- 6. Barabási A L, Author, Crandall R E. The New Science of Networks [J]. Physics Today, 2003, 6(5):243-270.
- 7. McCulloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity [J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-133.
- 8. Hebb D O. The organization of behavior: A neuropsychological approach [M]. John Wiley & Sons, 1949.
- 9. Rosenblatt F. Principles of neurodynamics [J]. 1962.
- 10. Girvan M, Newman MEJ. Community structure in social and biological networks. Proc. of the National Academy of Science, 2002, 9 (12):7821–7826.
- 11. Girvan M, Newman MEJ. Community structure in social and biological networks. Proc. of the National Academy of Science, 2002, 9 (12):7821–7826.
- 12. Newman M E J. Detecting community structure in networks [J]. European Physical Journal B, 2004, 38(2):321-330.
- 13. Newman M E J. Fast algorithm for detecting community structure in networks. [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2004, 69(6 Pt 2):066133.
- 14. Newman M E, Girvan M. Finding and evaluating community structure in networks [J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2004, 69(2):026113.
- 15. Reichardt J, Bornholdt S. Detecting fuzzy community structures in complex networks with a Potts model.[J]. Physical Review Letters, 2004, 93(21):218701.
- 16. Kleinberg J M. Authoritative Sources in a Hyperlinked Environment [J]. Journal of the Acm, 1998, 46(5):604-632.
- 17. Flake G W, Lawrence S, Giles C L, et al. Self-Organization and Identification of Web Communities [J]. Computer, 2002, 35(3):66-70.
- 18. Wu F, Huberman B A. Finding communities in linear time: a physics approach [J]. European Physical Journal B, 2003, 38(2):331-338