Journal of Theoretical and Applied Information Technology
15th December 2012. Vol. 46 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# COMPLEX ALGORITHM NETWORK AND ITS TOPLOGY ANALYSIS: A CASE STUDY

[1]MUCHOU WANG, [2]WEIFENG PAN (*corresponding author), [2]RONGYE WANG, [3]ZHUXIN HU

[1]Wenzhou University Library, Wenzhou University, Wenzhou 325035, Zhejiang, China

[2]School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, Zhejiang, China

[3]College of Teacher Education, Wenzhou University, Wenzhou 325035, Zhejiang, China

**ABSTRACT**

Complex networks are everywhere. This paper explores a new type of complex network, complex algorithm network constructed from evolutionary algorithms. The main aim of this paper is to investigate the topological properties underling the dynamics of the evolutionary algorithms. It uses complex network to describe individuals and their relationships while execution of a specific evolutionary algorithm. It introduces the main parameters in complex network theory to uncover properties of the dynamics of evolutionary algorithms. Our simulations are based on one selected evolutionary algorithm (Guo' algorithm in 10 versions) and one test functions. Data obtained through the simulations are processed graphically as well as statistically. Some novel properties of the complex algorithm network are uncovered, which has great implications for the design and improvement of existing evolutionary algorithms.

**Keywords:** *Evolutionary Algorithm, Complex Networks*

## 1. INTRODUCTION

Networks are everywhere. Over the past few years, complex networks have been intensively studied across many fields of science. Examples include social networks, bimolecular networks, software networks, the World Wide Web, etc [1]. Many real networks have been found to be neither regular graphs nor random graphs. They belong to a new type of graph, complex networks, which has completely different statistical properties than those of regular and random graphs [2]. Complex network theory can be used to study many different networks, the field of complex networks has been developing at a very fast pace and has brought together researchers from various areas such as computer science, mathematics, sociology, etc. However, there is still a type of network, algorithm network people seldom explore.

Evolutionary algorithm is a sub-discipline of computer science belonging to the bio-inspired computing area. Due to their effectiveness to cope with optimization problems, evolutionary algorithms have been widely applied in lots of fields. Evolutionary algorithms are population-based algorithms, using a population of chromosomes as candidate solutions to explore the search space. The candidate solutions are evaluated by a fitness function. Further, some reproduction operators such as crossover and mutation are used to promote the evolution of populations. The mechanism of the evolutionary algorithms is so simple, but it is very effective. However, there is little research work to explore the evolutionary process from a topological perspective underlying the parent and offspring individuals.

In this paper evolutionary algorithms are proposed to be represented as complex algorithm networks. The idea is to take advantage of the parameters in complex network theory to explore the evolution mechanism and properties hidden in the evolution of algorithms.

The rest of this paper is organized as follows. Section 2 discusses the preliminary knowledge with focus on complex algorithm network definition and an introduction of the main parameters in complex network theory. Section 3 describes a case study as an illustration to our approach. And some major structural properties have been uncovered. The paper closes with our conclusions in Section 4.

## 2. PRELIMINARY KNOWLEDGE

In this section, we define the so-called complex algorithm network, and introduce some

Journal of Theoretical and Applied Information Technology
15th December 2012. Vol. 46 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

preliminary knowledge of the complex networks [3], especially the main parameters in complex network research.

### 2.1 Complex Algorithm Network

Crossover and mutation are two main genetic operators used in a large set of evolutionary algorithms. Algorithms usually use crossover to combine two individuals to produce offspring. So obviously, there have strong relationships between the two or more parent individuals and the produced offspring. Similarly, though mutation takes place with a fairly low probability, it is applied by altering one or more gene values in an individual. So the produced new offspring also has a strong relationship to its parents. These two relationships should be captured when we are analyzing the performance of an evolutionary algorithm. In the following paragraphs, we will first give the formal definition of complex algorithm network.

**Definition 1** Complex Algorithm Network (CAN): CAN is a network (or graph) used to represent the individuals and their relationships during the population evolution process that promoted by the genetic operators such as crossover and mutation. Formally it can be represented as a graph

$$G = (V, E) \tag{1}$$

where $V$ is the set of nodes representing individuals undergoing genetic operations. And $E$ is the set of edges denoting all genetic relationships between two individuals, i.e. an edge between two nodes means there are some genes in one node inherited from another. Figure 1 gives an illustration of the details to construct a very simple CAN.

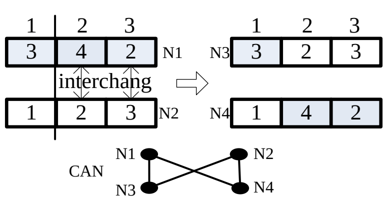



*Figure 1: Simple Illustration Of CAN*

### 2.2 Parameters of Complex Networks

In this section, we apply the complex network theory to analyze the structural quality of complex algorithm networks. We will introduce the main parameters of complex networks from the perspectives of complexity science.

#### 2.2.1 Degree and degree distribution

Degree is a basic characteristic of a specific node. It can be described as the number of links with one end on the node. Degree is usually used to quantify the importance of a node, i.e. the larger degree a node has, the more important it becomes. Additional information is provided by the degree distribution, $P(k)$, which expresses the probability of finding a node with a degree $k$ [3]. Mathematically, it can be described as $P(K = k) \sim k^{-\alpha}$ and used to check whether a network is of scale-free type or not. Networks with $P(k)$ obeying a power law tail are scale-free.

#### 2.2.2 Average path length

A shortest path between node $i$ and $j$, $d_{ij}$ , is one of the paths connecting two nodes with minimum links. Average path length, $L = <d_{ij}>$, is the average of shortest path length over all pairs of nodes in the network. It can be seen as an indirect reflection of the ability for two nodes to pass message with each other. Recently, it has been found that many real world complex networks have a small $L$ which slowly grows with the sizes of networks. For example, the $L$ of the method call network, class-class network and package-package network of Azureus is around 6, 3, and 2, respectively [1].

#### 2.2.3 Clustering coefficient

The clustering coefficient $C_i$ for node $i$ is a measure of degree to which its neighbors to be themselves neighbors in the network. Simply speaking, it is the mean probability that two nodes that are network neighbors of the same other nodes themselves will be neighbors [2]. If representing the number of links between neighbors of node $i$ as $l_i$ and the number of neighbors of node $i$ as $k_i$, $C_i$ can be calculated as $C_i = 2l_i/k_i(k_i - 1)$. Then the clustering coefficient for the network as a whole is given as the average of clustering coefficients of all the nodes, i.e. $C=<C_i>$.

#### 2.2.4 Density

The density $D$ f a network is defined as a ratio of the number of edges $E$, $nE$ to the number of possible edges, i.e.,

$$D = 2 \times nE / |V|(|V| - 1) \tag{2}$$

#### 2.2.5 Diameter

Diameter is another means of measuring network graphs. It is defined as the longest of all the calculated shortest paths in a network [3]. In other words, once the shortest path length from every node to all other nodes is calculated, the diameter is the longest of all the calculated path length. The diameter is representative of the linear size of a network.

Journal of Theoretical and Applied Information Technology
15th December 2012. Vol. 46 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

## 3. A CASE STUDY

In this section, for the illustration purpose, complex network theory will be applied to investigate the topology and its evolution of one evolutionary algorithm when represented by CANs. The parameters selected are of great importance to understand the evolution of algorithm. And the properties uncovered are useful for the design and development of new evolutionary algorithms.

### 3.1 Subject Algorithm

We defined a number of criteria to select the evolutionary algorithms used in this study, which take into consideration many aspects of the algorithms such as the availability of its source code, its effectiveness, and whether it has been used as subjects in case studies.

We manually queried the google search engine for candidate evolutionary algorithms using the mentioned selection criteria. More than 1000 algorithms returned. For limitation of time and resources, we only select Guo's algorithm as a case study to be investigated here, simply for it has been ever introduced in [4] by Liu and Zeng for the algorithm structure analysis.

Guo et al. proposed a new algorithm for solving function optimization problems with inequality constraints. And in many real applications, the performance of their algorithm is shown to be promising. Since many real-world problems can be described as a constraint optimization problem, Guo's algorithm has a wide application potential. According to [5], the Guo's algorithm can be described as Algorithm 1. It is just a simple version of their algorithm. For details, please refer to their paper.

**Algorithm 1**

**Step 1:** $t$:=1; initialize population $p(t)$; $p(t) = \{x_1(t), x_2(t), \ldots, x_N(t)\}, x_i(t) \in S$

**Step 2:** find $x_{best}$ and $x_{worst}$, such that □$\forall x \in P(t)$, better($x_{best}$, $x$) and better($x$, $x_{worst}$);

**Step 3:** while not better($x_{worst}$, $x_{best}$) do

Select $M$ individuals $x_1$, $x_2, \ldots, x_M$ from $p(t)$ randomly;

Produce a new individual $x \in V$, $V=\{x| x \in S$, $x = \sum_{i=1}^{M} a_i x_i, \sum_{i=1}^{M} a_i = 1, 0.5 \le a_i \le 1.5\}$;

If better($x$, $x_{worst}$) then $x_{worst}$:=$x$;

Find $x_{best}$ and $x_{worst}$ from $P(t)$

end do

**Step 4:** output $x_{best}$ and f($x_{best}$)

**Step 5:** end

We applied the Guo's algorithm to solve a function optimization problem with inequality constraints, which can be depicted as

**Minimize:**

$$f(x) = 5.3578547x_3^2 + 0.8356891x_1x_5 + 37.293239x_1 - 40792.141 \quad (3)$$

**Subject to:**

$$0 \le 85.334407 + 0.0056858x_2x_5 + 0.0006262x_1x_4 - 0.0022053x_3x_5 \le 92$$

$$90 \le 80.51249 + 0.0071317x_2x_5 + 0.0029955x_1x_2 + 0.0021813x_3^2 \le 110$$

$$20 \le 9.300961 + 0.0047026x_3x_5 + 0.0012547x_1x_3 + 0.0019085x_3x_4 \le 25$$

$$78 \le x_1 \le 102.33 \le x_2 \le 45.27 \le x_i \le 45, i = 3,4,5$$

Its optimum value is -3.0665.5.

The reason we chosen this minimization problem is it has been introduced in [4]. While using the Guo's algorithm, we take the same parameter setting, i.e. the population size $N$ is 50, $M$ is 8, and the maximum number of generations ($MNG$) is from 100 to 1000.

We construct the CANs according to the method shown in CAN definition. Table 1 shows the statistics of CANs constructed from Guo's algorithm when maximum number of generation is set to be 100 to 1,000. The data are in the form of number of nodes (|$V$|) and number of edges (|$E$|).

Table 1

*Table statistics of CANs*

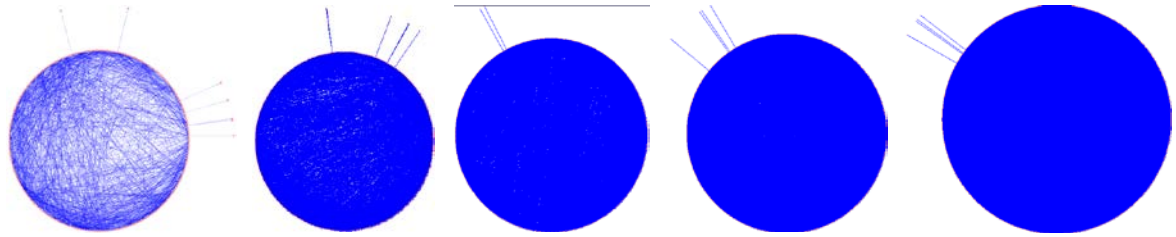
| *MNG* | \|*V*\| | \|*E*\| |
|---|---|---|
| 100 | 228 | 1468 |
| 200 | 527 | 3844 |
| 300 | 1087 | 8342 |
| 400 | 1566 | 12170 |
| 500 | 2153 | 16866 |
| 600 | 2624 | 20616 |
| 700 | 2836 | 22375 |
| 800 | 3887 | 30735 |
| 900 | 4247 | 33643 |
| 1,000 | 4333 | 34350 |

### 3.2 Subject Algorithm

Our experiments were carried out on a PC at 2.30GHz with 2 GB of RAM. We collect the running information of Guo's algorithm in all executions, and build their CANs. For illustration purpose, figure 2 shows the CANs for different setting of $MNG$ from 100 to 500. And for $MNG$ from 600 to 1,000 are omitted.

**Journal of Theoretical and Applied Information Technology**
15th December 2012. Vol. 46 No.1

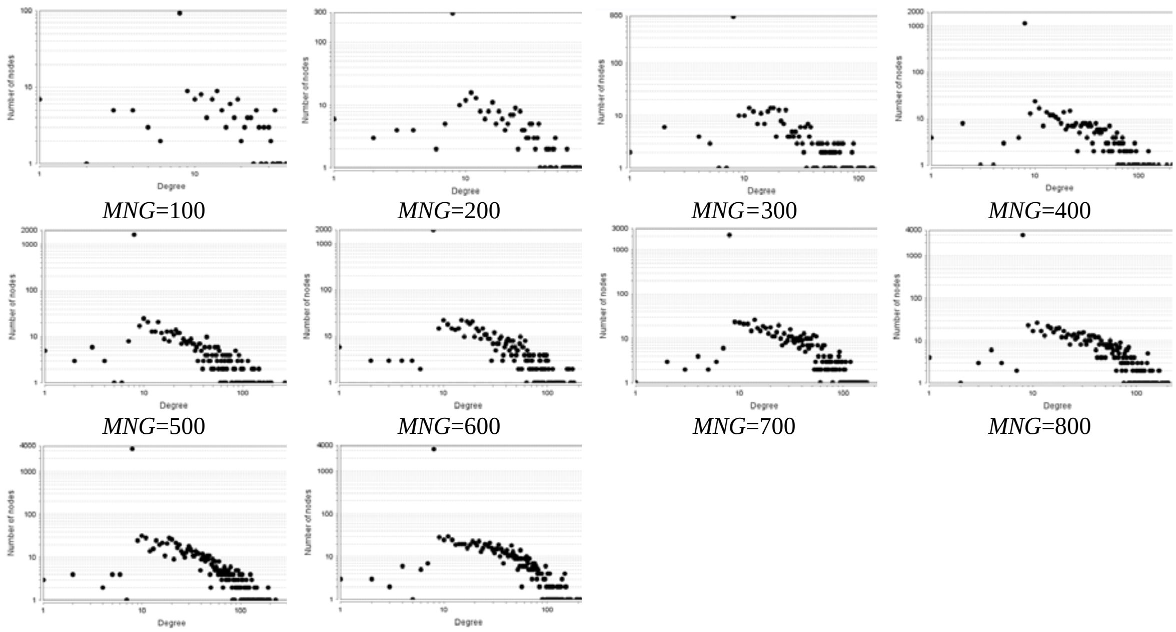
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



*MNG* = 100 *MNG* = 200 *MNG* = 300 *MNG* = 400 *MNG* = 500

*Figure 2: CANs*

The degree of a node denotes the times it has been used to generate new offspring. Figure 3 shows the degree distribution for different MNG settings. It can be easily found that, if we ignore the initial few nodes, the degree distributions roughly follow power law, especially with the increase of *MNG*. So the CANs are of nearly scare-free type networks.




*MNG*=100 *MNG*=200 *MNG*=300 *MNG*=400

*MNG*=500 *MNG*=600 *MNG*=700 *MNG*=800

*MNG*=900 *MNG*=1000

*Figure 3: Degree Distributions*

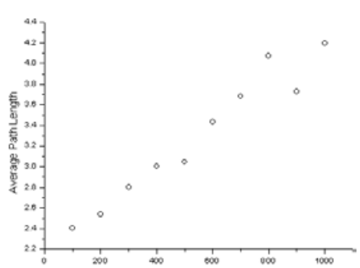



*Figure 4: Average Path Length Evolution*

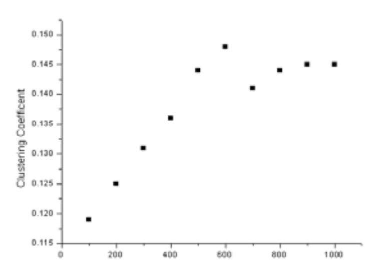



*Figure 5: Clustering Coefficient Evolution*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 4 shows the evolution of average path length over the 10 different settings of *MNG*. And Figure 5 shows the evolution of clustering coefficient. It can be found the average path length and clustering coefficient roughly grows with the increase of *MNG*. It means it becomes increasingly difficult for two nodes to pass message with each other, and there is a high probability that two nodes that are network neighbors of the same other nodes themselves will be neighbors.

$L$ and $C$ are usually used together to check a network is of small-world type or not [1]. Generally, a network can be viewed as small-world network if its $L$ is similar to the corresponding random network with the same $V$ and average degree $<k>$, and its clustering coefficient is larger. $L_{rand}$ is the average path length in the corresponding random network, which can be approximately calculated through $\ln|V|/\ln<k>$ and $C_{rand}$ through $<k>/|V|$.

*Table 2: Table statistics of CANs*

| *MNG* | *L* | *C* | $L_{rand}$ | $C_{rand}$ |
|---|---|---|---|---|
| 100 | 2.405 | 0.119 | 2.125 | 0.056 |
| 200 | 2.539 | 0.125 | 2.338 | 0.028 |
| 300 | 2.801 | 0.131 | 2.560 | 0.014 |
| 400 | 3.002 | 0.136 | 2.681 | 0.010 |
| 500 | 3.046 | 0.144 | 2.789 | 0.007 |
| 600 | 3.432 | 0.148 | 2.858 | 0.006 |
| 700 | 3.684 | 0.141 | 2.882 | 0.006 |
| 800 | 4.074 | 0.144 | 2.994 | 0.004 |
| 900 | 3.728 | 0.145 | 3.024 | 0.004 |
| 1000 | 4.195 | 0.145 | 3.030 | 0.004 |

Table II shows the $L$, $C$, $L_{rand}$ and $C_{rand}$ for each *MNG* setting. We can find that though the number of nodes and edges are very large, $L$ is very small. And its value is very similar to $L_{rand}$ of the corresponding random network, and its $C$ is much larger than that of the corresponding networks. So a conclusion can be made that CANs if of small world type.

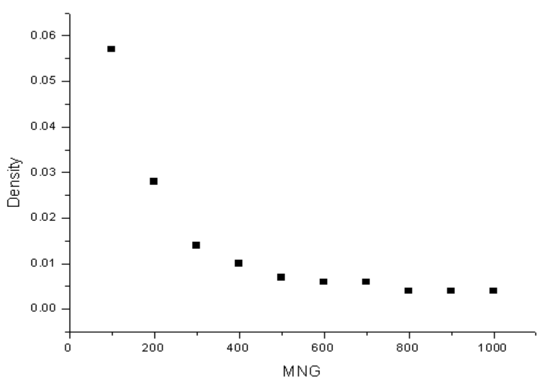



*Figure 6: Density Evolution*

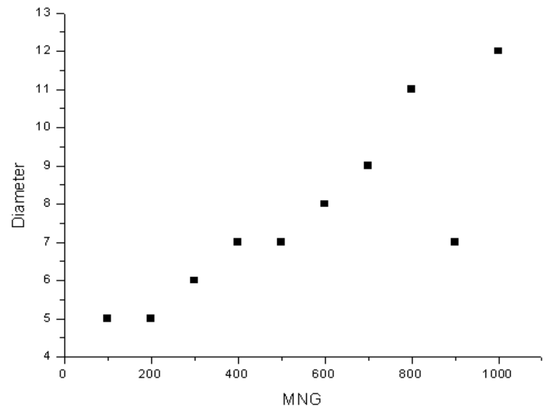



*Figure 7: Diameter Evolution*

Figure 6 and 7 show the evolution of density and diameter over the 10 different settings of *MNG*, respectively. It can be found the density declines with the increase of MSN, while the diameter grows.

## 4. A CASE STUDY

Complex algorithm network is a new type of complex network. It can also be analyzed by complex network theory. This paper represented an evolutionary algorithm, Guo's algorithm as a complex network, and used complex network theory to analyze its static property and its evolution process. And several properties have been uncovered. We found complex algorithm network are roughly a small-world type and scale-free type network. Further, average path length, clustering coefficient, and diameter grow with the increase of maximum number of generations, while diameter declines. These new results found in the current work should be taken into consideration when design or improve evolutionary algorithms. And we may propos some novel and efficient evolutionary algorithms by incorporating properties found in complex algorithm networks.

## ACKNOWLEDGEMENTS

This work was supported by Scientific Research Fund of Zhejiang Provincial Education Department (Nos. Y201224036 and Y201018266), Science Research Project of Zhejiang Gongshang University (No. 1130KU112021), and College Students Innovation Project of Zhejiang Gongshang University (No. 1130XJ1712275).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**REFERENCES:**

[1] W. F. Pan, B. Li, Y. T. Ma, J. Liu, "Multi-granularity evolution analysis of software using complex network theory", *Journal of Systems Science and Complexity*, Vol. 24, No. 6, 2011, pp. 1068-1082.

[2] M. E. J. Newman, "The structure and function of complex networks", *SIAM Review*, vol. 45, 2003, pp. 167-256.

[3] L. Costa, O. Oliveria, G. Travieso, "Analyzing and modelling real-world phenomena with complex networks: a survey of applications", 2008, *ArXiv*: 0711.3199v2.

[4] Y. Liu, Y. Zeng, "Characterizing evolutionary algorithm using complex networks theory: a case study", Proceedings of the International Conference of Internet Computing and Information Services, IEEE Conference Publishing Services, September 17-18, 2011, pp. 495-632.

[5] T. Guo, L. Kang, "A new evolutionary algorithm for function optimization", *Wuhan University Journal of Natural Science*, Vol. 4, No. 4, 1999, pp. 409-414.