Journal of Universal Computer Science, vol. 16, no. 6 (2010), 983-1003
submitted: 1/10/09, accepted: 27/2/10, appeared: 28/3/10 © J.UCS

Entropy Optimization of Social Networks Using an Evolutionary Algorithm

Maytham Safar
(Kuwait University, Kuwait
maytham.safar@ku.edu.kw)

Nosayba El-Sayed
(Kuwait University, Kuwait
nosayba.ae@ku.edu.kw)

Khaled Mahdi
(Kuwait University, Kuwait
khaled.mahdi@ku.edu.kw)

David Taniar
(Monash University, Melbourne, Australia
David.Taniar@infotech.monash.edu.au)

Abstract: Recent work on social networks has tackled the measurement and optimization of these networks' robustness and resilience to both failures and attacks. Different metrics have been used to quantitatively measure the robustness of a social network. In this work, we design and apply a Genetic Algorithm that maximizes the cyclic entropy of a social network model, hence optimizing its robustness to failures. Our social network model is a scale-free network created using Barabási and Albert's generative model, since it has been demonstrated recently that many large complex networks display a scale-free structure. We compare the cycles distribution of the optimally robust network generated by our algorithm to that belonging to a fully connected network. Moreover, we optimize the robustness of a scale-free network based on the links-degree entropy, and compare the outcomes to that which is based on cycles-entropy. We show that both cyclic and degree entropy optimization are equivalent and provide the same final optimal distribution. Hence, cyclic entropy optimization is justified in the search for the optimal network distribution.

Keywords: Social Networks, Entropy, Evolutionary Algorithm, Genetic Algorithm
Categories: L.6.0, L.6.1, L.6.2, K.4.2, J.4, G.1.6

1 Introduction

The current growth of information and communication technology has played an important role in changing the way people interact with each other [Shtykh, 08]. New forms of social interactions among people have emerged [Izumi, 09], including e-mail messages, online messages, bulletin boards and more recently, friendship networks (e.g. *Facebook, hi5, myspace*, etc). Such networks could be described as "Social Networks". Social networks consist mainly of groups of inter-connected people, where nodes represent individuals and edges represent the relationships between

them. Social Networks analysis is mainly focused on studying the patterns of communication and exchange of information between people. These patterns are believed to have great influence not only on the individuals who adopt them, but also on the societies and organizations that enclose them. The spread of diseases, news, rumors, even beliefs and convictions, has proven to become strongly dependent on the inherent structure of the social network. Moreover, the increasing number of people that adopt new technologies and interact via modern communication platforms has paved the way for Social Networks analysis. The availability of databases that store a variety of data about these networks, including the underlying topology, has facilitated the modeling and analysis of social networks to a great extent. Researchers have found great interest in utilizing this data to form a deeper understanding of how human-ties are created, and to compare them with online ties – also referred to as "tech-ties" [Green et al. 2007]. Consequently, applications on Social Networks were found in several domains including the study of epidemiology, sociology, biology, psychology and even politics.

An essential characteristic of any network is its resilience to failures or attacks, or what is known as the *robustness* of a network. The definition of a robust network is rather debatable. One interpretation of a robust network is that which assumes that social links connecting people together can experience dynamic changes, as is the case with many friendship networks such as *Facebook, Hi5*, etc. Individuals can easily delete a friend or add a new one, with and without constraints. Other networks, however, have rigid links that are not allowed to experience changes with time such in strong family network. Entropy of a network is proven to be a quantitative measure of its robustness. Therefore, the maximization of a network's entropy is equivalent to the optimization of its robustness.

Entropy is a very important characteristic that has been used to determine the degree of robustness in social networks [Mahdi et al. 2008] [Safar et al. 2008] [Sorkhoh et al. 2008] [Wang et al. 2005]. Entropy of a network is related to the probability of finding the network in a given state. For a system of moving molecules, the state is obviously the positions and the momentum of each molecule at a given instant. For a system of magnets, the state is defined through the magnets directed north or south. The entropy of a specific network shape was inverstigated before in [Simovici 2007], where the entropy of a Lattice network was studied. While this Lattice was a theoritcal one, our work studies the entropy of actual social networks with different network shapes.

In social networks, there are several choices that define the state of the network; one is the number of social links associated with a social actor, known as degree. This definition is commonly used by almost all researchers. In [Boella et al. 2009] they introduced an approach to iteratively design small social networks with methods analyzing the cooperation in the system using the relations between its nodes (based on degree distributions). However, we focused more on large networks with investigating their cycles and not degree distribution. Characterization of social network through the degree leads to different non-universal forms of distribution. For instance, random network has a Poisson distribution of the degree. Small-World network has generalized binomial distribution. Scale-Free network has a power law distribution form. There is no universality class reported.

In this work, we maximize the entropy of a scale-free network using an evolutionary algorithm, hence finding the topology of the optimal scale-free network that is robust to random failures. We calculated the network's entropy throughout our evolutionary approach as a function of cycles (cyclic-entropy). Hence, we find the optimal probability distribution function of the number of cycles that exist in the network and consequently the most probable cycle size. The originality of this work resides in the application of evolutionary algorithms to optimize networks' cyclic entropy. The scale-free network is created using Barabási and Albert's generative model.

2 Related Work

The study in [Albert et al. 2000] describe the effect of a network's heterogeneity on its degree of tolerance against either random node failures or intentional attacks. The three models of social networks are analyzed and compared: Scale-free (SF), Random Networks (RN) and Small-World (SW). Scale-free networks, which include social networks, were found to display a high degree of robustness against random failures but great vulnerability against targeted attacks. The study and analysis of resilience in complex networks was further investigated by many researchers, who mostly used percolation theory to study the resilience of different complex network topologies.

Methods based on Percolation Theory focus on analyzing the threshold value pc, which represents the number of nodes that must be removed from a network before it disconnects into smaller, separate networks. Conversely, [Wang et al. 2005] studied the robustness of scale-free networks to random failures using entropy of the degree distribution in the network, hence the level of its heterogeneity. An optimal design of a robust network was achieved through the maximization of its entropy, following a nonlinear mixed integer programming approach.

The authors in [Mahdi et al. 2008] propose a universal distribution function form based the degree of loops or cycles existing in the network instead of the degree of links in the network. The network configuration state was thus defined as the degree of cycles within the network rather than the common definition of the network state as the degree of links associated with the actors in the social network. This new distribution form was found applicable to all types of social networks (scale-free, small world, and random networks). The same definition of the system state was used in [Safar et al. 2008] on a fully connected social network for the purpose of finding the maximum entropy value, hence identifying the equilibrium state of the social network, the state of maximum entropy. In other words, finding the point where the system is most stable. The maximum value of the cyclic entropy was found to be 1.42 in the fully connected network when the number of vertices in the network was 7. The classification of social networks was achieved by calculating different social network parameters (average vertex degree, clustering coefficient and average path) for a virtual friendship network [Bhatnagar, 09] [Luo, 09] [Sorkhoh et al. 2008] [Nikulin, 08]. The results implied that the general model network considered was closest to being a small world network rather than a scale-free network, as what other studies predicted. Also, the calculation of the number of *triads* (smallest cycle size in the network) was done to both directed and undirected networks, with a varying number of nodes (1000 to 15000). The increase in the number of triads as the nodes increased

in an undirected network was found to be one magnitude of order higher than that of a directed network.

3 Cyclic Entropy of Networks

Loops were one of the major concerns in social network field. In [Scott 2000], they mentioned that the loops (cycles) can be considered as the major aspect that can separate the graph to sub-graphs or components. A cyclic component is a group of intersecting cycles. They intersect with each other by lines or points. Other researches proved that there is a strong relation between the structural balance of a social network and the loops in the network. Balancing in social network is the collection of rules [Ashrafi, 04] [Ashrafi, 07] [Taniar, 08] [Tjioe, 05] that defines the normal relations between the clients. In [Batagelj 2005] and [Wasserman and Faust 1994] they modeled some social networks as a signed directed graphs (graphs with a sign associated with each arc) and they claimed that a network is balanced if it contains only semi-cycles (a cycle that contained arcs in different directions) even number of negative arcs. Also they found that the cycle or semi-cycle is cluster able if it does not contain exactly one negative link [Kwok, 02]. Some networks must be acyclic. That is, it does not contain any cycles. Such a network is called a hierarchy. So, some researches were interested in detecting the invalid relation by counting the cycles in the graph as in [Batagelj 2005]. This process, gives each level in the hierarchy a number, a ranking process.

Based on the conclusion made by the work in [Safar et al. 2008], it is evident that degree-based classification parameters are questionable and should be used with caution. The authors in [Albert and Barabasi 2002] as well as others, tabulated social networks based on node's degrees calculation due to their simplicity, direct connection to societal behavior and ease of calculation in addition to the distinguished distribution functional form that helps recognize the type of social networks. Although the later is seen as an advantage in classification, we believe it is a disadvantage. There is a built-in commonality of all social networks that degree-based parameters and classification ignore. This property is the connectedness of the network and the return of the information to the source, as represented by a cycle.

The cycle-based approach although computationally intensive provided us with useful information about social behavior and how well the network is connected. A social network can store information in form of periodic orbits (cycles) existing in the network. A node in the network belongs to many cycles of different size; such a node is recognized as a knowledgeable node that has access to all information stored in all the intersected cycles. If we compared it with node's degree, all we can conclude is that the node has relations and not necessary access to information transferred within the links.

The authors hypothesize that understanding the cycles patterns and distribution in the network will lead to characteristic finding on the nature of social network. Social networks in most works are treated like any complex network with minimal sociological features modeled. The social actor is a simple node and the relationships are simple links connecting the nodes with specific and non-specific directions. Nothing said on the information accessible to the nodes as they are irrelevant to the network topology. If links show basic relationships only, cycles represent both the

relationships and information storage within the network. In other words, the more cycles in the network implies intrinsic property of information storage within the network. Hence, the cycle distribution will suggest the true capacity of this social network to store information. Furthermore, the calculation of statistical entropy associated with the cycle distribution becomes necessary to quantify the capacity of information storage and the maximum entropy represent the state of network useful information equilibrium [Safar et al. 2008].

Cyclic entropy was first applied in [Mahdi et al. 2009] - a study related to the dynamic change of social network during chatting [Pirrone, 08]. Interestingly, the study showed discrepancy of the values of cyclic entropy greatly related to the social behavior of the social actors when Kuwaitis and Saudis chatting rooms are compared. This conclusion is no way can be obtained looking at the nodes degrees distribution where the degree entropy of both social compared social networks is the same.

In previous work [Sorkhoh et al. 2008], extended here, we found the fully connected network will have maximum cyclic entropy when the size of the network is seven. This interesting finding, clearly suggests the economy of social network. A fully connected network of seven nodes should be sufficient to act as an information storage social network better than a network of bigger size connected as well. Such result motivated the authors to further explore and investigate cycle-based calculations.

The advantages of Cycles-based social networks versus degree-based calculation can be summarized in the following points:

- Information flow and storage is included as a major component
- Cyclic entropy distinguishes different social networks behavior
- Universal mathematical functional form of distribution

The biggest disadvantage of the cycle approach is cycles' computations, however, the existence of a universal distribution representing social networks may be used to facilitate courting the cycles in the network.

From statistical mechanics, the entropy can be calculated from a given probability distribution *P(k)* of the system in state *k*:

$$S = -\sum_{k} P(k) \ln P(k) \tag{1}$$

where *P(k)* is the probability of finding a cycle of length *k* in the network. Based on the definition of entropy in Eq 1, we need to generate an initial distribution to start the optimization and let the evolutionary algorithm search for the optimal distribution that maximizes the total entropy of the network.

4 Cyclic Entropy Optimization

Optimization algorithms are algorithms designed to find an optimal value of a variable *x* for a given function *f(x)*, such that *f(x)* is either maximized or minimized, under a possible set of constraints. They are applied in solving optimization problems,

which are mainly defined as problems of finding an optimal (best) solution through searching within the set of all feasible solutions – also known as the *solution space*. Optimization problems are categorized as NP-hard (nondeterministic polynomial-time hard), where an exact solution for the problem is difficult to find. Alternatively, heuristics are used to find a *good* solution that is reasonably close to the optimal solution of the problem, within a reasonable amount of time.

Genetic Algorithms (GAs), one type of Optimization Algorithms, are adaptive heuristic search algorithms, which are based on mimicking the powerful process of natural evolution. GAs are used in solving optimization problems to produce an optimal or near-optimal solution. The basic techniques of the GAs are designed such that the solution search process is directed into the region of better performance. This is accomplished by imitating the "survival of the fittest" principle, during the selection stage of the GA. Several generic selection algorithms exist, such as tournament selection, fitness proportionate selection and stochastic universal sampling.

In this work, we use a different, evolutionary approach to optimize the entropy of the cyclic distribution in social networks: Genetic Algorithms (GA). We follow a comprehensive set of guidelines regarding the usage and design of a genetic algorithm for various problems, in [Baeck et al. 2000]. Aspects such as fitness evaluation, constraint handling issues [Pardede, 05], population sizing and structuring, mutation parameters and parameter control are explained within the scope of conducted research in evolutionary algorithms. Moreover, the authors in [Baeck et al. 2000] discuss the efficient implementation of evolutionary algorithms for different sets of problems, using different parameters. [Tan et al. 2005] presents another comprehensive treatment on designing multi-objective evolutionary algorithms and their applications in various areas, such as control systems and evolutionary scheduling.

In order to apply genetic algorithms effectively on scale-free networks, the structure of this type of networks should be examined first. Scale-free networks are networks that follow a power law degree distribution. In other words, the number of links "k" originating from a given node in a scale-free network exhibits a power law distribution: $p(k) \sim k^{\gamma}$, such that $2<\gamma<3$. The power law distribution of scale-free networks affects its topology, where the nodes with highest degrees (hubs) are followed by the smaller ones, and so on. Due to the fact that the majority of nodes are small ones, scale-free networks are robust to random failures because the probability of a hub failure is very small. Examples of networks that are categorized as scale-free networks are the World Wide Web (where the web pages are considered nodes and the links connecting them are the hyper-links), protein interaction networks, the network of Hollywood actors and various Social Networks. Scale-free networks can be generated using Barabási and Albert's generative model [Albert and Barabasi 2000]. We begin with an initial network that consists of m_0 nodes, where $m_0 \geq 2$. Next, new nodes are progressively added to the existing network using preferential attachment mechanism. The probability p_i of linking a new node to an existing node i is proportional to the number of links k already attached to this (existing) node, as follows:

$$p_i = \frac{k_i}{\sum_i k_i} \tag{2}$$

where k_i is the degree of node i.

The optimization starts with an initial distribution derived from Barabási and Albert's scale-free generative model that randomly produces the initial population of scale-free social networks, which is given as an input to our genetic algorithm. Each chromosome in the list of chromosomes formulating the population represents the *directed* graph that corresponds to a generated scale-free network. The nodes in the graph represent individuals and the links represent relationships among them. For instance, (A → B) implies that A is a friend of B. The objective function of our genetic algorithm is to maximize the cyclic entropy of the scale-free network, where entropy is defined using eq. 1. The objective function is expressed as [Tan, 06], [Tan, 07]:

$$\max\left(-\sum_k P(k)\ln P(k)\right) \tag{3}$$

Accordingly, the fitness function that quantifies the value of the each chromosome in our solution space is simply equal to the value of the cyclic entropy of its corresponding graph. Our objective function is constrained with the restriction that all graphs in the population remain connected during the optimization loop. That is, to ensure that the different mutation operators we implemented would never cause the graph selected for mutation to disintegrate into disconnected sub-graphs throughout all the GA generations. This constraint is important due to the nature of the mutation operators we implemented in the genetic algorithm: 1) Adding an edge between two randomly chosen vertices in the graph, and 2) Removing a random edge that connects two vertices in the graph. These mutation operations were chosen such that they ensure the evolvement of the cyclic structure of each graph, thus evolving the chromosome holding it to reproduce better generations. Naturally, adding and removing edges randomly in the graph will modify the number and lengths of cycles existing in it, therefore altering its cyclic entropy value.

The mutation rate parameter in our genetic algorithm is a dynamic rate that varies according to a deterministic time-dependent function $p_m(t)$ suggested by Back and Schlutz [Thomas and Martin 1996]:

$$p_m(t) = (2 + \frac{L-2}{T}t)^{-1} \qquad 0 \le t \le T \tag{4}$$

where t is the generation number, T is the maximum number of generations and L is the length of the chromosome (in our case, the size of the graph). This rule changes the probability of mutation as the number of generations increase, such that $p_m(0) = 0.5$ and $p_m(T) = 1/L$. Previous studies [Elben et al. 1999] [Hinterding et al. 1997]

[Thomas and Martin 1996] proved that using a dynamic deterministic mutation rate instead of a fixed mutation rate improves the evolution process, and avoids over-mutating mature chromosomes as it reduces the mutation rate with the increase of time. The pseudo code of our cyclic-entropy optimization algorithm for the design of a robust social network is shown in fig. 1.

```
1  Initialize a population of 30 scale-free network graphs:
      1.1    m0 = 3 , m = 3 ;     //parameters of Barabasi and
         Albert's algorithm
      1.2    n = 30 ;               //size of population
      1.3    Population_List = create_empty_list() ;
      1.4    for( 1 → n )
           1.4.1 new chromosome = Barabasi and Albert's
                 generative model(m0 , m, n);
           1.4.2 Population_List.add( chromosome);

2  Evaluate_generation( Population_List , n):
      2.1   for( i = 1 → n )
      2.2   Population_List( i ).Calculate_Cyclic_Entropy();

3  For ( i = 1 → number_of_generations )
      3.1    Sort( Population_List)
      3.2    int k = Random( n - 1 );        //the index of
         the chromosome to be mutated
      3.3    mutation_operation = Random( add , remove );
         //choose mutation operation randomly
      3.4    if( mutation_operation = add )
           3.4.1 Population_List(k).add_random_edge();
      3.5    Else
           3.5.1 Population_List(k).remove_random_edge();
      3.6    Evaluate current generation:
           3.6.1 for( i  = 1 → n )
           3.6.2 Population_List( i
                 ).Calculate_Cyclic_Entropy();
4  End_for_loop  //end of Genetic Algorithm's Optimization
   Loop
```

Figure 1: Genetic Algorithm's Optimization Loop Pseudocode

To implement the constraint of ensuring graph connectivity in the genetic algorithm, we utilized the *Floyd Warshall* algorithm [Weisstein], which follows a dynamic programming approach to calculate the shortest paths between all vertices of

a graph. The pseudo code of Floyd Warshall algorithm, which assumes an input graph of N vertices is shown in figure 2 below.

```
for i = 1 to N
    for j = 1 to N
       if there is an edge from i to j
       dist[0][i][j] = the length of the edge from i to j
        else
         dist[0][i][j] = INFINITY

 for k = 1 to N
  for i = 1 to N
    for j = 1 to N
       dist[k][i][j] = min(dist[k-1][i][j], dist[k-1][i][k] +
                           dist[k-1][k][j])
```

Figure 2: Floyd Warshall Algorithm's Pseudocode

One of the applications of Floyd Warshall algorithm in graph analysis is finding the *transitive closure* of a directed graph: a data structure that enables determining in O(1) time possible reachability questions, such as: Can we reach node "b" from "a" in one or more hops? The data structure is stored as a matrix where each entry, matrix[i][j], answers whether node *i* can reach node *j* in one or more hops. Using this approach, we can check by examining the elements of the constructed matrix if the graph has any disconnected parts; i.e., if there are at least two nodes in the graph that are not able to reach each other. If this case occurs, then we must undo the mutation operation and randomly apply another operation, and so on.

5 Experimental Design

We have implemented the algorithm using JAVA as a programming language. In our experiment, we ran the genetic algorithm with an initial population of 30 chromosomes, where each chromosome represents a scale-free network graph of size 20 nodes. Each graph was constructed using Barabási and Albert's scale-free generative model. In order to evaluate the chromosomes using entropy as a function of cycles, we utilize the following the algorithm stated in [Mahdi et al. 2008] to compute the number and lengths of cycles in each graph, hence its cyclic entropy according to equation (1). The "cycles counting" algorithm is based on the iterative loop counting algorithm (ILCA) from MathWorks®. We modified this algorithm to meet our network criteria. ILCA is developed to find all the loops in any connected undirected graph by starting from any vertex or a vertex that has the most links and then search all the paths from it. This approach will not work with directed graph because a vertex that can reach all the edges must be found first. Also the way it distinguishes the cycles from reach other is not efficient in a directed graph. That's

because in undirected graph, the cycle A→B→C is the same as A→C→B. We solved those two problems by making the vertex that we start extracting the network from as the start point and changing the algorithm that it uses for cycle distinguishing.

We ran the genetic algorithm on a machine of 2GB RAM and 2.4 GHz processor, for 200 generations. In each generation *t* a number of chromosomes equal to ($p_m(t)$* *population size)* is selected for mutation, where $p_m(t)$ is the mutation rate according to the time-varying dynamic formula stated in equation (4). For each selected chromosome, the mutation operation to be applied was chosen randomly between 1) adding an edge that connects two random vertices in the graph, or 2) removing an existing edge that connects two vertices together in the graph. Afterwards, the mutated graphs' new entropy values are calculated as their cyclic structures have been altered, reproducing the new generation of scale-free networks, and so on.

In the second experiment we conducted, the same initial population which consisted of 30 scale-free networks was used as an input to the Genetic Algorithm; only this time we used Entropy of the links degree distribution to define the fitness function, instead of the cyclic-entropy. Therefore, each chromosome was evaluated according to the Entropy of the link degree distribution of its corresponding graph. We ran the genetic algorithm with the same parameters used in the first experiment, in terms of the time-varying mutation rate, number of generations, input population and mutation operations. Table 1 shows the parameters setting used within both experiments.

Parameter	Value
Initial GA population	30
Chromosome (graph) size	20
m_0 (BA model)	3
m (BA model)	3

Table 1: Parameters settings for the entropy optimization experiments

6 Experimental Results

6.1 Cyclic Entropy Optimization Results

Figure 3 shows the best entropy reached in each generation of the genetic algorithm, for the first experiment. The optimum value reached was **2.519** at generation 121, where it remained constant until generation 160, followed by a relatively large drop in its value. Figure 4 keeps track of the overall best entropy reached in all the generations, which is of great significance to us since we are interested in observing the behavior of the corresponding network; the overall best chromosome represents the optimal robust scale-free network that has the best entropy within all possible solutions in the solution space.

For this optimal solution, we have plotted its initial cycle length distribution before the evolution process and the final cycle length distribution after running the genetic algorithm, as shown in figures 5 and 6, respectively. The cyclic entropy that corresponds to its initial status was equal to **2.187**, where the optimal entropy corresponding to its evolved final status is equal to **2.519**. Figure 7 compares both plots together in order to observe the regions where the major changes took part, i.e. the different cycle length regions that the evolution process altered the most, in order for this solution to evolve into the optimum solution.

Figure 8 clarifies, numerically, those regions where changes of different cycle length occurrences take place the most. It is obvious that cycles of lengths (7 to 9), 14 and 15 experienced more alteration than the rest. Cycles of lengths (2 to 5), 11, 12, (17 to 19) experienced less alteration. Figure 9 demonstrates the distribution of cycle lengths of our optimal solution throughout the different generations in the genetic algorithm. Starting from the first generation, we can observe how the cyclic structure of our solution evolved through the generations, eventually reaching its optimal design, at generation 158. It is noticeable that most of the evolvement to the cyclic structure occurs in regions (8 to 10) and (14 to 15). On the other hand, regions (2 to 6) and (17 to 19) experienced almost negligible changes. The number of mutation operations this chromosome was subject to is **20** operations: **9** random edge removals and **11** additions of random edges. The total time taken by this experiment to execute is 204640.562 seconds = 56.844 hours.

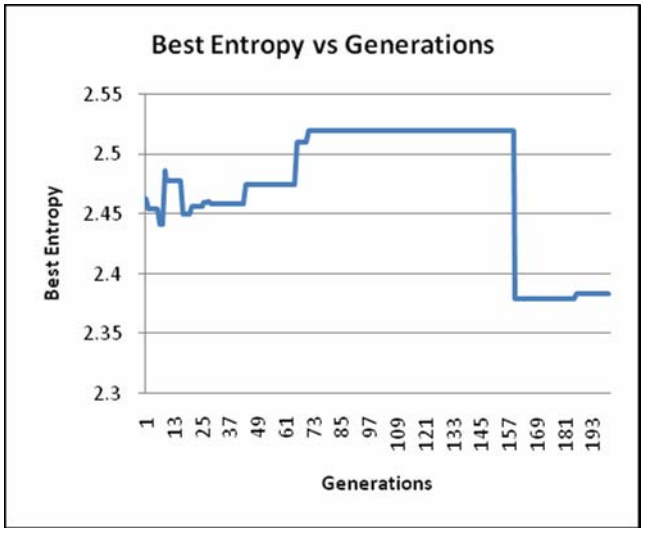

Figure 3: Best Entropy vs. Generations

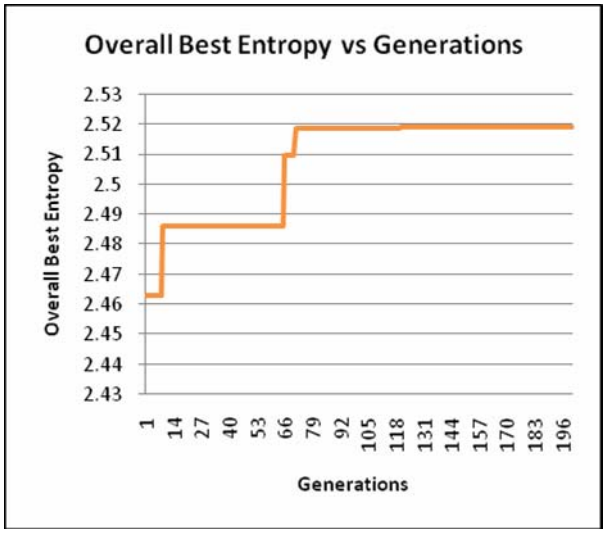

Figure 4: Overall Best Entropy vs. Generations

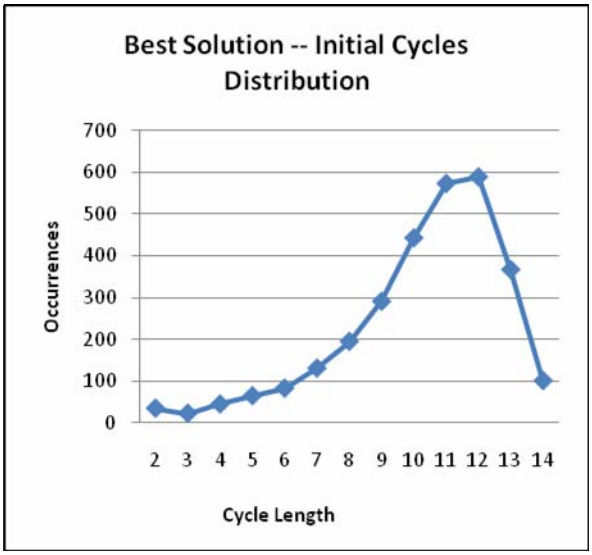

Figure 5: The initial cycles distribution of the optimal solution

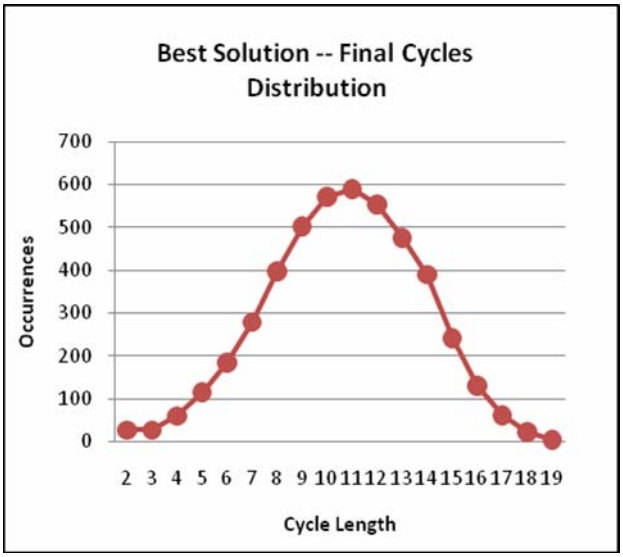

Figure 6: The final cycles distribution of the optimal solution

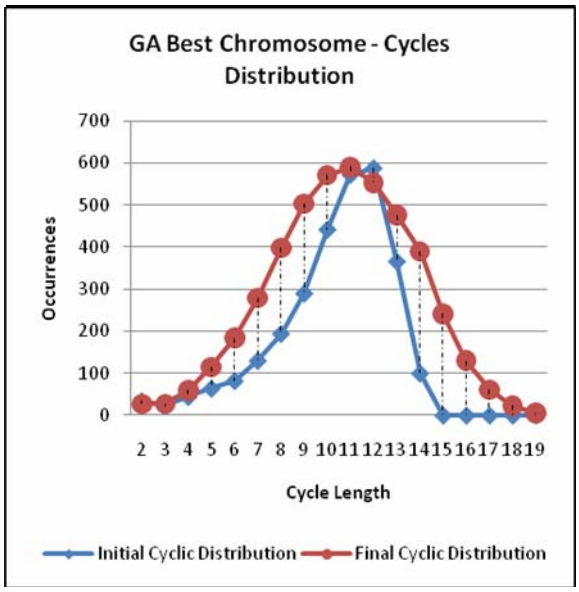

Figure 7: Initial VS Final cycles distribution of the optimal solution

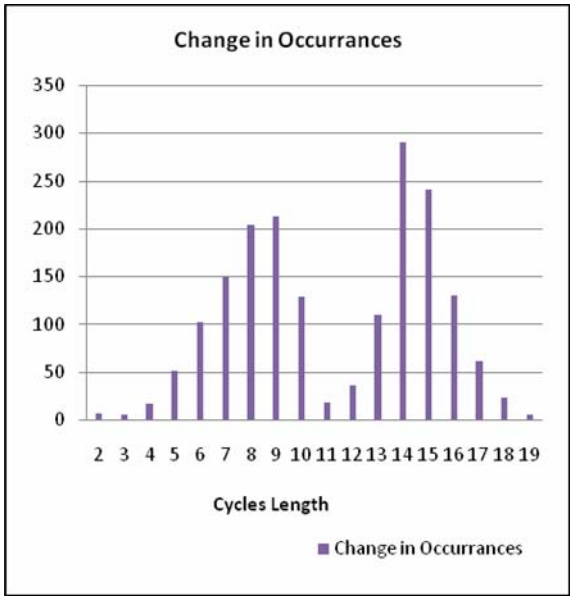

Figure 8: Absolute change in occurrences of different cycle lengths

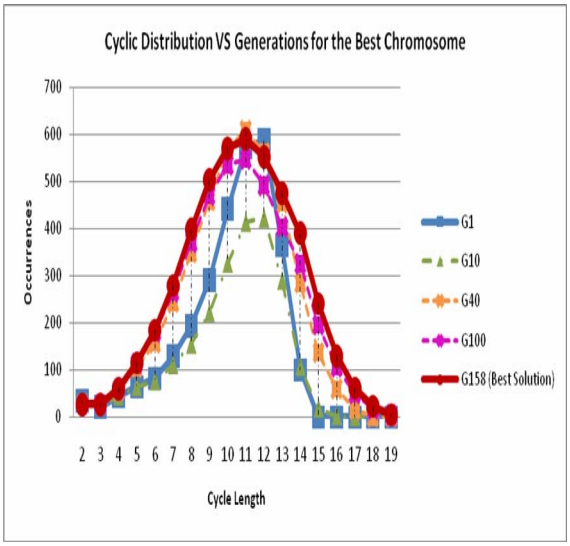

Figure 9: Variation of the cycle length distribution over the genetic algorithm generations

6.2 Degree Entropy Optimization Results

In the second experiment which uses the Entropy of the links degree distribution to define the fitness function, the optimal degree entropy value reached was **2.207** at generation **102**. Figure 10 shows the best degree entropy reached in each generation of the genetic algorithm. Figure 11 shows the overall best entropy within all generations, which reached its highest (optimal) value at generation 102.

The initial degree distribution of the optimal solution is shown in figure 12. It is obvious that the distribution follows a power-law degree distribution, which is a property of scale-free networks. The degree entropy corresponding to this distribution was found to be **1.69574**. After running the genetic algorithm, the optimal degree entropy of this chromosome after its evolution is equal to **2.2071079**. Figure 13 shows the degree distribution that corresponds to this optimal entropy. Figure 14 compares both initial and final degree distributions of the best solution. The number of mutation operations this chromosome was subject to is **27** operations: **11** random edge removals and **16** additions of random edges. The total elapsed time taken to execute this experiment is equal to 99.437 seconds only.

For the optimal solution resulting from the second experiment -which has the maximum degree entropy-, we have applied the cycles counting algorithm to find the distribution of cycles within it, as shown in figure 15. The most probable cycle size in this solution is a cycle of **10** nodes, and the corresponding cyclic entropy was found equal to **2.18976**. On the other hand, the degree distribution was calculated for the optimal cyclic-entropy solution that resulted from the *first* experiment. The distribution, shown in figure 16, indicates a scale-free power law distribution, with a degree-entropy value of **1.85136**.

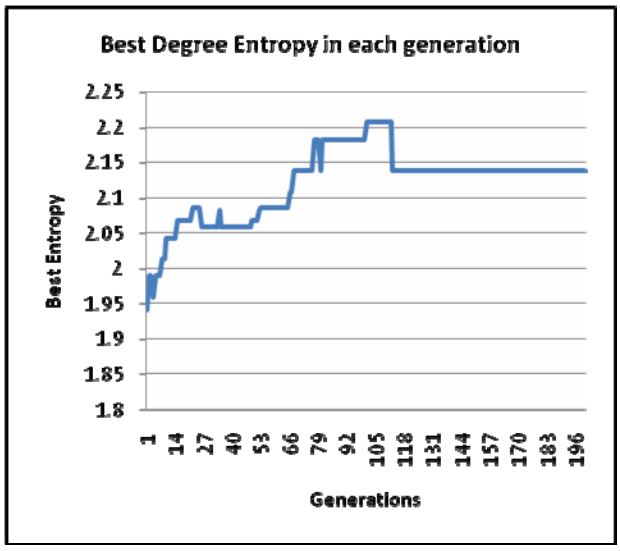

Figure 10: Best Degree Entropy vs. Generations

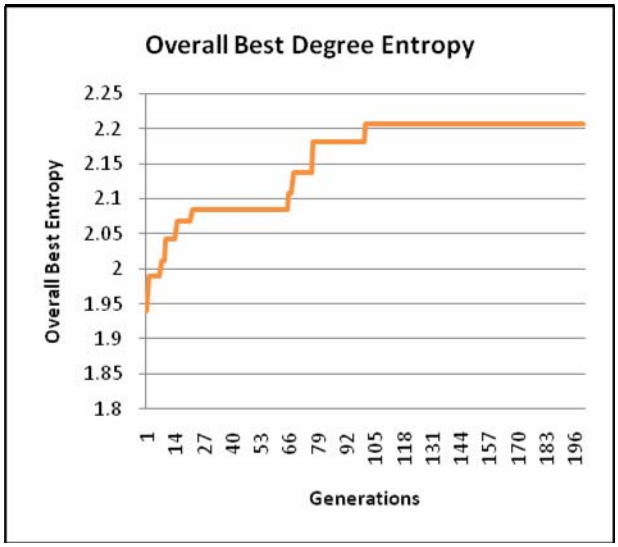

Figure 11: Overall Best Degree Entropy vs. Generations

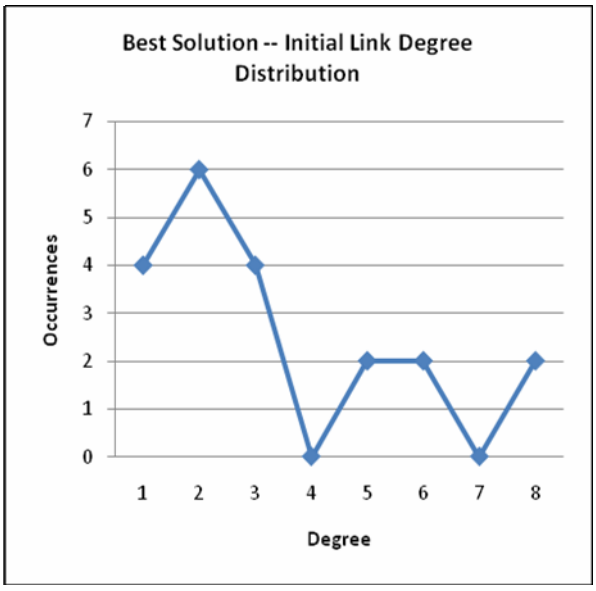

Figure 12: Initial degree distribution of the optimal solution

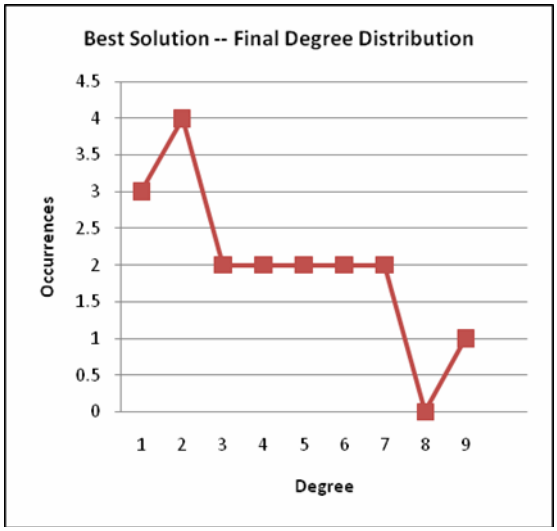

Figure 13: Final degree distribution of the optimal solution

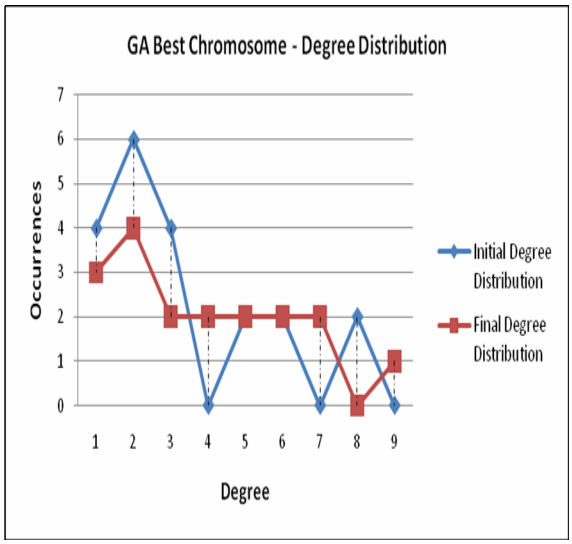

Figure 14: Initial vs. Final degrees distribution of the optimal solution

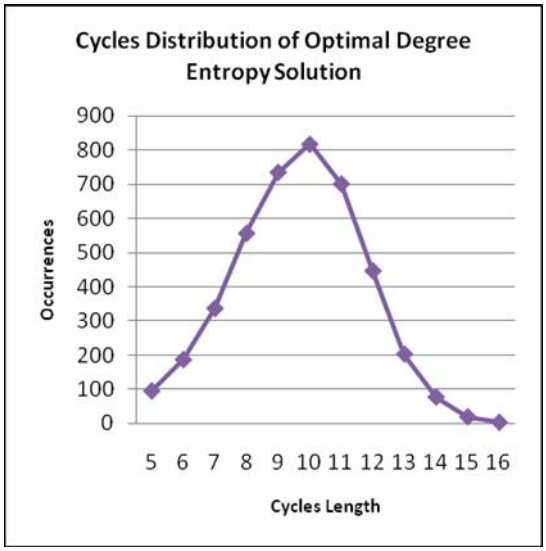

Figure 15: Cycles distribution of the optimal degree entropy solution

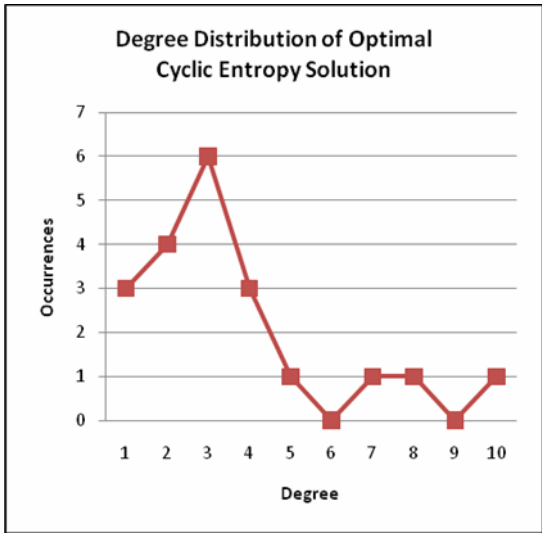

Figure 16: The degree distribution of the optimal cyclic entropy solution.

7 Conclusions and Future Work

The experiment we conducted in this work demonstrates the use of the genetic algorithm in the design of robust, scale-free social networks. The optimization routine finds the maximum entropy equals **2.519** corresponding to an optimal network distribution found when the initial distribution is subject to 9 random edge removals and **11** additions of random edges regardless of the initial distribution. In addition, the most probable cycle size in the optimal solution is a cycle of **11** nodes. In a maximum fully connected network the optimal cycles distribution has a its most probable cycle size at 7 with a maximum entropy equals 1.42 much lower than the scale free network studied in this work where a fully connected network of 11 nodes is less than the maximum [Mahdi et al. 2008]. According to the definition of entropy, we conclude that a scale-free network of 11 nodes has more robustness and will be more resilient to failure.

In the search for a design of most robust network, we intend to extend the work to examine other types of social network model such as small world. We propose the use of cycle distribution instead of degree distribution for many reasons. Degree distribution is one dimensional hence it suggests little information on the nature of the network. As for cycles distribution, it is a two dimensional problem that provides more elaborate information about the network. In previous work, the authors showed that cycle distribution provide solid and unique evidence of the type of actual social network through the analysis of its cyclic entropy. In addition, cycles distribution is found to have one universal mathematical representation where degree distribution is mathematically specific depending on the type of social network. Such uniformity of cycles distribution allows better characterization of social networks, the cycle distribution is given by the probability function, $p(l_i)=a\ exp(-(l_i-b)^2/c^2)$ where l_i is the cycle length size, and a, b and c are positive real numbers that have unique values for each network. With the result of this work, we proceed in performing optimization of the universal cycle distribution function in order to accurately find the most robust social network and the most resilient to failure.

References

[Albert, 00] Albert, R., Jeong, H., Barabasi, A.L.: Error and attack tolerance of complex networks, Nature, vol. 406, pp. 378-382, 2000.

[Albert, 02] Albert, R., Barabási, A.L.: Statistical mechanics of complex networks, Reviews of Modern Physics, vol. 74, January 2002.

[Ashrafi, 04] Ashrafi, M.Z., Taniar, D., Smith, K.A.: ODAM: An Optimized Distributed Association Rule Mining Algorithm. IEEE Distributed Systems Online 5(3), 2004

[Ashrafi, 07] Ashrafi, M.Z., Taniar, D., Smith, K.A.: Redundant association rules reduction techniques, International Journal of Business Intelligence and Data Mining 2(1): 29-63, 2007

[Baeck, 00] Baeck, T., Fogel, D.B., Michalewicz, Z.: Evolutionary Computation 2: Advanced algorithms and operations, vol. 2. Bristol: IoP Publishing, 2000.

[Batagelj, 05] Batagelj, V.: Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Science, England: Cambridge University Press, 2005.

[Bhatnagar, 09] Bhatnagar, V., Kaur, S., Mignet, L.: A Parameterized Framework for Clustering Streams, International Journal of Data Warehousing and Mining 5(1): 36-56 (2009)

[Boella, 09] Boella, G., Torre, L., Villata, S.: Analyzing Cooperation in Iterative Social Network Design, Journal of Universal Computer Science 15(13): 2676-2700, 2009.

[Elben, 99] Elben, A.E., Hinterding, R., Michalewicz, Z.: Parameter control in evolutionary algorithms, IEEE Transactions on Evolutionary Computation 3: 124-141, July 1999.

[Green, 07] Green, D.T., Martin, N.L., Pearson, M.: Tech ties: making online connections in social networks, International Journal of Web Based Communities, vol. 3, pp. 460-467, 2007.

[Hinterding, 97] Hinterding, R., Michalewicz, Z., Eiben, A.E.: Adaptation in evolutionary computation: a survey, IEEE International Conference on Evolutionary Computation Indianapolis, IN, USA, pp. 65-69, 1997.

[Izumi, 09] Izumi, S., Yamanaka, K., Tokairin, Y., Takahashi, H., T., Shiratori, N.: Ubiquitous supervisory system based on social contexts using ontology, Mobile Information Systems 5(2): 141-163 (2009)

[Kwok, 02] Kwok, T., Smith, K.A., Lozano, S., Taniar, D.: Parallel Fuzzy c-Means Clustering for Large Data Sets. In Proceedings of the 8th International Euro-Par Conference, Lecture Notes in Computer Science LNCS 2400, Springer, pp: 365-374, 2002

[Luo, 09] Luo, J., Ni, X.: A clustering analysis and agent-based trust model in a grid environment supporting virtual organisations, International Journal of Web and Grid Services 5(1): 3-16 (2009)

[Mahdi, 08] Mahdi, K., Safar, M., Sorkhoh, I.: Entropy of Robust Social Networks, Proceedings of the International Association for Development of the Information Society (IADIS) International e-Society Conference, 2008.

[Mahdi, 09] Mahdi, K., Safar, M., Farahat, H.: Analysis of Temporal Evolution of Social Networks, Journal of Mobile Multimedia (JMM), vol. 5, pp. 333-350, 2009.

[Nikulin, 08] Nikulin, V.: Classification of Imbalanced Data with Random sets and Mean-Variance Filtering, International Journal of Data Warehousing and Mining 4(2): 63-78 (2008)

[Pardede, 05] Pardede, E., Rahayu, W., Taniar, D.: Preserving Conceptual Constraints During XML Updates, International Journal of Web Information Systems 1(2): 65-82, 2005

[Pirrone, 08] Pirrone, R., Russo, G., Cannella, V., Peri, D.: GAIML: A new language for verbal and graphical interaction in chatbots, Mobile Information Systems 4(3): 195-209 (2008)

[Safar, 08] Safar, M., Mahdi, M., Sorkhoh, I.: Maximum Entropy of Fully Connected Social Network, Proceedings of the International Association for Development of the Information Society (IADIS) International Conference on Web Based Communities, 2008.

[Scott, 00] Scott, J.: Social Network Analysis: a Handbook, Second ed. U.S.A: Sage Publications Ltd, 2000.

[Shtykh, 08] Shtykh, R.Y., Jin, Q.: Harnessing user contributions and dynamic profiling to better satisfy individual information search needs, International Journal of Web and Grid Services 4(1): 63-79 (2008)

[Simovici, 07] Simovici, D.: Metric-Entropy Pairs on Lattices, Journal of Universal Computer Science 13(11): 1767-1778, 2007.

[Sorkhoh, 08] Sorkhoh, I., Safar, M., Mahdi, K.: Classification of Social Networks, Proceedings of the International Association for Development of the Information Society (IADIS) WWW/Internet Conference 2008.

[Tan, 05] Tan, K.C., Khor, E.F., Lee, T.H.: Multi-objective Evolutionary Algorithms and Applications: Advanced information and knowledge processing. London: Springer Verlag, 2005.

[Tan, 06] Tan, L., Taniar, D., Smith, K.A.: Maximum-entropy estimated distribution model for classification problems, International Journal of Hybrid Intelligent Systems 3(1): 1-10, 2006

[Tan, 07] Tan, L., Taniar, D.: Adaptive estimated maximum-entropy distribution model. Information Sciences 177(15): 3110-3128, 2007

[Taniar, 08] Taniar, D., Rahayu, J.W., Lee, V.C.S., Daly, O.: Exception rules in association rule mining. Applied Mathematics and Computation 205(2): 735-750, 2008

[Thomas, 96] Thomas, B., Martin, S., Intelligent Mutation Rate Control in Canonical Genetic Algorithms, Proceedings of the 9th International Symposium on Foundations of Intelligent Systems: Springer-Verlag, 1996.

[Tjioe, 05] Tjioe, H.C., Taniar, D.: Mining Association Rules in Data Warehouses, International Journal of Data Warehousing and Mining 1(3): 28-62, 2005

[Wang, 05] Wang, B., Tang, H., Guo, C., Xiu, Z.: Entropy optimization of scale-free networks' robustness to random failures, Physics A, vol. 363, pp. 591-596, 2005.

[Wasserman, 94] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. England: Cambridge University Press, 1994.

[Weisstein] Weisstein, E.: Floyd-Warshall Algorithm, from MathWorld – A Wolfram Web Resource, http://mathworld.wolfram.com/Floyd-WarshallAlgorithm.html.