2015 17th UKSIM-AMSS International Conference on Modelling and Simulation

The Effect of State Space Clustering on the Performance of Simulated Annealing and Its Topology-Aware Variant

Said Kerrache, Hafida Benhidour
College of Computer and Information Sciences
King Saud University
Riyadh, KSA
skerrache@ksu.edu.sa, hbenhidour@ksu.edu.sa

***Abstract*—Simulated annealing is one of the most widely used algorithms for global optimization. Due to its success, several variants of classical simulated annealing have been proposed. These variants may use more sophisticated neighborhood selection strategies or may employ different acceptance probabilities. Topology-aware simulated annealing is one such variant that takes into consideration the branching factor of states when performing uphill moves. The experimental evaluation done on topology-aware simulated annealing suggests the potential effect of clustering on performance. In this paper, we experimentally investigate the effect of the state space clustering on the performance of classical simulated annealing and its topology-aware variant. This is achieved through the use of networks with different degrees of clustering as search spaces. These networks are generated using the hidden metric model, a recently proposed complex network model. The results show that the effects are indeed nontrivial, and that there exist certain clustering levels that cause an improvement in the performance. These results are more pronounced in spaces with multiple deep local minima, where the performance falls off if the clustering of the network is made smaller or larger than a certain optimal value.**

Keywords-Simulated annealing; topology-aware simulated annealing; scale-free networks; hidden metric model; clustering

I. Introduction

Simulated annealing is a well known stochastic optimization algorithm [1], [2], [3] that is inspired by the dynamics of large systems of particles in a thermodynamic system [4]. Simulated annealing has been successfully used for solving global optimization problems in a wide range of disciplines in science, engineering and operation research [3], [5], [6], [7]. This is due to its good performance, low computational cost and memory requirements, simple implementation as well as its solid theoretical foundation. Simulated annealing uses a stochastic decision rule that allows it to escape local minima and hence improve the quality of the solution. Transition towards nodes with higher objective value (also called energy) are made according to an acceptance probability p that depends on the difference ΔE in the objective value between the current state and the neighbor as well as a parameter T called temperature:

$$p = \exp\left(\frac{\Delta E}{T}\right). \quad (1)$$

This form ensures that transitions towards higher energy states only take place when the difference in energy is relatively small and the temperature is high enough.

Several variations of the acceptance probability have been proposed in the literature in attempt to improve the performance or reduce the computational cost. Some of these variations are presented in Section II. In [8], we proposed an acceptance probability that takes into consideration the degree of the next state, which is a topological feature of the state space. Indeed, the state space can be considered as a network and can be analyzed using the recent techniques introduced in the field of network science [9]. The latter is an interdisciplinary research field concerned with the analysis of large scale real life networks such as the Internet, social networks and biological networks. Its main goal is analyzing and modeling networks as well understanding the effect of topology on the different processes taking place on the network. The experimental analysis that we conducted in [8] showed that the topology of the network affects the performance even though the former is not correlated with the objective function. In particular, we observed that a scale-free topology gives better results than a random topology. This suggests that clustering may be the key feature causing the improvement. In this paper, we experimentally investigate the effect of clustering on the performance of simulated annealing and its topology-aware variant.

The rest of this paper is organized as follows. In Section II, we give an overview of the related work on simulated annealing and its topology-aware variant. Section III contains a review of the most important models of complex networks as well as clustering in networks. The model used in this paper to generate the search spaces, the hidden metric model, is presented in Section IV. The experimental analysis is presented in Section V. Finally, Section VI concludes the paper and gives some future research directions.

978-1-4799-8713-9/15 $31.00 © 2015 IEEE
DOI 10.1109/UKSim.2015.97

II. Related Work

Simulated annealing is one of the most studied optimization algorithms. Over the years, several variations have been proposed in the literature [3], [10]. Certain variations of simulated annealing are concerned with the choice of neighbors, whereas other variants deal with the acceptance probability, that is, the probability of making an uphill move. Examples of the latter is the thresholding acceptance method [11], [12], where the acceptance probability is chosen to be:

$$p = \begin{cases} 1 & \text{if } -\Delta E \leq T, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

This acceptance rule relinquishes the stochastic nature of the original simulated annealing and replaces it with a deterministic transition rule. It is shown that this can lead to improvement in the performance for certain classes of search spaces. Another example of acceptance rules is the one proposed in [13] (see also [14], [15]), called the Tsallis acceptance probability:

$$p = \begin{cases} 1 & \text{if } \Delta E \geq 0, \\ \left(1 - q\dfrac{\Delta E}{T}\right)^{1/q} & \text{if: } \Delta E < 0, -q\Delta E/T \leq 1, \\ 0 & \text{if: } \Delta E < 0, -q\Delta E/T > 1, \end{cases} \quad (3)$$

where $q \neq 0$.

In [8], we proposed a new acceptance probability that takes into consideration the topology of the network through the incorporation of the states' degrees (or branching factors) into the acceptance probability. The latter is defined as:

$$p = \left(1 - \frac{\Delta E \bar{k}}{T k_{next}}\right)^{-\alpha'}, \quad (4)$$

where k_{next} is the degree of the candidate state under consideration, $\bar{k}$ is the average degree of all states in the search space and $\alpha' > 1$ is a parameter of the algorithm. By giving more weight to states that have larger degrees, this acceptance rule tends to accept states that offer a better chance of exploring larger parts of the state space. This strategy facilitates escaping from local minima, since highly connected nodes are more likely to be connected to nodes that are outside the local minimum in which the algorithm might be stuck.

The experiments show that the topology-aware variant improves the performance in many situations compared to simulated annealing, especially when the objective function has several deep local minima. Furthermore, from a computational perspective, the new acceptance rule does not add significant computational cost or memory requirements. The results showed also that the performance of both algorithms depended on the type of networks used as a state space. In particular, the results were better when using a scale-free network. Since the main the difference between random networks and scale-free networks is that scale-free networks have higher clustering, we decided to investigate the effect it has on the performance of simulated annealing and its topology-aware variant.

III. Networks and Clustering

The study of networks (or graphs) as mathematical structures and the associated algorithms is a relatively old mathematical discipline. However, the study of real networks, such as food webs, protein-protein interaction networks, social networks, and the Internet is a relatively recent endeavor [9]. The study of networks has been for years dominated by oversimplifying assumptions that were meant to facilitate the analysis, but which, unfortunately, kept the application of the obtained results and models on real-life networks infeasible. With the increasing availability of data and the proliferation of powerful computers, attention to the analysis of natural networks has emerged. The early studies on social and technological networks revealed, surprisingly, features and topological properties that are different from what the models existing at the time predicted. Nevertheless, many real life networks have been found to share common properties, such as small diameters and power-law degree distributions, even though their individual components (the nodes) are of completely different nature [9]. This triggered a large research effort to analyze and model networks and their properties.

Several models have been proposed over the years to capture the essential properties of the topology of networks and their growth. The most widely known models are:

- The random network model [16]: This is the oldest network model and has been used for several decades. Its premise is that any pair of nodes in the network are connected randomly with a fixed probability. This model is quite simple and thus easily amenable to theoretical analysis. However, it fails to capture some important properties of real networks such as clustering and power-law degree distributions.
- The small world model [17]: In this model, the network is constructed by starting from a regular grid, then rerouting links randomly with a given probability p. The resulting networks constitute a middle ground between regular grids and random networks and possess a small diameter (the small-world property).
- The preferential attachment model [18]: In this model, nodes tend to link with higher probability to nodes that have already a high degree. This mimics the "rich-gt-richer" phenomenon observed in many real-life networks.
- The hidden metric model [19]: In this model, the topology of the network is dictated by a hidden metric space were the nodes reside. The existence of the links is affected by the distance between the nodes and their popularity. This model allow to generate scale-free networks that have small diameter and are well

clustered. This model is presented in more details in the next section.

One of the salient properties of real life networks is the combination of the small world property and good clustering. The latter can be seen as the tendency of nodes to form communities. Several measures have been proposed to quantify the level of clustering in the network. Among these, the most known measure is the clustering coefficient, which is defined locally for a node n_i in a network (V, E) as:

$$C_i = \frac{|\{e_{jk}|n_j, n_k \in N_i, e_{jk} \in E|}{k_i(k_i - 1)}, \tag{5}$$

where e_{jk} is the edge linking nodes n_j and n_k, N_i is the set of neighbors of node n_i and k_i is its degree. The coefficient C_i takes value in the interval $[0, 1]$. It reaches the maximum value 1 when all nodes adjacent to n_i are connected, and falls to 0 when no links exist between them. The network average clustering coefficient is obtained by taking the average of local clustering coefficients over all nodes.

IV. The Hidden Metric Model

One of the most successful models of complex networks is the hidden metric model proposed in [20] as a special case of a more general class of networks having a topology controlled by hidden variables [21]. The model assumes the existence of an underlying metric space that controls the topology of the network. Each node has a position in this metric space, so that nearby nodes have high similarity. It is assumed that similar nodes tend to connect, as it is the case in real life networks. For example, people living nearby tend to form social ties and computers that are geographical close tend to be interconnected. In addition to similarity, the model incorporates the concept of popularity, which suggests that nodes differ in their capability to connect to other nodes. Some nodes have higher tendency to connect to other nodes, even if these are not similar to them. Such nodes are called popular and play the role of the hubs in the network.

In the hidden metric model, two nodes i, j are connected with probability p_{ij} defined as follows:

$$p_{ij} = \left(1 + \frac{d_{ij}}{\beta \kappa_i \kappa_j}\right)^{-\alpha}, \tag{6}$$

where d_{ij} is the distance between the two nodes, $\beta > 0$ is a parameter that controls the scale, κ_i and κ_j are the expected degrees of the two nodes i and j, and $\alpha > 1$ is a parameter that controls clustering. Notice that changing the value of β amounts to rescaling all the distance by a constant. The parameter α has a more critical role as shown later in this section.

The form of p_{ij} is such that nodes with small degrees can only be connected if they are highly similar (the distance d_{ij} is small). On the other hand, nodes with high degrees can be connected even if they are far away. Fig. 1 shows an example of a network generated by a hidden metric model. The network consists essentially of two clusters connected by a single link. Within each cluster, the nodes are highly connect, because they are close to each other. The two hubs (drawn with a larger size in the figure) are connected to nearby nodes and also to each other, despite the distance between the two being relatively large. Notice that navigation within each cluster involves small displacements, whereas moving between the hubs allows to cross large distances.

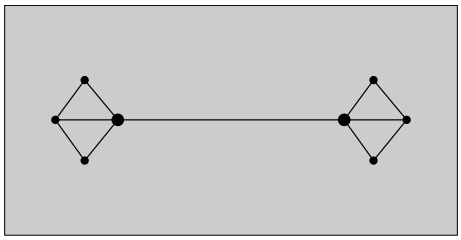


Figure 1: Illustration of a network generated by a hidden metric model.

The parameter α affects the clustering in the network [19]. Small values of α result in little clustering, whereas large values cause high clustering. Fig. 2 shows some examples of networks generated using different values of α. As α increases, the number of long range connections decreases leaving place to local connections, which indicates that the network is highly clustered.

V. Experimental Results

In this section, we investigate the effect of the clustering on the performance of simulated annealing and its topology-aware variant. The tests are conducted on scale-free networks generated by the hidden metric model as follows:

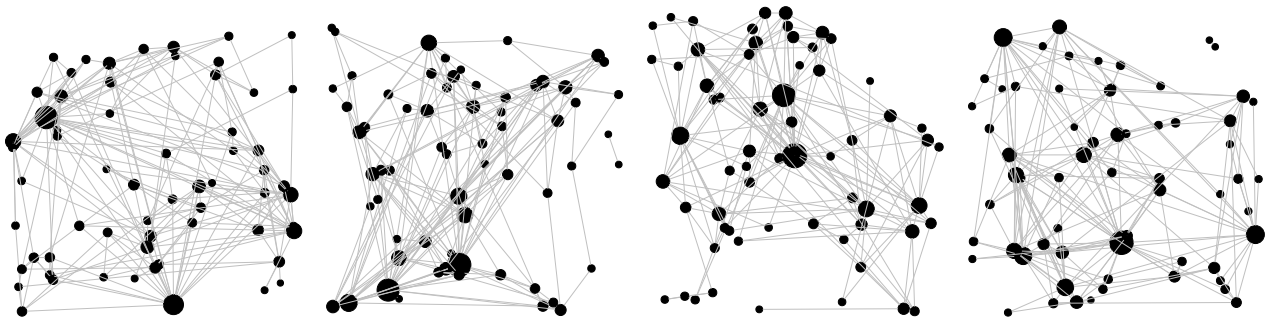
1) First, each node is assigned a position in the two dimensional domain $[0, 1]^2$. The coordinates of node i are denoted by (x_i, y_i).
2) In the second step, each node is assigned an expected (or hidden) degree. The expected degrees of the nodes are drawn from a power law distribution $\sim k^{\gamma}$.
3) Finally, the nodes are connected according to the probability given in Eq. (6). The parameter α is given several values to induce different levels of clustering. The parameter β is chosen to approximately achieve a target average degree.

Each node i is assigned an objective value computed as $f(x_i, y_i)$, where f is a continuous objective function on $[0, 1]^2$. Four well known objective functions in the field of global optimization are used in the experiment:

- Easom function: this function is defined as :

$$-\cos(x)\cos(y)\, e^{\left(-(x-\pi)^2-(y-\pi)^2\right)}, \tag{7}$$

and defined on the interval $[-10, 10]^2$. It has several local minima and global minimum at $f(\pi, \pi) = -1$.



(a) $\alpha = 2.0$ (b) $\alpha = 3.0$ (c) $\alpha = 4.0$ (d) $\alpha = 5.0$

Figure 2: Example of small networks generated with different values of α but with almost similar average degree.

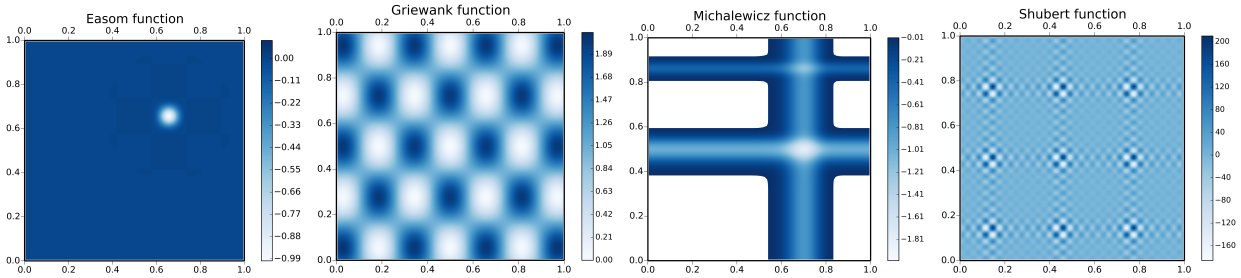
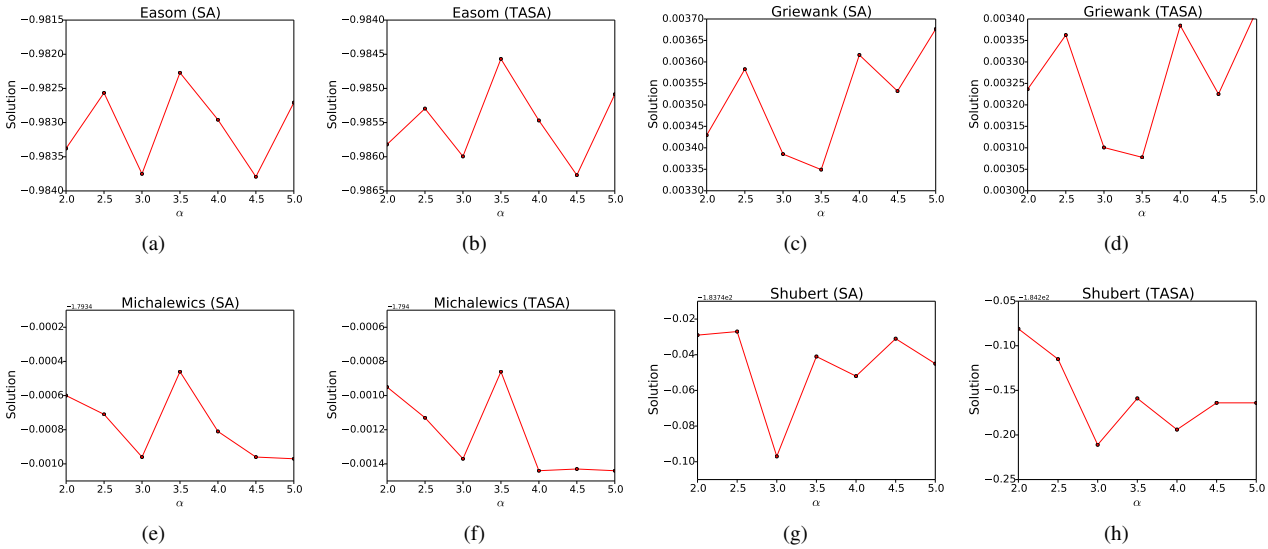


Figure 3: The four objective functions used in performance evaluation.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: The average value of the solution obtained by simulated annealing (SA) and topology-aware simulated annealing (TASA) for all four objective functions when using different values of α.

- Griewank function: this function is defined as :

$$\left(x^2+y^2\right)/4000-\cos\left(x\right)\cos\left(y/\sqrt{2}\right)+1, \quad (8)$$

  and defined on the interval $[-10,10]^2$. It has several local minima and a global minimum at $f\left(0,0\right)=0$.
- Michalewicz function: this function is defined as :

$$-\sin\left(x\right)\sin\left(x^2/\pi\right)^{20}-\sin\left(y\right)\sin\left(2y^2/\pi\right)^{20}, \quad (9)$$

  and defined on the interval $[0,\pi]^2$. It has several local minima and a global minimum at $f\left(2.20,1.57\right)=-1.80$.
- Shubert function: this function is defined as :

$$\left(\sum_{i=1}^{5} i\cos\left(\left(i+1\right)x+i\right)\right)\left(\sum_{i=1}^{5} i\cos\left(\left(i+1\right)y+i\right)\right), \quad (10)$$

  and defined on the interval $[-10,10]^2$. It has several local minima and several global minima with value -186.73.

The domain of the coordinates, that is $[0,1]^2$, is scaled and translated to match the domain of definition of each of these objective functions.

Since the nodes are not linked only based on locality, the landscape of the objective function over the network may be different from the landscape of the function as it appears normally in the domain $[0,1]^2$. The landscape over the network may be far more disordered than the initial function. The objective value may vary largely between two neighboring nodes. This can make the optimization problem more difficult, but at the same time it offers shortcuts that can be used to escape local minima.

In the experiment, 100 network having 10^5 nodes are created using the hidden metric model for different values of α ranging from 2.0 to 5.0. The parameter γ is assigned the value 2.5. This is a value that is usually encountered in natural complex networks [9]. The scale parameter β is chosen so that an average degree $\bar{k}\simeq 12$. For each objective function, the two algorithms are run using 10^2 different initial states selected randomly.

Both algorithms are run on the same problem instances. This includes the network, the initial state, the temperature schedule and the random seed. Since the temperature schedule is deterministic, the algorithms are run for the exact same number of iterations. The temperature is initialized to 10^4, and it is decreased at each iteration by multiplying by 0.99. The schedule terminates when reaching the temperature 10^{-2}. Topology-aware simulated annealing is run with the value $\alpha'=2$ (see Eq. (4)).

The average value of the solution obtained by both algorithms for all four objective functions as α varies are shown in Fig. 4. Table I shows the results of the one-sided two-sample t-test for the mean. The "t" column shows the t-test score, whereas the "p" column gives the test p-value. In this t-test, the null hypothesis is that the mean solution obtained using the best value of α is not smaller than the mean solution obtained using the other values of α. The goal is to test whether the change in α significantly impacts the solution quality.

The results show that the value of α affects the performance of the two algorithms in a non-trivial way. For certain values of α, that is for certain levels of clustering, the algorithms achieve better performance than other values. Furthermore, this effect is statistically significant as can be seen from Table I and the results of the t-test therein. This is interesting given the fact that the topology of the network is highly independent of the objective function. This shows that for certain levels of clustering, the exploration of the state space is achieved more efficiently.

The dependence of performance on α is clear in all four functions, but the results are more interesting for Griewank and Shubert functions. From Fig. 4, it can be seen that for Griewank and Shubert functions, there exists an intermediate value of α for which the best performance is achieved, $\alpha=3.5$ for Griewank and $\alpha=3.0$ for Shubert. Beyond this value, increasing α causes the performance to decrease. These two functions are characterized by several deep local minima. Low clustering can cause poor local exploration of the state space and hence can result in low performance. High clustering, on the other hand, can cause the algorithm to get stuck in local minima, since all nearby nodes are connected. State spaces with several deep local minima are particularly hard to deal with when using simulated annealing. This is in contrast to spaces with shallow local minima (as it is the case for Easom and Michalewicz functions). The possibility of improving the performance of simulated annealing and its topology-aware variants by merely changing the clustering of the network (independently of the objective function) can serve as an important tool for designing better optimization algorithms that balance state space exploration and local improvement in a principled way.

VI. Conclusion

In this paper, we investigated the effect of state space clustering on the performance of simulated annealing and its topology-aware variant. The hidden metric model was used for testing this effect. This model has the benefit of allowing to create scale-free networks that are clustered and possess the small world property. The clustering level can be controlled by varying the parameter α of the model. The results show that the dependence of the performance on the clustering level is nontrivial. This means that there are certain degrees of clustering that can boost the performance of the algorithms. These clustering degrees depend of course on the landscape of the objective function.

As a future work, we propose to study the combined effect of the shape of the objective function and network

Table I: THE t-TEST RESULTS OF COMPARING THE MEAN OF THE SOLUTION OBTAINED USING THE BEST VALUE OF α AGAINST THE MEAN SOLUTION OBTAINED USING THE OTHER VALUES.

Function	Algorithm	α													
		2.0		2.5		3.0		3.5		4.0		4.5		5.0	
		p	t	p	t	p	t	p	t	p	t	p	t	p	t
Easom	SA	0.04	-1.77	0.00	-5.16	0.43	-0.19	0.00	-6.53	0.00	-3.68	-	-	0.00	-4.64
	TASA	0.01	-2.22	0.00	-4.78	0.08	-1.40	0.00	-8.51	0.00	-4.10	-	-	0.00	-5.88
Griewank	SA	0.02	-2.06	0.00	-6.04	0.17	-0.95	-	-	0.00	-6.91	0.00	-4.66	0.00	-8.52
	TASA	0.00	-4.25	0.00	-7.67	0.26	-0.63	-	-	0.00	-8.22	0.00	-3.94	0.00	-9.29
Michalewicz	SA	0.00	-3.81	0.00	-2.63	0.46	-0.09	0.00	-5.25	0.05	-1.62	0.44	-0.15	-	-
	TASA	0.00	-5.89	0.00	-3.82	0.20	-0.84	0.00	-7.04	-	-	0.43	-0.18	0.50	-0.00
Shubert	SA	0.05	-1.66	0.05	-1.67	-	-	0.09	-1.36	0.14	-1.08	0.06	-1.59	0.10	-1.26
	TASA	0.00	-3.93	0.00	-2.91	-	-	0.06	-1.57	0.31	-0.51	0.08	-1.43	0.07	-1.46

clustering on the performance of the two algorithms (simulated annealing and topology-ware simulated annealing). Capturing quantitatively the interaction between the objective function variations and network clustering can allow to dynamically adjust the network in response to the information collected on the objective function during the optimization process.

REFERENCES

[1] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi *et al.*, "Optimization by simmulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[2] V. Černỳ, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," *Journal of optimization theory and applications*, vol. 45, no. 1, pp. 41–51, 1985.

[3] P. Salamon, P. Sibani, and R. Frost, *Facts, conjectures, and improvements for simulated annealing*. Siam, 2002.

[4] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[5] X. Geng, Z. Chen, W. Yang, D. Shi, and K. Zhao, "Solving the traveling salesman problem based on an adaptive simulated annealing algorithm with greedy search," *Applied Soft Computing*, vol. 11, no. 4, pp. 3680–3689, 2011.

[6] R. Romero, R. Gallego, and A. Monticelli, "Transmission system expansion planning by simulated annealing," *Power Systems, IEEE Transactions on*, vol. 11, no. 1, pp. 364–369, 1996.

[7] C. Koulamas, S. Antony, and R. Jaen, "A survey of simulated annealing applications to operations research problems," *Omega*, vol. 22, no. 1, pp. 41–56, 1994.

[8] S. Kerrache and H. Benhidour, "Topology-aware simulated annealing," in *International Conference on Artificial Intelligence, Modelling and Simulation*, 2014.

[9] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[10] E. Aarts, J. Korst, and W. Michiels, "Simulated annealing," in *Search methodologies*. Springer, 2005, pp. 187–210.

[11] P. Moscato and J. F. Fontanari, "Convergence and Finite-time Behavior of Simulated Annealing," *Advances in Applied Probability*, vol. 18, pp. 747–771, 1990.

[12] G. Dueck and T. Scheuer, "Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing," *Journal of Computational Physics*, vol. 90, no. 1, pp. 161–175, Sep. 1990.

[13] C. Tsallis and D. A. Stariolo, "Generalized simulated annealing," *Physica A: Statistical Mechanics and its Applications*, vol. 233, no. 1-2, pp. 395–406, Nov. 1996.

[14] A. Franz and K. H. Hoffmann, "Optimal annealing schedules for a modified tsallis statistics," *J. Comput. Phys.*, vol. 176, no. 1, pp. 196–204, Feb. 2002.

[15] T. J. P. Penna, "Traveling salesman problem and Tsallis statistics," *Physical Review E*, vol. 51, pp. R1–R3, Jan. 1995.

[16] P. Erdös and A. Rényi, "On the Evolution of Random Graphs," in *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 1960, pp. 17–61.

[17] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.

[18] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.

[19] M. Boguna, D. Krioukov, and K. C. Claffy, "Navigability of complex networks," *Nature Physics*, vol. 5, no. 1, pp. 74–80, Nov. 2008.

[20] M. Serrano, D. Krioukov, and M. Boguna, "Self-similarity of complex networks and hidden metric spaces," *Physical Review Letters*, vol. 100, no. 078701, Feb 2008.

[21] M. Boguñá and R. Pastor-Satorras, "Class of correlated random networks with hidden variables," *Phys. Rev. E*, vol. 68, no. 3, p. 36112, 2003.