Ranking important nodes in complex networks by simulated annealing*

Yu Sun(孙昱)[1,†], Pei-Yang Yao(姚佩阳)[1], Lu-Jun Wan(万路军)[2], Jian Shen(申健)[1], and Yun Zhong(钟赟)[1]

[1] *Information and Navigation College, Air Force Engineering University, Xi'an* 710077, *China*
[2] *Air Traffic Control and Navigation College, Air Force Engineering University, Xi'an* 710077, *China*

(Received 15 July 2016; revised manuscript received 31 October 2016; published online 20 December 2016)

In this paper, based on simulated annealing a new method to rank important nodes in complex networks is presented. First, the concept of an importance sequence (IS) to describe the relative importance of nodes in complex networks is defined. Then, a measure used to evaluate the reasonability of an IS is designed. By comparing an IS and the measure of its reasonability to a state of complex networks and the energy of the state, respectively, the method finds the ground state of complex networks by simulated annealing. In other words, the method can construct a most reasonable IS. The results of experiments on real and artificial networks show that this ranking method not only is effective but also can be applied to different kinds of complex networks.

Keywords: complex networks, node importance, ranking method, simulated annealing

PACS: 02.10.Ox, 89.20.Ff, 89.75.Fb **DOI:** 10.1088/1674-1056/26/2/020201

1. Introduction

It is of great significance in real life to take full advantage of important nodes in complex networks. For example, protecting and backing up key routers and servers in the internet will reduce the damage caused by network attacks, building an integrated transportation hub for main traffic junctions in road networks will alleviate traffic congestion, and issuing information through authorized media or high-profile persons in social networks will guide public opinion more effectively. Due to their huge utility, how to identify important nodes in complex networks has been a wide concern.[1–3]

Methods of ranking nodes in complex networks usually use the topological properties of nodes to judge their importance. Commonly, these ranking methods are divided into a single-index ranking method (SRM) and multiple-index ranking method (MRM). As the name suggests, SRM evaluates the importance of nodes with only one topological property of these nodes, such as degree centrality, betweenness centrality, and closeness centrality. Besides, Singh *et al.* proposed structural centrality to measure the importance of nodes, which was defined on the basis of a generalized inverse matrix of a Laplacian matrix of a complex network and reflected the distance from a node to other nodes.[4] Nie *et al.* constructed a new index to rank nodes according to the view that the more important the neighbor nodes of a node, the more important the node was.[5] Unlike what was done in Ref. [5], Hu *et al.* considered the contributions made by both neighbor nodes and non-neighbor nodes of a node to the importance of this node[6] while Liu *et al.* took the contribution of each edge of a node into account.[7] Compared with SRM, MRM is thought to be very accurate because it uses several topological properties of nodes to assess their importance. The main step of MRM is to integrate typical topological properties of nodes into a comprehensive measure of node importance through linear or nonlinear transition. The way of integrating topological properties is divided into a multi-attribute decision making method and a data dimension reduction method. The widely adopted multi-attribute decision making methods in ranking nodes include an ideal point method[8–11] and a hierarchy analytic method[12] while the commonly used data dimension reduction methods are the principal component analysis method[13] and the local linear embedded method.[14] Except for the ranking methods mentioned above, some researchers explored other ways. For example, Ventresca *et al.* searched for important nodes through a multi-objective evolutionary algorithm[15] and Xiao established a super-network model for evaluating the node importance.[16]

It can be seen from most researches that before ranking nodes, an index set constituted by one or several topological properties of nodes should be determined. But unfortunately, it is difficult to prove the completeness of an index set used in measuring the node importance. Thus, in this paper we try a different way to rank nodes by bringing in the simulated annealing method so that constructing an index set of nodes is not needed any more. At the end of this paper, the effectiveness of the new ranking method is demonstrated through an

*Project supported by the National Natural Science Foundation of China (Grant No. 61573017) and the Natural Science Foundation of Shaanxi Province, China (Grant No. 2016JQ6062).

†Corresponding author. E-mail: suny.z@qq.com

© 2017 Chinese Physical Society and IOP Publishing Ltd

experiment on real and artificial complex networks.

2. Theoretical foundation

In the field of condensed matter physics, annealing is an important way to obtain a solid substance in a low-energy state. The detailed process of annealing is that a solid substance is heated until it melts to liquid and then the temperature of the liquid is controlled to drop slowly until the liquid freeze to solid. During the fusion of the solid substance, the particles of the substance are transformed from a highly structured state to a disordered state, which is helpful to eliminate the inhomogeneity of particle distribution in the substance. During the solidification of a liquid substance, the particle distribution of the substance turns out to be uniform. As a result, the finally solidified substance will stay in a low-energy state called the ground state, which means that it is stable and has the least energy.

In order to simulate the state change of a substance during the solidification of a liquid, Metropolis *et al.* proposed an effective way.[17] Suppose that the current state of a substance is $S = s_i$ and a new state s_j is obtained from a slight disturbance of s_i, such as moving the position of a particle of the substance that stays in s_i then the next state of the substance will be

$$S = \begin{cases} s_i, & x > p, \\ s_j, & x \leq p, \end{cases} \tag{1}$$

where x is a random number in the interval (0,1) and p is called the new state acceptance probability (NSAP) that is defined as

$$p = \min\left(\exp\left(\frac{E_i - E_j}{k_B T}\right), 1\right), \tag{2}$$

where E_i and E_j represent the energies of the states s_i and s_j, respectively; k_{B} is the Boltzmann constant; T is the current temperature.

If the initial state of a liquid substance is known, then a state sequence can be constructed according to the rule introduced above. Since the next state of the substance is only related to its current state, a state sequence constructed can be treated as a Markov chain. Under a certain temperature T, the steady distribution of the Markov chain is the Boltzmann distribution. In other words, if the set $H = \{s_1, s_2, \ldots\}$ is constituted by all states that may appear in the annealing process, then the possibility of the event that the substance state changes from s_x $(s_x \in H)$ to s_y $(s_y \in H)$ after a series of state transitions is

$$P(S = s_y) = \frac{\exp(-E_y/k_{\mathrm{B}}T)}{\sum\limits_{s_k \in H} \exp(-E_k/k_{\mathrm{B}}T)}, \tag{3}$$

where E_k and E_y are the energies of states s_k and s_y.

Suppose that the state s_z $(s_z \in H)$ is the ground state of the substance, then according to formula (3), the smaller the temperature T, the larger the possibility P $(S = s_z)$ is. During the solidification of a liquid substance, the temperature T decreases continuously and the state sequence becomes longer and longer. As a result, the final state of the substance will stay in the ground state s_z.

3. Ranking method based on simulated annealing

The rank result of important nodes in a complex network can be represented by a node sequence. In the sequence, that the position of a node A is in front of the position of a node B indicates that node A is more important than node B. As a result, the most important node is the first one in the sequence while the least important node is the last one in the sequence. Such a node sequence of a complex network is called an importance sequence in this paper.

Since different ranking methods result in different important sequences, how to judge the reasonability of an important sequence becomes quite critical. It is well known that removing important nodes in complex networks will cause great damage to the networks. In other words, if nodes in a complex network are removed one by one according to their order in an important sequence, then the complex network will collapse quickly. The more reasonable the important sequence, the faster the collapse speed of the network is. Suppose that G is a complex network containing N nodes and Q is an important sequence of G. If the nodes in G are removed one by one according to their order in Q, then the collapse speed of G is defined as

$$M(Q) = \frac{1}{N}\sum_{i=1}^{N} C_i, \tag{4}$$

where C_i is the size of the largest connected component (LCC) of G after i nodes in G have been removed. The smaller the $M(Q)$, the faster the collapse speed of the network is, which also indicates how more reasonable the importance sequence Q is.

If a complex network G is compared to a substance, an important sequence Q of G is compared to a state of the substance and the reasonability measure $M(Q)$ is compared to the energy of the state, then the ground state of G can be found by means of the simulated annealing introduced in Section 2. That is, the most reasonable important sequence of G can be constructed.

The detailed flow chart of our ranking method based on simulated annealing is shown in Fig. 1. In this figure, Q is an initial important sequence generated randomly, T_0 is the initial temperature, V is the times of temperature decrease, α is the temperature drop coefficient, and L is the times of state transition at a fixed temperature.

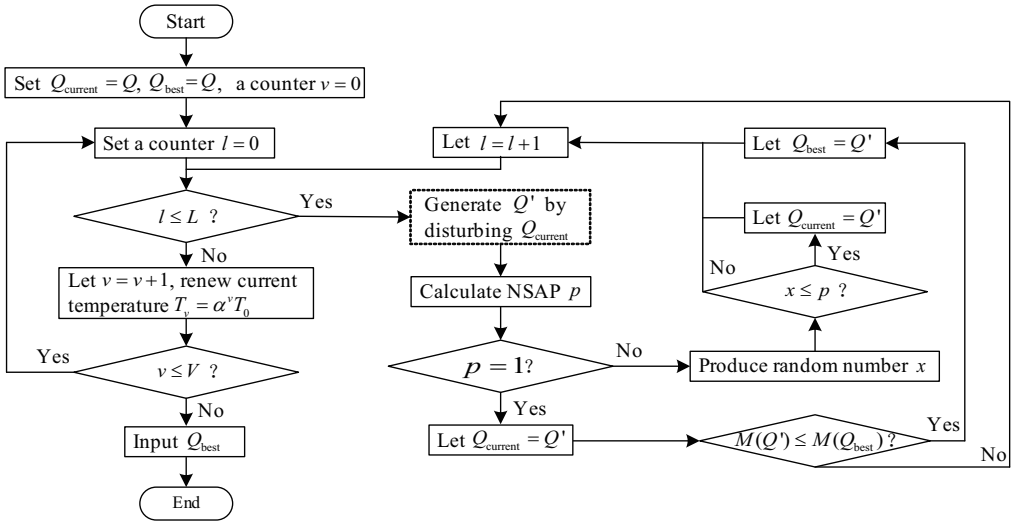


Fig. 1. (color online) Flow chart of the proposed ranking method based on simulated annealing.

The detailed process of generating Q' by disturbing Q_{current} in the dotted line box in the figure is as follows. Choose a pair of nodes in Q_{current} randomly and then exchange their positions. A new importance sequence obtained just through this operation is Q'.

If the parameters of the ranking method based on simulated annealing are well set, the most reasonable important sequence will be found finally no matter what the initial important sequence is. Thus, we set the initial temperature T_0 to be

$$T_0 = \frac{E^{\max} - E^{\min}}{k_{\text{B}}}, \tag{5}$$

which is suggested in Ref. [18]. In formula (5), $E^{\min}$ and $E^{\max}$ represent the smallest energy and the largest energy that a state may have during the simulated annealing. For a network containing N nodes, the extreme collapse scenes are shown in Fig. 2.

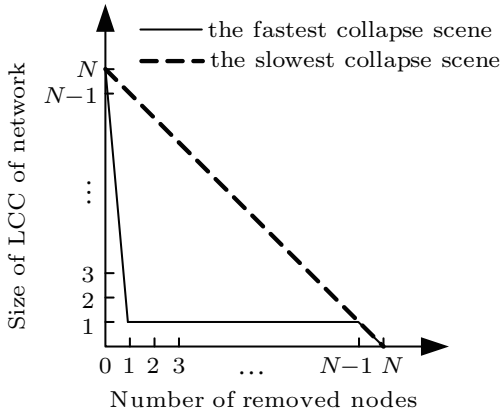


Fig. 2. Extreme collapse scene of networks.

The fastest collapse scene appears if the hub node of a star network is first removed from the network while the slowest collapse scene happens if the nodes of a network which is a complete graph are removed. According to Fig. 2 and formula (4), the fastest collapse speed of a network containing N nodes is $(N-1)/N$ while the slowest collapse speed is $(N-1)/2$. Since the smallest energy $E^{\min}$ and the largest energy $E^{\max}$ are equal to the fastest collapse speed and the slowest collapse speed in value in this paper respectively, the initial temperature T_0 is $(N-1)(N-2)/(2Nk_{\text{B}})$. Besides, the times of temperature decrease V is set to 100 to make the final temperature low enough, the temperature drop coefficient α is set to be between 0.8 and 0.99 to speed up the convergence of the ranking method, and the times of state transition at a fixed temperature L is set to be $10N$ which is related to the number of nodes needed to be ranked.

4. Experiment results

To validate the effectiveness of the proposed ranking method, it is tested on different real networks and artificial networks. The used real networks are the network of Jazz musicians (NJM),[19] the network of the Beijing subway (NBS),[20] and the network of American airlines (NAA),[21] whose information is shown in Table 1.

Table 1. Information of real networks.

Network name	Number of nodes	Number of edges
NJM	198	2742
NBS	245	280
NAA	332	2126

We apply the proposed method to the real networks mentioned above. The results are shown in Fig. 3 in which the top 10% of important nodes of each network are marked with a red color.

If the nodes marked in Fig. 3 are removed, then the LCC of NJM decreases from 198 to 125, that of NBS decreases

from 245 to 14, and that of NAA decreases from 332 to 69. To demonstrate the effectiveness of the ranking method based on simulated annealing (MSA), it is compared with two other newly developed methods which are the ranking method based on mapping entropy (MME)[6] and the ranking method based on multi-attribute (MMA).[9] According to different methods, different importance sequences are obtained. If one removes nodes from these networks one by one according to their order in the important sequences, then the size change of LCC of these real networks will be as shown in Fig. 4. For comparison, the effect based on node random removal (NRR) is also displayed in this figure.

The more reasonable the important sequence, the faster the collapse speed of a corresponding network is. According to Fig. 4, it is known that all the three node ranking methods have some kind of rationality for the collapse speed caused by them which is faster than that cased by NRR. Among the three node ranking methods, it can be observed from Fig. 4 that the efficiency of MSA is much better than those of MME and MMA. In other words, when used to identify important nodes of complex networks, the proposed method is feasible and effective.

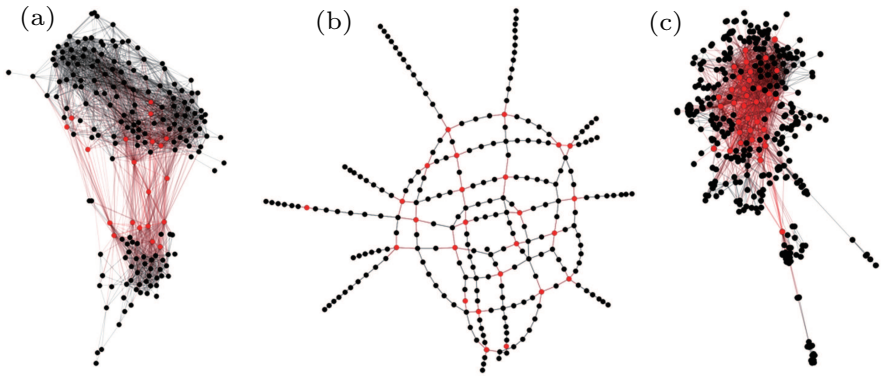


Fig. 3. (color online) Important nodes of real networks of (a) NJM, (b) NBS, and (c) NAA.

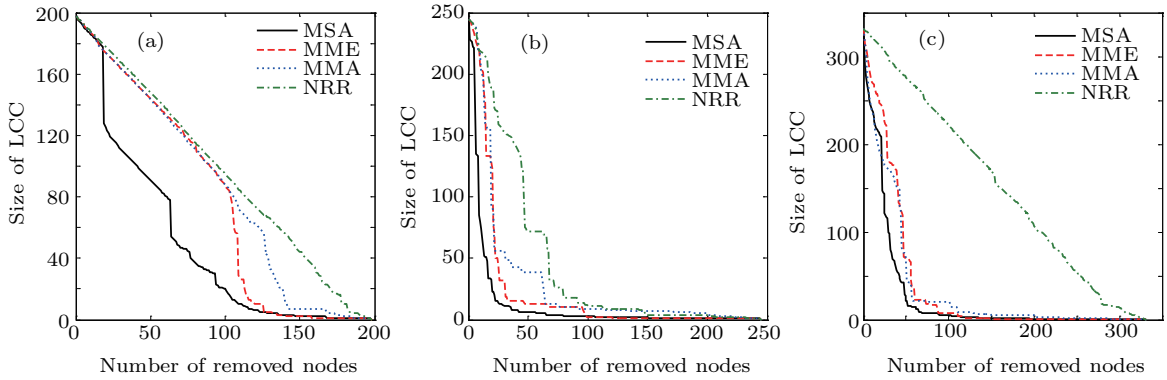


Fig. 4. (color online) Variations of size of LCC with the number of removed nodes of real networks: (a) NJM, (b) NBS, (c) NAA.

Table 2. Information of artificial networks.

Network name	AN_1	AN_2	AN_3	AN_4	AN_5	AN_6	AN_7	AN_8	AN_9
Type	random	random	random	small-world	small-world	small-world	scale-free	scale-free	scale-free
Number of nodes	200	200	200	200	200	200	200	200	200
Number of edges	400	600	800	400	600	800	400	600	800

To compare these methods further, in this paper we run them on 9 artificial networks (ANs) whose information is shown in Table 2. The experimental results are shown in Fig. 5.

It is well known that the scale-free network tends to collapse more quickly than the random network when both of them are under malicious node attack. The experimental results in Fig. 5 demonstrate the conclusion once again. From Fig. 5, it can also be seen that the performance of MSA is superior to those of other methods in each type of complex network. In detail, the MSA is the best, the MME takes the second place, the MMA is the third, and of course the NRR is the last. In addition, it should be pointed out that the greatest advantage of MSA is that it does not need to construct an index set before ranking nodes, thereby it may avoid the bad effect brought by the incompleteness of the index set chosen.

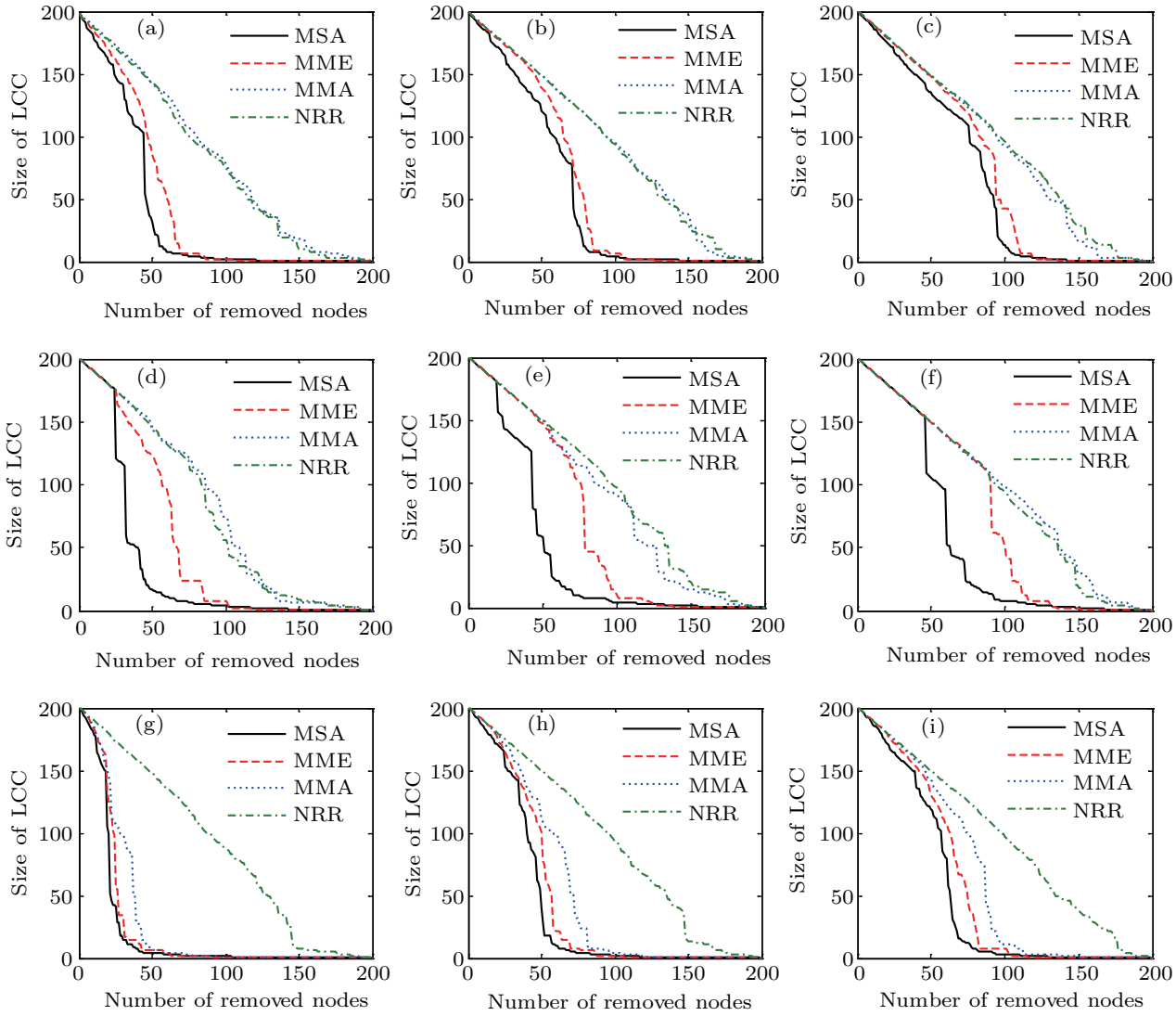


Fig. 5. (color online) Variations of size change of LCC with the number of removed nodes of artificial networks: (a) AN_1, (b) AN_2, (c) AN_3, (d) AN_4, (e) AN_5, (f) AN_6, (g) AN_7, (h) AN_8, (i) AN_9.

5. Discussion

Though the ranking method based on simulated annealing (MSA) performs well in our experiment, its computational complexity reaches $O(LVN^3)$ while those of MME and the ranking method based on simulated annealing are $O(N)$ and $O(N^3)$. As a result, the average time cost of the ranking method based on simulated annealing is 3.7 h when running on the artificial complex networks shown in Table 2, which is quite a bit larger than those of MME and MMA. Thus, the good performance of the ranking method based on simulated annealing is achieved at the expense of high computational complexity. Considering that the real complex networks contain thousands or even millions of nodes, the practicability of the ranking method based on simulated annealing still needs to be improved and one available way is to reconstruct the ranking method based on simulated annealing with the idea of parallel design. Since our method of solving the node rank problem is different from traditional ones, we believe that it will promote the development and improvement of future methods.

6. Conclusions

In this paper we study the nodes ranking problem in complex networks and propose a new ranking method based on simulated annealing. The proposed method describes the concept of node importance sequence and defines a reasonability measure for it. By comparing an important sequence and its reasonability measure to a state of complex network and the energy of the state, respectively, this method constructs the most reasonable important sequence with the help of simulated annealing. At the end of this paper, the effectiveness of the new method is validated through experiments on different types of complex networks. In the future, we will be devoted

to reducing the computational complexity of this method for its practicability in large-scale complex networks.

References

[1] Sheikhahmadia A, Nematbakhsha M A and Shokrollahi A 2015 *Physica A* **436** 833

[2] Zhao J, Yu L, Li J R and Zhou P 2015 *Chin. Phys. B* **24** 058904

[3] Yu Y and Fan S H 2015 *J. Inform. Comput. Sci.* **12** 1281

[4] Singh A, Kumar R and Singh Y N 2015 *Acta Phys. Polon. B* **46** 305

[5] Nie T Y, Guo Z, Zhao K and Lu Z M 2016 *Physica A* **453** 290

[6] Hu P, Fan W L and Mei S W 2015 *Physica A* **429** 169

[7] Liu J, Xiong Q Y, Shi W R, Shi X and Wang K 2016 *Physica A* **452** 209

[8] Du Y X, Gao C, Hu Y, Mahadevanc S and Deng Y 2014 *Physica A* **399** 57

[9] Liu Z H, Jiang C, Wang J Y and Yu H 2015 *Knowledge-Based Systems* **84** 56

[10] Yang Y Y and Xie G 2016 *Inform. Process. Manag.* **52** 911

[11] Hu J T, Du Y X, Mo H M, Wei D J and Deng Y 2016 *Physica A* **444** 73

[12] Xu Y J, Gao Z, Xiao B, Meng F Y and Lin Z Q 2013 *Proc. IEEE International Conference on Broadband Network & Multimedia Technology*, November 17–19, 2013, Guilin, China, p. 105

[13] Wang P, Yu X H, Lv J H and Chen A M 2014 *Proc. IEEE International Symposium on Circuits and Systems*, June 1–5, 2014, Melbourne, Australia, p. 1267

[14] Hu F, Liu Y H and Jin J Z 2014 *Proc. International Symposium on Distributed Computing and Applications to Business, Engineering and Science*, November 24–27, 2014, Xianning, China, p. 138

[15] Ventresca M, Harrison K R and Ombuki-Berman B M 2015 *Lecture Notes in Computer Science* **9028** 164

[16] Xiao Q 2016 *Tehnički Vjesnik* **23** 397

[17] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087

[18] Burke E K and Kendall G 2005 *Search methodologies: introductory tutorials in optimization and decision support techniques* (New York: Springer) p. 187

[19] Pablo G and Leon D 2003 *Advances in Complex Systems* **6** 565

[20] Han C F and Liu L 2009 *Proc. IEEE International Conference on Engineering Management and Service Science*, September 20–22, 2009, Wuhan, China, p. 2

[21] Hidefumi S 2012 *Proc. IEEE World Congress on Computational Intelligence*, June 10–15, 2012, Brisbane, Australia, p. 1