A Novel Social Network Structural Balance Based on the Particle Swarm Optimization Algorithm

Li Zhao Xing[1], *He Li Le*[1], *Zhang Hui*[2]

[1] *School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China*
[2] *School of Information Engineering, Yulin University, Yulin 719000, China*
Emails: 39812767@qq.com   hllnh2013@163.com

Abstract: *Exploration of the structural balance of social networks is of great importance for theoretical analysis and practical use. This study modeled the structural balance of social networks as a mathematical optimization problem by using swarm intelligence, and an efficient discrete particle swarm optimization algorithm was proposed to solve the modeled optimization problem. To take advantage of the topologies of social networks in the algorithm design, the discrete representation of the particle was redefined, and the discrete particle update principles were redesigned. To validate the efficiency of the proposed algorithm, experiments were conducted using synthetic and real-world social networks. The experiments demonstrate that the proposed algorithm not only achieves a balanced social network structure, but also automatically detects the community topology of networks.*

Keywords: *Particle Swarm Optimization, social network, structural balance, community structure.*

1. Introduction

The prevalence of social networks is changing the way people live. Social media provide individuals with an outlet to express their viewpoints and concerns, and social relationships can be developed through social platforms. Analysis of social networks not only discloses their structural features, but can also predict the propagation of information along networks and its possible disasters. Therefore, network analysis is of great importance for the perspectives of theoretical research and practical applications [1].

One efficient method of analyzing a social network is to model it as a complex network composed of nodes and edges. The nodes represent the objects in the network and the edges denote the relationships between the objects. Analysis of the basic features of complex networks can help people get knowledge of the social network better and contributes to a great deal to decision making. Because friendship and antagonism generally exist between objects, the social networks are normally modeled as signed networks comprised of positive and negative edges, where a positive edge represents friendship and a negative edge denotes antagonism.

Complex networks [2, 3] have been found to exhibit numerous features of interest, such as small world features and scale-free features. In recent years, scholars have established another important feature exhibited by complex networks, i.e, the community structure characteristics [4]. The network community structure feature embodies the idea associated with the phrase "birds of a feather flock together". Research regarding network community structure has attracted considerable attention from scholars [5].

From the perspective of sociology, the network community structure feature lays particular stress on the macrostructure of a network, where a complex network can be viewed as an aggregation of several communities. On the other hand, other features, such as the structural balance feature, reveal the microstructure of a network. The network structural balance is also known as social balance. The concept was first proposed by Heider [6], and was later generalized by Cartwright and Harary from the perspective of graph theory [7, 8]. According to the theory of structural balance, a complex signed network is structurally balanced if it can be divided into two subsets, where the relations are friendly within each subset, and the relations are hostile between the subsets. More than 70 years have elapsed since the advent of the concept of structural balance; however, the study of structural balance still maintains the interest of scholars worldwide [9]. The theory of structural balance has been widely applied in fields such as social psychology, international relations and Internet.

Early researches on the network structural balance have focused on establishing the mathematical definition of the balance structure and formulating the balance criteria. However, this earlier definition easily leads to polarization between two sharply divided subsets, and later researchers have pointed out that real social networks can be divided into numerous subsets rather than just two. This indicates that the structural balance definition given in early studies is too strict. The general idea of recent network structural balance methods is to first determine the imbalanced edges, and to change them so as to make the network structurally balanced. However, these methods are not sufficiently sophisticated. Some of these methods must artificially specify the network division numbers in advance, but some methods, while achieving a balanced network structure, are unable to determine the community topology of the network.

The present paper considers the structural balance problem of social networks as a mathematical optimization problem, where the network structural balance is established by the minimization of an energy function. An efficient optimization

method based on a discrete particle swarm optimization algorithm is proposed to solve the minimization problem. Combined with the social network topology information, we have redefined the discrete particle representation schema and have also redesigned the particle update principles based on these schemas. In contrast to the existing network structural balance methods, the proposed algorithm not only achieves a balanced network structure, but also detects the community topology of the network. The method proposed can automatically determine the optimum community partitions. The efficiency of the proposed algorithm is validated using synthetic and real-world signed networks.

2. Network structural balance

The structural balance is an important part of social network research. For the networks composed of three nodes, shown in Fig. 1, structural balance theory defines the structures given in Figs 1(a) and (b) as balanced, whereas those given in Figs 1(c) and (d) are imbalanced. The "+" symbol in the figure represents a positive relationship, such as love or friendship, whereas a "−" denotes a negative relationship, such as dislike or animosity.
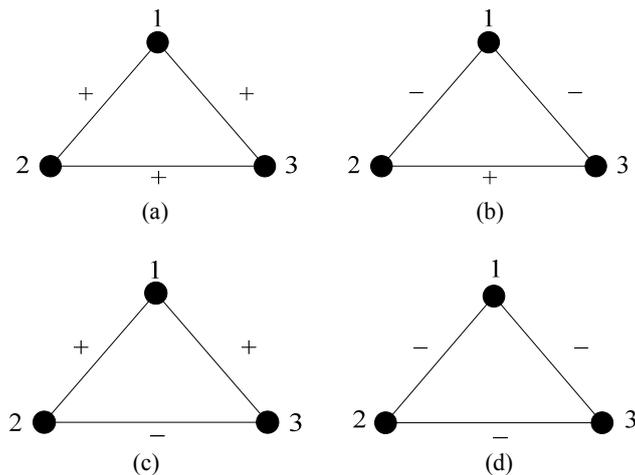

Fig. 1. An illustration of the structural balance where (a) and (b) are considered balanced networks, and (c) and (d) are imbalanced

For an arbitrary complex network, if the nodes of the network can be divided into two subsets X and Y, and if X and Y meet the relationship described in Fig. 2, then the network is considered structurally balanced.

However, a large number of studies have shown that the structural balance condition given in Fig. 2 is too strict, and that the majority of real-world networks can be divided into more than two subsets that can also satisfy the condition of X and Y. Real-world social networks are rarely structurally balanced, and most of them require changes in the network edge attributes or an increase or reduction of the number of network edges to achieve a balanced structure.
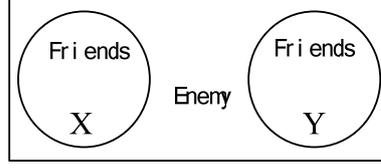
25

Fig. 2. A schematic describing of the condition for structural balance

3. Particle swarm optimization

Particle Swarm Optimization (PSO) involves the artificial modeling and simulation of the behavior of swarm intelligence attributed to the foraging and predator avoidance behavior of social animals [11]. PSO involves simple principles, few parameters, and rapid convergence, and has been widely used to solve complex optimization problems [12, 13].

PSO is used to optimize problems according to a set of particles, where each particle represents a candidate solution to the problem, and each particle iteratively adjusts their movement within the search space to reflect their own learning in accordance with some simple updating rules, ensuring that the particles tend to swarm towards the globally optimal solution of the problem.

We suppose that a particle swarm is comprised of n particles, the particle dimension of the search space is d, and $V_i = \{v_i^1, v_i^2, \cdots, v_i^d\}$ represents the velocity vector of i-th particle and $X_i = \{x_i^1, x_i^2, \cdots, x_i^d\}$ indicates its position vector. Therefore, the updating rules can be expressed as:

(1) $$v_i^{t+1} = v_i^t + c_1 r_1 (p_i^t - x_i^t) + c_2 r_2 (g^t - x_i^t),$$

(2) $$x_i^{t+1} = x_i^t + v_i^{t+1},$$

where $P_i = \{p_i^1, p_i^2, \cdots, p_i^d\}$ and $G = \{g^1, g^2, \cdots, g^d\}$ respectively represent the present and global optimal solutions of particle i during motion, c_1 and c_2 are the cognitive and social components, respectively, and r_1 and r_2 are two random numbers within the range from 0 up to 1.

PSO is an iterative algorithm, where each particle constantly adjusts their path of motion, and thus attains a position close to the optimal solution of the problem based on (1) and (2).

4. Particle Swarm Optimization based network structural balance

4.1. Algorithm framework

In this paper we propose a discrete PSO by minimizing an energy function to locate the unbalanced edges that exist in the network, and then change their signs to obtain a structurally balanced network. A flow chart of the proposed algorithm, denoted as PSOSB is shown in Fig. 3.
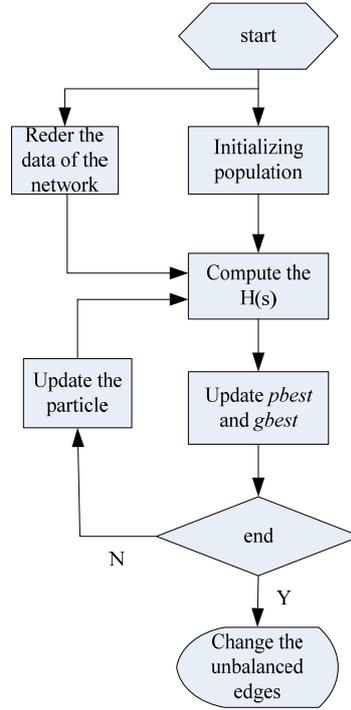
26

Fig. 3. A flow chart of the proposed PSOSB algorithm

4.2. Particle fitness function

The fitness function is used to evaluate the quality of the particle status, which determines the degree of structural balance of a social network. A previously proposed [14] energy function $H(s)$ used to measure the degree of balance of a network is defined as follows:

$$(3) \qquad\qquad H(s) = \sum_{i,j} (1 - a_{ij} s_i s_j)/2,$$

where $a_{ij} \in \{\pm 1\}$ is an element of the adjacency matrix of the network, and s_i and s_j characterize the relationships of node i and node j, respectively. If node i and node j are friends, then $s_i \cdot s_j = 1$, otherwise $s_i \cdot s_j = -1$.

When a network is balanced, $H(s) = 0$ and the network progressively draws closer to a balanced structure as the magnitude of $H(s)$ decreases. In this paper the network having the smallest $H(s)$ is identified using PSO, and the unbalanced edges are correspondingly changed to produce a balanced network structure.

4.3. Particle representation

Encoding and decoding serve as bridges in the optimization problems and optimization algorithms. Because the proposed algorithm seeks to minimize $H(s)$ for structural balance problems, we use encoding based on a string due to its simple encoding and convenient decoding operation. The encoding proposed in this paper is illustrated in Fig. 4.
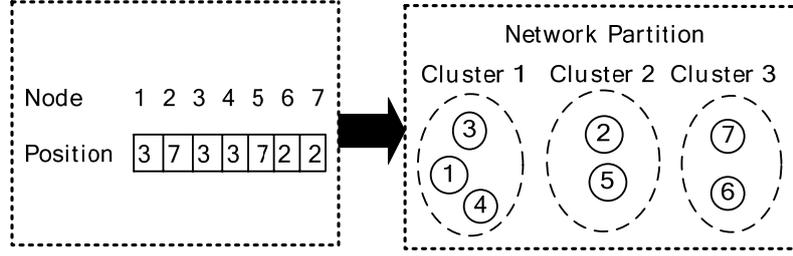
27

Fig. 4. A schematic diagram of the particles used in the proposed algorithm

As shown in Fig. 4, the position of the particle graph is a sequence of integers and each digit represents a classification corresponding to the position of the node. When decoded, the nodes with the same community number are divided into the same community. Thus, the coding automatically determines how the network is divided into several communities, without the need for manually pre-specified community numbers.

When decoded, if node i and node j belong to the same community, $s_i \cdot s_j = 1$, whereas if node I and node j belong to different communities, $s_i \cdot s_j = -1$.

4.4. Particle status update rules

From the representation of the particle, we know that the position of the particle is an integer. Therefore, the traditional updating rules, given by equations (1) and (2) are modified to apply PSO to the particles as follows:

$$(4) \qquad v_i^{t+1} = \zeta(\omega v_i^t + c_1 r_1 (p_i^t \oplus x_i^t) + c_2 r_2 (g^t \oplus x_i^t)),$$

$$(5) \qquad x_i^{t+1} = x_i^t \otimes v_i^{t+1},$$

Where ω is the inertial weight, and $\oplus$ is defined as the XOR operator.

The term $\zeta(x)$ is a bounding function that transforms its argument x into a discrete binary form for operation by equation (4) with its position vector. The function $\zeta(x)$ is defined by the following expression:

$$(6) \qquad \zeta(x) = \begin{cases} 1 & \text{if} \quad x \geq 1, \\ 0 & \text{if} \quad x < 1. \end{cases}$$

The operation $\otimes$ is the core of the particle updating procedure in Equation (5), which directly affects the convergence of the algorithm and the performance of mining communities.

For a network, the probability of two unconnected nodes belonging to a community is less than that for two connected nodes, so, based on this fact, we define the operation $\otimes$ by:

$$(7) \qquad x_i^{t+1} = x_i^t \otimes v_i^{t+1} = \begin{cases} \text{rand}(x_i^t) & \text{if} \quad v_i^{t+1} = 1, \\ x_i^t & \text{if} \quad v_i^{t+1} = 0. \end{cases}$$

Here, rand(x_t^i) represents the community label of the node, which is selected from the neighborhood of the node x_t^i.

28

5. Experimental tests and analyses

5.1. Experimental settings

We test the performance of the proposed algorithm in this section on both synthetic networks and real-world social networks. The parameter settings of the PSOSB algorithm are as follows: the inertial coefficient ω is set to the classical value of 0.729, the particle swarm learning factors c_1 and c_2 are both set to 1.494, and both the particle swarm population size pop and the maximum number of iterations g_{max} are set to 100. The performance of the PSOSB algorithm is compared with that of the previously proposed memetic computing based method [15], which is denoted as MemeSB. For this method the crossover probability and mutation probability are set to 0.8 and 0.2, respectively, and the population size and the maximum number of iterations for the algorithm are set to 100. Because both the PSOSB and MemeSB are iterative algorithms, 30 independent trials for each algorithm were conducted in the experiments.

The synthetic network data employed in our experiments are the AN1 and AN2 networks used previously [16]. The networks are comprised of 28 nodes, divided into three communities, and the network structures are balanced. The real-world network data used are the SPP [17] network, the GGS [18] network, the Yeast network [20], and the EGFR network [19]. The attributes of the four real-word networks are listed in Table 1.

Table 1. The properties of the real-world signed social networks employed in the tests

Networks	Nodes	Edges	Positive edges	Negative edges
SPP	10	45	18	27
GGS	16	58	29	29
EGFR	330	779	515	264
Yeast	690	1080	860	220

5.2. Simulated data tests

The experiments demonstrated that when testing the AN1 and AN2 networks, the PSOSB and MemeSB algorithms obtained equivalent results. Moreover, the results were both rather stable.
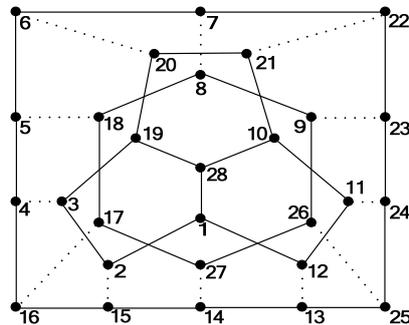

Fig. 5. Balanced structure of the ANI network obtained by the PSOSB and MemeSB algorithms, where the solid and dotted lines indicate positive and negative edges, respectively

The *H*(*s*) values obtained by the PSOSB and MemeSB algorithms after 30 independent trials were 0, which indicates that both algorithms obtained the balanced structures of AN1 and AN2 networks, as shown in Figs 5 and 6, respectively.

As it can be seen from Figs 5 and 6, the AN1 and AN2 networks are divided into three communities, and, within each community, the nodes are connected with positive edges, whereas, between communities, the nodes are connected with all negative edges. These results are in complete accordance with the definition of the network structural balance.

5.3. Real-world data tests

The results of the 30 independent trials of PSOSB and MemeSB algorithms on real-world networks are recorded in Table 2. For the SPP and GGS networks, both algorithms have obtained equivalent results because the scales of these two networks are small, and consequently, both algorithms readily obtain the optimal solutions of the problem. However, while both PSOSB and MemeSB obtained minimum *H*(*s*) values, the computational speed of PSOSB was faster than that of MemeSB.

Fig. 7 shows the actual community structure of the SPP network. The PSOSB obtains the minimum *H*(*s*) value of the network. The corresponding network topology is evaluated and we change all the imbalanced edges to obtain the balanced network topology. The balanced network topology is shown in Fig. 8. It can be seen from Figs 7 and 8, that in order to achieve the balanced structure of the network, the negative edges between the SNS node and the DS and ZS-ESS nodes have been changed to positive edges.
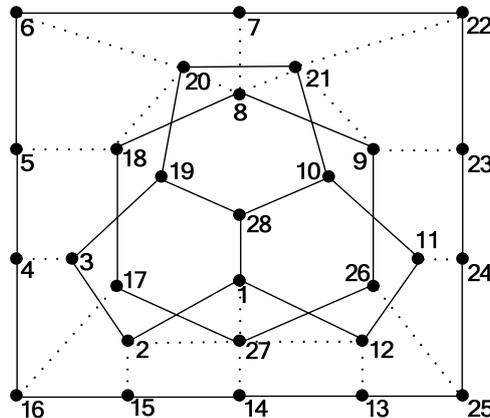

Fig. 6. Balanced structure of the AN2 network, obtained by PSOSB and MemeSB algorithms, where the solid and dotted lines indicate positive and negative edges, respectively

For the GGS network, it can be seen from Table 2 that PSOSB has obtained smaller values of *H*(*s*) than those of MemeSB, and the computational speed of PSOSB was still faster than of the MemeSB algorithm. Figs 9 and 10 respectively show the actual community structure of the GGS network and the balanced

structure obtained by changing the imbalanced edges. Comparison of the figures indicates that four positive edges of the GGS network have been changed to negative edges, as depicted by the black dashed lines in Fig. 10.
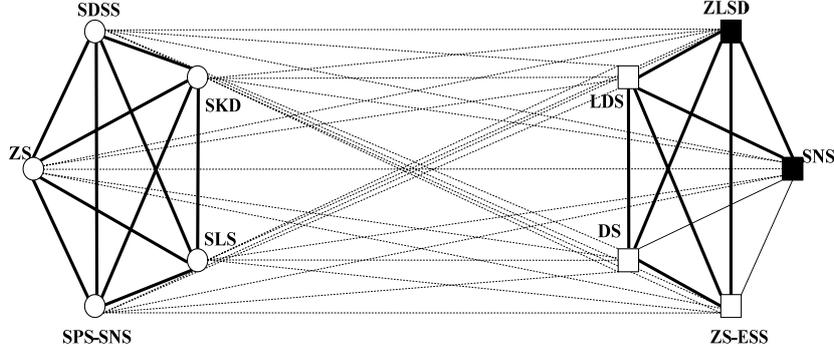

Fig. 7. Actual community structure of the SPP network, where the solid and dotted lines indicate positive and negative edges, respectively
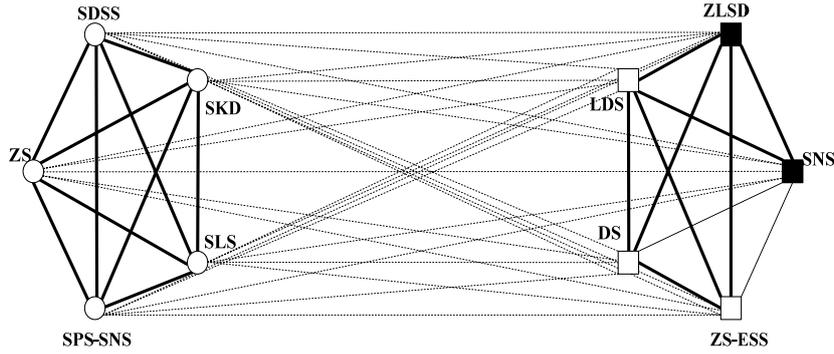

Fig. 8. Balanced structure of the SPP network, obtained by PSOSB and MemeSB algorithms, where the solid and dotted lines indicate positive and negative edges, respectively

Table 2. Statistical results for the real-world networks conducted in 30 independent trials

Index	$H(s)_{min}$		$H(s)_{avg}$		Communities		Time (s)	
Algorithm	PSOSB	MemeSB	PSOSB	MemeSB	PSOSB	MemeSB	PSOSB	MemeSB
SPP	2	2	2	2	2	2	3.3632×10^{-3}	4.4575×10^{-1}
GGS	4	4	4	4	4	4	3.7983×10^{-3}	5.2521×10^{-1}
EGFR	186	202	194.1333	232	73	86	7.3712	8.1589×10^{2}
Yeast	40	52	44.6	62	108	112	1.3988×10	1.0813×10^{4}

As indicated in Table 2 for the SPP and the GGS networks, we note that the PSOSB algorithm demonstrates significantly better computational performance than the MemeSB algorithm for the EGFR and Yeast networks as well. For these more complex networks, PSOSB not only obtains smaller $H(s)$ values than does MemeSB, but the computational time is significantly faster than that of MemeSB.
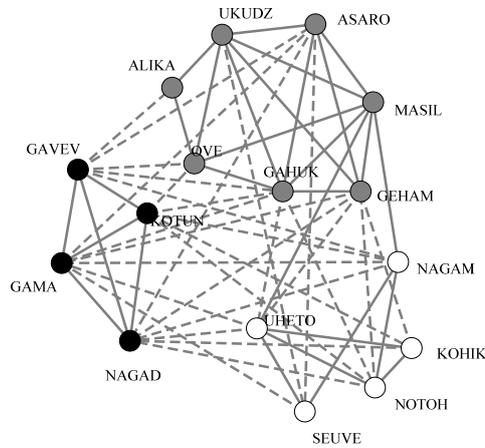
Fig. 9. Actual community structure of the GGS network, where the solid and dotted lines indicate positive and negative edges, respectively
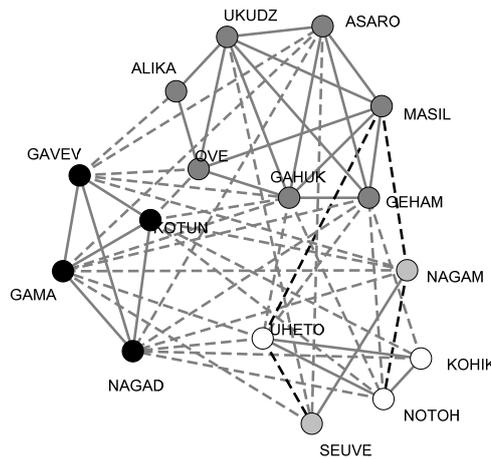

Fig. 10. Balanced structure of the GGS network after changing the imbalanced edges, given by the black dashed lines, where the solid and dashed lines indicate positive and negative edges, respectively

To highlight the relative convergence speeds of the PSOSB and MemeSB algorithms, Figs 11 and 12 respectively show the convergence curves of the two algorithms when operating upon the EGFR and Yeast networks for 100 iterations. The convergence results indicate that the PSOSB algorithm essentially convergences after about 40 iterations, while the MemeSB algorithm roughly requires 80 iterations to attain convergence.

In addition to employing PSO, where the globally optimal particle guides the population rapidly to approximate to the optimal solution of the problem, we have made full use of the network structure information to design our algorithm. The proposed particle status update rules can accelerate the particle search. Although the MemeSB algorithm used for comparison relies upon a local search strategy, the strategy is based on a greedy mechanism, which easily falls into a local optimum. Moreover, the greedy mechanism is computationally expensive.

32

Because PSOSB and MemeSB algorithms are random search algorithms, i.e, the results for each independent trial of the algorithm are different, it is necessary to discuss the stability of the algorithm. Figs 13(a) and (b) show boxplots of the results obtained by PSOSB and MemeSB on the EGFR and the Yeast networks, respectively, after 30 independent trials.
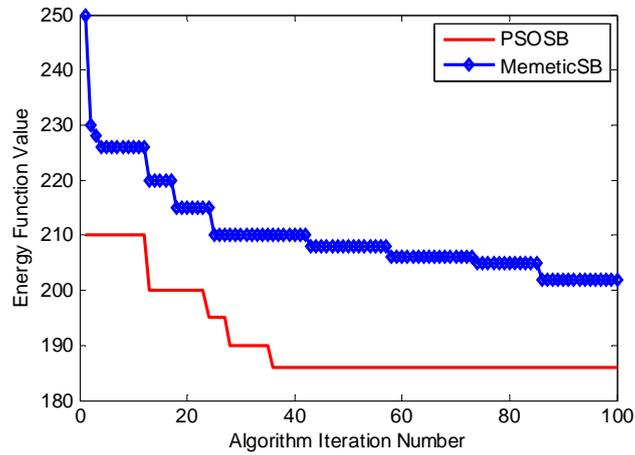

Fig. 11. Comparison between the convergence of PSOB and MemeSB algorithms for the EGFR network
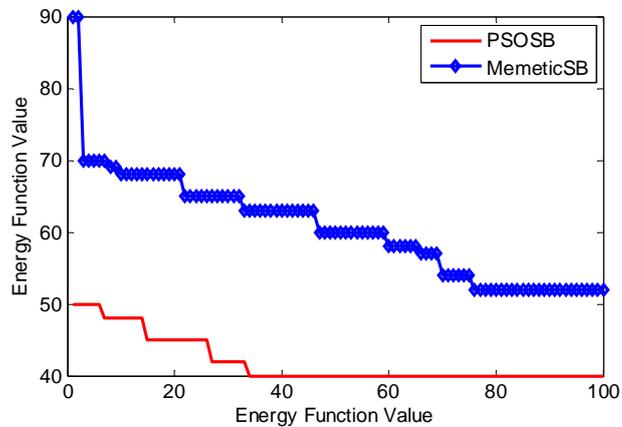

Fig. 12. Comparison between the convergence of PSOB and MemeSB algorithms for the Yeast network

A boxplot reflects the statistical distribution of the results obtained by an algorithm after many trials, and the length of the box reflects the stability of the algorithm. For a stable algorithm, the length of the box is relatively short. It can be observed from Fig. 13 that the length of the box obtained by PSOSB is shorter than that of MemeSB for the two networks considered, which suggests that PSOSB algorithm is relatively more stable than MemeSB algorithm. The experiments on real-world networks again verify the efficiency of the algorithm proposed in this paper.
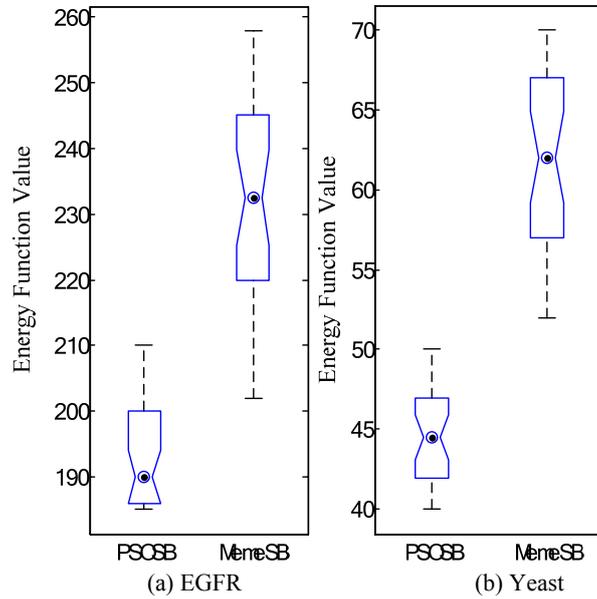
Fig. 13. Comparison between the stability of PSOSB and MemeSB algorithms on the EGFR (a) and the Yeast (b) networks, as demonstrated by boxplots

6. Conclusion

The study of the structural balance of social networks is of great importance. To achieve a balanced social network structure, the present study transformed the network structural balance problem into an optimization problem, and proposed an algorithm based on particle swarm optimization to estimate its solution. The proposed algorithm not only achieves a balanced structure for a network, but also discloses the hidden network community structure. The efficiency of the proposed algorithm was validated using synthetic and real-world network data sets. Subsequent work will focus on two aspects: on one hand, we will conduct research on designing efficient local learning strategies to enhance the global searching abilities of the algorithm; and on the other hand, we will seek to realize the parallelization of the algorithm so that be able to address large scale networks.

R e f e r e n c e s

1. W a s s e r m a n, S. Social Network Analysis: Methods and Applications. Cambridge University Press, 1994.
2. N e w m a n, M. Networks: An Introduction. Oxford University Press, 2010.
3. W a n g, X., X. L i, G. C h e n. Complex Network Theory and Its Application. Tsinghua University Press, 2006.

34

4. G i r v a n, M., M. E. J. N e w m a n. Community Structure in Social and Biological Networks. – Proceedings of the National Academy of Sciences, Vol. **99**, 2002, No 12, pp. 7821-7826.
5. F o r t u n a t o, S. Community Detection in Graphs. – Physics Reports, Vol. **486**, 2010, No 3, pp. 75-174.
6. H e i d e r, F. Social Perception and Phenomenal Causality. – Psychological Review, Vol. **51**, 1944, No 6, p. 358.
7. H a r a r y, F. On Local Balance and N-Balance in Signed Graphs. – The Michigan Mathematical Journal, Vol. **3**, 1955, No 1, pp. 37-41.
8. H a r a r y, F. A Matrix Criterion for Structural Balance. – Naval Research Logistics Quarterly, Vol. **7**, 1960, No 2, pp. 195-199.
9. Z h e n g, X., D. Z e n g, F. Y. W a n g. Social Balance in Signed Networks. – Information Systems Frontiers, 2014, pp. 1-19.
10. A n t a l, T., P. L. K r a p i v s k y, S. R e d n e r. Dynamics of Social Balance on Networks. – Physical Review E, Vol. **72**, 2005, No 3, pp. 36-121.
11. K e n n e d y, J. Particle Swarm Optimization. Encyclopedia of Machine Learning. Springer, US, 2010, pp. 760-766.
12. H u a n g, Y., H. W u, Y. X i a o et al. Logistics RFID Network Optimization Based on Robust PSO under Uncertain Conditions. – Control Theory & Applications, Vol. **31**, 2014, No 3, pp. 319-326.
13. W a n g, G., J. C h e n, T. C a i, P. L i. A Multi-Objective Decomposition-Based Stochastic Particle Swarm Optimization Algorithm and its Application to Optimal Design for Linear Motor. – Control Theory & Applications, Vol. **30**, 2013, No 6, pp. 693-701.
14. F a c c h e t t i, G., G. I a c o n o, C. A l t a f i n i. Computing Global Structural Balance in Large-Scale Signed Social Networks. – Proceedings of the National Academy of Sciences, Vol. **108**, 2011, No 52, pp. 20953-20958.
15. S u n, Y., H. D u, M. G o n g et al. Fast Computing Global Structural Balance in Signed Networks Based on Memetic Algorithm. – Physica A: Statistical Mechanics and Its Applications, Vol. **415**, 2014, pp. 261-272.
16. Y a n g, B., W. K. C h e u n g, J. L i u. Community Mining from Signed Social Networks. – IEEE Transactions on Knowledge and Data Engineering, Vol. **19**, 2007, No 10, pp. 1333-1348.
17. F e r l i g o j, A., A. K r a m b e r g e r. An Analysis of the Slovene Parliamentary Parties Network. – Developments in Statistics and Methodology, 1996, pp. 209-216.
18. R e a d, K. E. Cultures of the Central Highlands, New Guinea. – Southwestern Journal of Anthropology, 1954, pp. 1-43.
19. M i l o, R., S. S h e n-O r r, S. I t z k o v i t z et al. Network Motifs: Simple Building Blocks of Complex Networks. – Science, Vol. **298**, 2002, No 5594, pp. 824-827.
20. O d a, K., Y. M a t s u o k a, A. F u n a h a s h i et al. A Comprehensive Pathway Map of Epidermal Growth Factor Receptor Signaling. – Molecular Systems Biology, Vol. **1**, 2005, No 1, pp. 1-13.