Physica A 415 (2014) 261–272

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Fast computing global structural balance in signed networks based on memetic algorithm

Yixiang Sun [a], Haifeng Du [b], Maoguo Gong [a,*], Lijia Ma [a], Shanfeng Wang [a]

[a] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, Shaanxi Province 710071, China

[b] Center for Administration and Complexity Science, Xi'an Jiaotong University, Xi'an, Shaanxi Province 710049, China

HIGHLIGHTS

- The computation of global structural balance in signed networks is modeled as an optimization problem.
- We propose a memetic algorithm to optimize an evaluation function which is used to compute a distance to exact balance.
- Experiments show the excellent effectiveness and efficiency of the proposed method.

ARTICLE INFO

Article history:
Received 13 March 2014
Received in revised form 25 June 2014
Available online 12 August 2014

Keywords:
Signed network
Structural balance
Global property
Memetic algorithm
Genetic algorithm
Local search

ABSTRACT

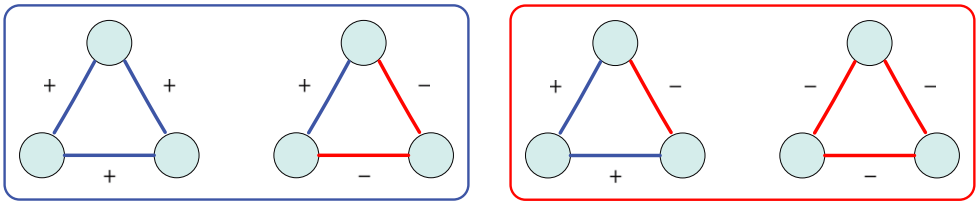
Structural balance is a large area of study in signed networks, and it is intrinsically a global property of the whole network. Computing global structural balance in signed networks, which has attracted some attention in recent years, is to measure how unbalanced a signed network is and it is a nondeterministic polynomial-time hard problem. Many approaches are developed to compute global balance. However, the results obtained by them are partial and unsatisfactory. In this study, the computation of global structural balance is solved as an optimization problem by using the Memetic Algorithm. The optimization algorithm, named Meme-SB, is proposed to optimize an evaluation function, energy function, which is used to compute a distance to exact balance. Our proposed algorithm combines Genetic Algorithm and a greedy strategy as the local search procedure. Experiments on social and biological networks show the excellent effectiveness and efficiency of the proposed method.

1. Introduction

Many real-world complex systems can be represented as networks, such as social network, biological network, collaboration network, the World-Wide-Web, and power grid. Networks are usually represented by graphs, where nodes (or vertices) represent the objects and edges represent the interactions among these objects [1]. In most cases, the edges typically indicate friendship, collaboration, sharing information, and all such things which have positive connotations. Social networks reflect a largely similar view that the connections mean friends, fans, followers, and so on. But in most networks, there are also negative effects at work. Some relations are friendly, but others are hostile; interactions between people or groups are regularly beset by controversy, disagreement, and sometimes outright conflict [2]. Social networks which have the characteristic mentioned above are called as signed social networks. For such complex networks, the connection between global properties and local interactions is a nice subject.

* Corresponding author. Tel.: +86 029 88202661; fax: +86 029 88201023.
E-mail addresses: gong@ieee.org, gongieee@gmail.com (M. Gong).

http://dx.doi.org/10.1016/j.physa.2014.07.071
0378-4371/



(a) Balanced triangles. (b) Unbalanced triangles.

Fig. 1. Simplest cases of balance and unbalance: signed triangles. (a) The two triangles which have an even number of negative edges are balanced, because they satisfy the adages that "the friend of my friend is my friend" and "the enemy of my friend is my enemy". (b) The two triangles which have an odd number of negative edges are considered unbalanced, because they do not comply with the logic of friendship.

A global property determining the structural balance of a signed network has received much attention in recent years [3–6]. In the dynamical evolution of networks, the signed networks implicitly seek out structural balance. Structural balance is an important notion in the analysis of the structure of signed networks, and the structural balance theory has been applied in many fields, such as international relationships and social psychology. For instance, in the study of international relationships [7], international politics represents a setting in which it is natural to assume that a collection of nodes all have opinions (positive or negative) about one another. Here the nodes are nations, and the opinions indicate alliances or animosity. Research in political science has shown that structural balance can provide an effective explanation for the behavior of nations during various international crises. In the field of social psychology [8], the on-line social media sites such as the technology news site Slashdot and the on-line product-rating site Epinions, allow users to register subjective evaluations of each other, and the balance theory is important for understanding the role the positive and negative sentiments play on social platforms. Structural balance theory was originally introduced by Heider [9] for understanding the origin of tensions and conflicts in a signed social network. Cartwright and Harary [10] modeled Heider's ideas by using signed graphs. In a signed graph, nodes represent the objects, and positive/negative edges represent the friendly/hostile relationships. According to the relevant studies, structural balance is a global property of the whole network. In many cases, in order to understand the structure of signed networks, we need to measure the unbalanced degree (i.e., measure a distance to exact balance). Computation of global structural balance offers an approach to solve this problem. More generally, computing global structural balance corresponds to computing the ground-state of a (nonplanar) Ising spin glass [11], which is a well-known nondeterministic polynomial-time hard (NP-hard) problem.

Before computing the structural balance, we shall find a function which is used to compute a distance to exact balance. A viable option is an energy function introduced by Facchetti et al. [12]. The smaller the energy function, the more balanced the network is. Computing global balance means assigning some signs (or values, whose meaning will be described in Section 3) to all the nodes so as to minimize the energy function. So, it is an optimization problem. Of course, using optimization algorithms to minimize this objective function is feasible. In this paper, in order to fast compute the global structural balance of the signed networks, we propose a novel approach which tries to optimize the energy function by employing Memetic Algorithm [13]. The proposed approach, named as Meme-SB, combines Genetic Algorithm and a greedy strategy as the local search procedure. Experiments on social and biological networks show the effectiveness and efficiency of the proposed method.

The rest of the paper is organized as follows: In the next section, a description of the related background including introduction of structural balance, computation of global balance and evolutionary algorithms is given. In Section 3, we describe the proposed Meme-SB in detail. Experimental studies are presented in Section 4. The concluding remarks are given in the last section.

2. Related background

2.1. Structural balance

According to the relevant studies [2,10], we know that the source of tensions comes from the cycles of the signed graph (i.e., the closed paths beginning and ending on the same node), especially the cycles which have an odd number of negative edges (hereafter called negative cycles). It follows the concept that balance is not related to the actual number of negative edges on the cycles but only to their parity. If a cycle has an odd number of negative edges, then this implies the cycle is unbalanced. Fig. 1 illustrates the simplest cases of balance and unbalance: signed triangles. In particular, a signed graph is exactly balanced if and only if it contains no cycle with an odd number of negative edges (the proof is described in detail in Refs. [2,10]). As such, structural balance is a global property of the whole network.

In addition, we have the Balance Theorem [10,14]: as shown in Fig. 2, if a signed graph is exactly balanced, the nodes can be divided into two sets (or communities), A and B. Every adjacent pair of nodes in A has a positive edge, every adjacent pair of nodes in B has a positive edge, and edges between A and B are negative. In turn, if a signed graph can be divided into two sets, and satisfies the Balance Theorem above simultaneously, then we can affirm that the graph is exactly balanced.

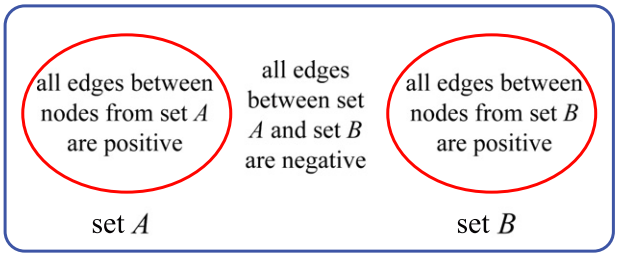


Fig. 2. Illustration of the division of a balanced signed graph.

2.2. Computation of global balance

In many cases, in order to understand the structure of signed networks, we need to measure how unbalanced a network is, and computing a distance to exact balance is a feasible way. The most plausible distance is given by the least number of edges which must be changed of signs in order for the graph to become exactly balanced [10,15,16]. More generally, computing this distance is equivalent to computing the ground-state of an Ising spin glass, which is a well-known NP-hard problem. The equivalence with energy minimization of a spin glass has, for example, been highlighted recently in Refs. [17–19]. As matter of fact, in related studies, in order to overcome this difficulty, a common approach has been to focus on the simplest cyclic motif potentially encoding frustration (so-called negative cycle), namely triangles of pairwise relations, comparing the frequency of positive and negative triangles [3–5]. Another alternative approach is introduced in Ref. [20], where spectral properties of the Laplacian are investigated. In their work, they use the Laplacians smallest eigenvalue to measure how unbalanced a signed network is (i.e., how much frustration is encoded in the cycles of the networks). Although both approaches can provide useful information to understand the structural balance of signed networks, this information is partial and unsatisfactory. The small motif analysis only identifies the frustration on the smallest possible groups of interacting objects, but overlooks more long-range conflicts associated to longer cycles. The spectral approach, on the contrary, gives an idea of the overall amount of frustration of the network, but does not provide any information on which relationships remain unbalanced [12].

By exploiting the analogy between signed networks and Ising spin glasses, Facchetti et al. have developed an efficient method for computing global structural balance [12,21]. The algorithm they use is inspired by the literature on ground-state search and has a heuristic character. It is based on the application of equivalence transformations which is called gauge transformations in the spin class literature [22]. These transformations aim to eliminate the so-called apparent disorder from the signed graph, and preserve the original frustration. In practice, these transformations lead to the reduction of the number of negative edges in the graph, which can simplify the calculation of global balance. Their Experimental studies have proved that it can get very low energies on large-scale signed social networks and its calculations are essentially exact. In their work, they minimize the energy function (which will be described in detail in Section 3) to compute the global structural balance. According to the minimum value of the energy function, we can get the information that the least number of unbalanced pairwise relationships between nodes. We believe that the energy function minimization is an optimization problem. So, using optimization algorithms to minimize the energy function is feasible.

2.3. Evolutionary algorithms

In the field of optimization, there are many algorithms that can be employed to solve optimization problems. Simulated annealing, tabu search, artificial neural networks and genetic algorithm are some examples. We all know that Evolutionary Algorithms (EAs) have become more and more popular, which are parallel in nature and do not require differentiability of objective functions and constraints, and also which deal with a set of possible solutions in a single run. As one method of EAs, genetic algorithms (GAs) have been widely used. In standard GAs, a population of strings (called chromosomes), which encode candidate solution (called individuals) to an optimization problem, evolves toward better solutions. The evolution usually starts from a random set of individuals. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population based on their fitness, and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. After several generations only solutions with large fitness survive.

Memetic Algorithms (MAs) are inspired by the concept of a meme originally introduced by Richard Dawkin, and they combine the characters of EAs with local search algorithms. Meme represents a unit of cultural evolution that can exhibit local refinement [23]. MAs are also named Hybrid Genetic Algorithms, Genetic Local Searchers, Lamarckian Genetic Algorithms, etc. From the view of optimization, MAs have been proved to be more efficient and more effective than traditional EAs for a few problem domains, especially for NP-hard problems. In the field of physics, the authors [24] developed a matching-based recombination algorithm to enhance the performance of MA for solving the Min Number Partition problem. In a study by Daolio et al. [25], the use of this type of methods can reveal the structure of low-lying minima in a disordered system. In our precious work [1,26–29], EAs have been successfully applied to handle community analysis problems. Recent works in

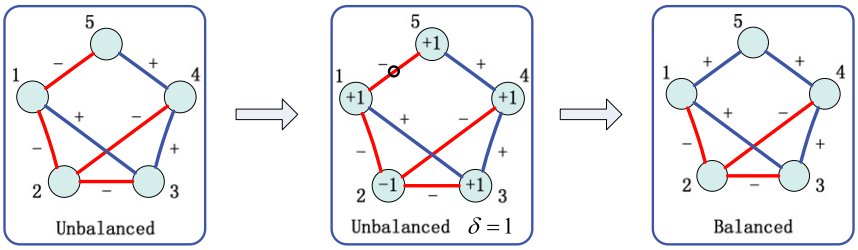


Fig. 3. Illustration of the meaning of the energy function's minimum value.

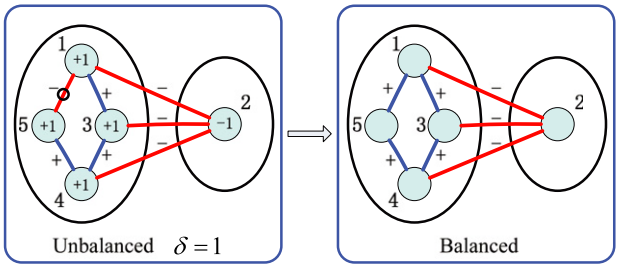


Fig. 4. Illustration of the division of a signed graph.

Refs. [1,29] point out that MAs have also been used for uncovering communities in social networks. For instance, in Ref. [1], we proposed a memetic algorithm, named as Meme-Net, to uncover communities at different hierarchical levels. Meme-Net shows its effectiveness. In Ref. [29], the authors proposed a fast memetic algorithm with multi-level learning strategies for community detection by optimizing modularity.

3. The proposed algorithm

In this section, we will describe our Meme-SB algorithm in detail. First, we describe the conception of energy function. Second, the main loop of the proposed algorithm is introduced. Then, the initialization procedure, genetic operation, and the local search strategy are described.

3.1. Objective function

Before introducing the proposed algorithm in detail, in this subsection, following the work by Facchetti et al. [12,30], we give a brief description to the energy function.

A signed network is a graph in which the nodes represent the objects and the edges represent the relationships among these objects. In this paper, the relationships are symmetric (i.e., the edges are undirected).

Consider $\mathbf{S} = [s_1 \cdots s_n]^T$ with $s_i \in \{\pm 1\} = \mathbb{B}_2$, $i = 1, \ldots, n$, where n is the number of nodes. Call $\mathcal{J}$ the $n \times n$ matrix of entries J_{ij} which represent the relationships between nodes s_i and s_j (friendship: $J_{ij} = +1$; hostility: $J_{ij} = -1$). As the signed graph is undirected, its adjacency matrix $\mathcal{J}$ is symmetric: $J_{ij} = J_{ji}$. Computing the global structural balance means assigning some signs (i.e., $+1$ and -1, which serve as the cluster identifiers) to all the nodes in such a way to minimize the energy function:

$$h(\mathbf{S}) = \sum_{(i,j)} (1 - J_{ij} s_i s_j)/2, \tag{1}$$

where the summation runs over all adjacent pairs of nodes. The minimum value of $h(\mathbf{S})$ corresponds to the least number of edges which lead to the unbalance of the network. If the minimum value is equal to 0, then we can affirm that the signed network is exactly balanced.

As shown in Fig. 3, we consider a signed graph which has 5 nodes and 7 edges. By optimizing the energy function and assigning a $+1$ or a -1 to all the 5 nodes, we obtain the minimum value of the function which equals 1. It indicates that there is just 1 edge (marked by circle) which leads to the unbalance of the network. If we change the sign of this edge into opposite, the network will be exactly balanced.

According to the signs of the nodes, we can divide them into two communities, which are shown in Fig. 4. From the classification result, we can find that there is 1 edge (marked by circle) which destroys the Balance Theorem: a balanced signed graph can be divided into two sets whose relations are all positive, while the relations between the two sets are all negative. Changing its sign and removing the "stain", we can make the graph balanced.

3.2. Main loop of the proposed algorithm

The framework of Meme-SB is given as Algorithm 1 and it requires some explanations. The GenerateInitialPopulation() function is used to create the initial population. The Selection() procedure is responsible for selecting parental population for mating in GA. Here, we use tournament selection. The GeneticOperation() function is used to perform crossover and mutation operation. The UpdatePopulation() procedure is used to reconstruct the current population. Here, the current population is constructed taken the best S_{pop} individuals from $\mathbf{P} \bigcup \mathbf{P}_{new}$. The TerminationCriterion() function is used to terminate the algorithm, which can be defined as setting a limit of the total number of iterations, reaching a maximum number of iterations without improvement, etc.

Algorithm 1 The algorithm framework of Meme-SB

Input: Maximum number of generations: G_{max}; Population size: S_{pop}; Size of mating pool: S_{pool}; Tournament size: S_{tour}; Crossover probability: P_c; Mutation probability: P_m.
$\mathbf{P} \leftarrow$ GenerateInitialPopulation(S_{pop});
repeat
  $\mathbf{P}_{parent} \leftarrow$ Selection($\mathbf{P}$, S_{pool}, S_{tour});
  $\mathbf{P}_{child} \leftarrow$ GeneticOperation($\mathbf{P}_{parent}$, P_c, P_m);
  $\mathbf{P}_{new} \leftarrow$ LocalSearch($\mathbf{P}_{child}$);
  $\mathbf{P} \leftarrow$ UpdatePopulation($\mathbf{P}$, $\mathbf{P}_{new}$);
until TerminationCriterion(G_{max})
Output: the fittest chromosome in $\mathbf{P}$.

3.3. Representation and initialization

Each chromosome in the population consists of genes, in which each gene corresponds to a node in the network. A chromosome is encoded as a string:

$$\mathbf{X} = \left\{x^1\ x^2\ \cdots\ x^n\right\},$$

where n is the number of nodes, and x^i is the sign of node s_i, which can be set to $+1$ or -1 in order to calculate the energy function. Fig. 5 displays an example of the representation.

Algorithm 2 The Population Initialization Procedure

Input: Population size: S_{pop}.
Generate a population $\mathbf{P}$ randomly, each chromosome $\mathbf{X}_k$ of which is a string consisting of $+1$ and -1, where $k = 1, 2, \ldots, S_{pop}$.
for each chromosome $\mathbf{X}_k$ **do**
  Generate a random sequence (i.e., $\{r_1\ r_2\ \cdots\ r_n\}$);
  $t_{counter} \leftarrow 1$;
  repeat
    select a node s_i, where $i = r_{t_{counter}}$;
    repeat
      randomly select a node s_j;
    until there is an edge J_{ij} between s_i and s_j
    if $J_{ij} = +1$ **then**
      $x_k^i \leftarrow x_k^j$;
    else
      $x_k^i \leftarrow -x_k^j$;
    end if
    $t_{counter} \leftarrow t_{counter} + 1$;
  until $t_{counter} = n$ (i.e., all the nodes are selected)
end for
Output: Population $\mathbf{P}$.

The population initialization procedure is given as Algorithm 2. Initially, each chromosome of the population is generated randomly. However, all of these solutions are of low quality. In GAs, it is common to initialize a high-quality population to speed up the convergence. Here, we employ a simple heuristic: our goal is to minimize the energy function, so, if the adjacent pair of nodes have a positive/negative edge, they should have the same/opposite signs. For each chromosome, we employ a simple greedy strategy: we select a node (or gene) randomly and reassign a value (or sign) to it to ensure that it makes no contribution to the energy function with another random node which is adjacent to it. We repeat this operation for n times, and at each time we randomly select a node which has never been selected before.

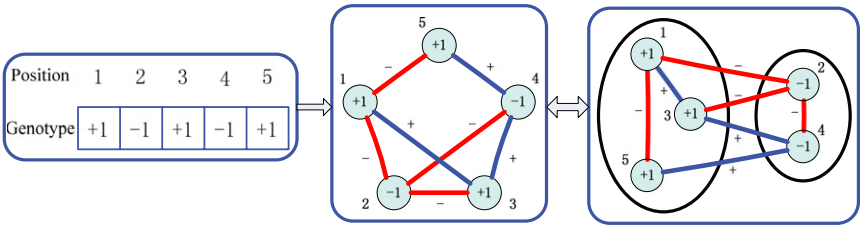


Fig. 5. Illustration of the representation. Left: one possible genotype. Middle: translate the genotype into the graph. Right: classification result.

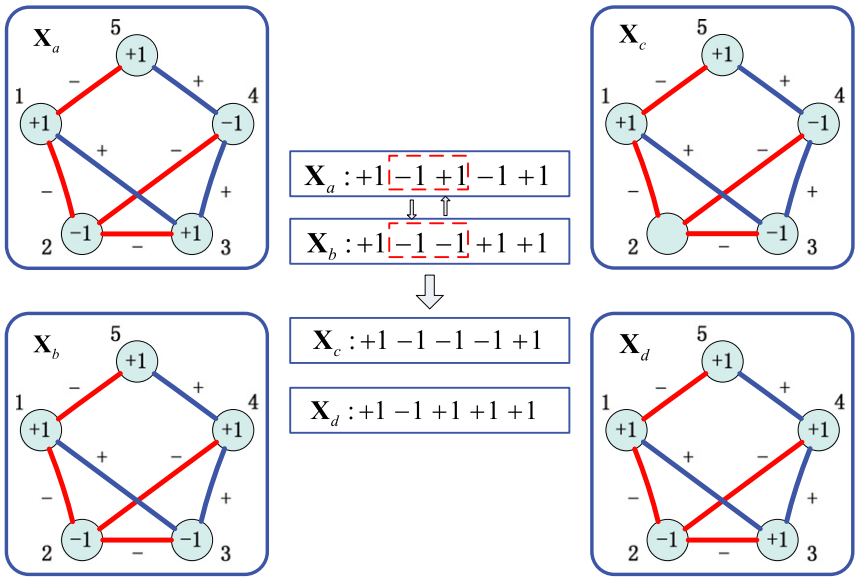


Fig. 6. Illustration of the two-point crossover.

3.4. Genetic operators

Crossover. Here, in order to improve the search capacity, we choose the two-point crossover [26] because this method is easy and effective. The crossing over procedure is defined as follows. Given two parents $\mathbf{X}_a$ and $\mathbf{X}_b$ at random, we first randomly select two points i and j (i.e., $1 \leq i \leq j \leq n$), and then everything between the two points of the parents is swapped (i.e., $x_a^k \leftrightarrow x_b^k$, $\forall k \in \{k \,|\, i \leq k \leq j\}$). This procedure returns two new chromosomes $\mathbf{X}_c$ and $\mathbf{X}_d$. An example of the operation of two-point crossover on the encoding step is shown in Fig. 6.

Mutation. In this process, we randomly pick a chromosome from $\mathbf{P}_{parent}$ to be mutated. Here we employ one-point mutation on this chromosome: a node is picked randomly on the chromosome, then we reassign a value to it to ensure that it makes no contribution to the energy function with another random node which is adjacent to it. This procedure follows the simple heuristic mentioned above. This operation is repeated n times on the chromosome. The specialized mutation operator has the abilities of strengthening the local search and keeping the diversity of the population.

3.5. The local search procedure

Local search is a means for algorithms to accelerate the convergence to an optimum and find global optimal solutions rather than local ones. The widely used local search means are hill climbing, simulated annealing, and tabu search. In this paper, a greedy strategy is devised to realize the local search procedure.

Before describing the local search procedure, we need to define the neighbors of a chromosome. Given a chromosome $\mathbf{X} = \{x^1\ x^2\ \cdots\ x^n\}$, we first randomly select a gene x^i ($i \in 1, \ldots, n$), and then change the value of the gene into the opposite (i.e., change it into -1 if $x^i = +1$ and $+1$ otherwise). The new chromosome $\mathbf{X}'$ after this change is called a neighbor of the chromosome $\mathbf{X}$. It needs to be noted that we can get a neighbor by changing several genes more than one.

The local search procedure we used in Meme-SB is a greedy strategy which is an iterative algorithm. This optimization technique starts with an arbitrary solution to a problem, then attempts to find a better solution by incrementally changing several elements of the solution. If the change can produce a better solution, an incremental change is made to the new

Table 1
Some parameters in the algorithm.

Parameter	Meaning	Value
G_{max}	The number of iterations	50
S_{pop}	Population size	500
S_{pool}	Size of the mating pool	150
S_{tour}	Tournament size	2
P_c	Crossover probability	0.9
P_m	Mutation probability	0.1

solution, repeating until no further improvement can be made. The detailed implementation is given as Algorithm 3. Here, we apply this technique to $\mathbf{P}_{child}$, which is the population after crossover and mutation. We only find the fittest chromosome in $\mathbf{P}_{child}$ and perform local search on it, until no improvement can be found. In Algorithm 3, the FindBest() function is used to evaluate the fitness of each chromosome in the input population, and return the chromosome having maximum fitness (or minimum energy), on which the local search procedure will be performed. The Energy() function is responsible for evaluating the energy of a solution. The FindBestNeighbor() function is used to find the best neighbor of a chromosome, which can be done easily according to the definition of the neighbors of a chromosome. Here, we need to explain the procedure of the FindBestNeighbor() function: given a chromosome, we select a gene randomly, and then change the value of the gene into the opposite. If this change is an improvement, we make this change to the new chromosome; if this change cannot produce a better solution, we repeal this change. We repeat this operation n times, and each time we randomly select a gene which has never been selected before.

Algorithm 3 The Local Search Procedure

1: **Input**: $\mathbf{P}_{child}$.
2: $\mathbf{N}_{current} \leftarrow$ FindBest($\mathbf{P}_{child}$);
3: *islocal* $\leftarrow$ FALSE;
4: **repeat**
5: $\quad\mathbf{N}_{next} \leftarrow$ FindBestNeighbor($\mathbf{N}_{current}$);
6: $\quad$**if** Energy($\mathbf{N}_{next}$)<Energy($\mathbf{N}_{current}$) **then**
7: $\quad\quad\mathbf{N}_{current} \leftarrow \mathbf{N}_{next}$;
8: $\quad$**else**
9: $\quad\quad$*islocal* $\leftarrow$ TRUE;
10: $\quad$**end if**
11: **until** *islocal* is TRUE
12: **Output**: $\mathbf{P}_{child}$.

4. Experimental results

In this section, we will test Meme-SB on 4 social networks and 4 biological networks. The compared algorithm is a GA version of this algorithm, termed as GA-SB, which is simply developed by removing the LocalSearch() function in Algorithm 1. The results obtained by the algorithm (described in detail in Refs. [12,21], and termed as HRT-SB here) introduced by Facchetti et al. are also given for comparison. The experiments on these networks show the effectiveness and efficiency of our proposed algorithm on computation of global structural balance.

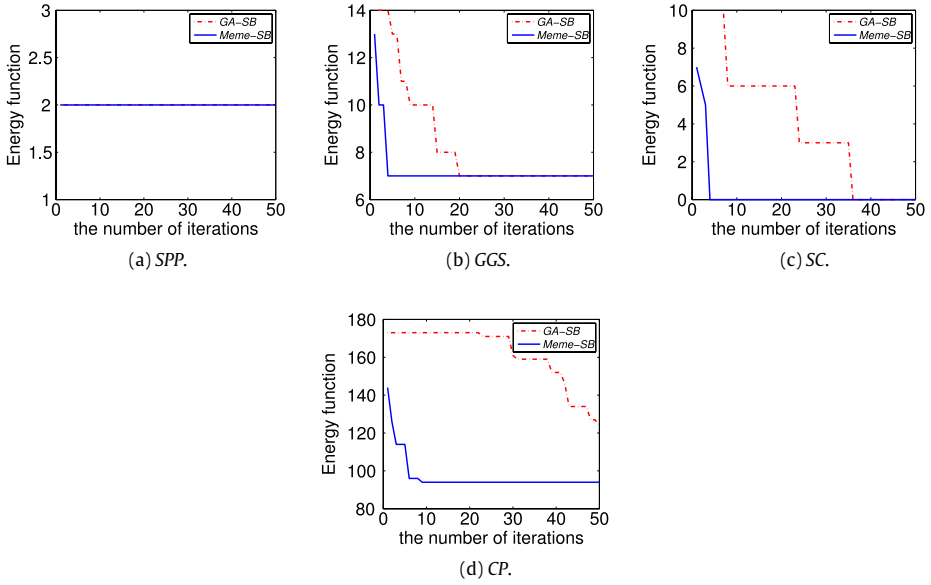
Because the algorithm we proposed is not parameter-free, we must set some values to them in advance such as population size, number of generation, crossover probability, and mutation probability. Some of these parameters are set to values we found by trail and error in order to ensure that the proposed algorithm has excellent performance. As shown in Table 1, some parameters we have mentioned and their value we used in the experiments are given. We conduct all the experiments on a personal computer with Intel Core2 Duo 2.33 GHz CPU and 2.00 GB Memory, running on Windows XP. In the following experiments, the reported data are the statistical results based on 30 independent runs on each dataset.

4.1. Social network

In this subsection, we test our method on 4 small-scale social networks, the *Slovene Parliamentary Party* network (*SPP*), the *Gahuku-Gama Subtribes* network (*GGS*), the *Social Circle* network (*SC*), and the *Cooperation Performance* network (*CP*). For each social network, the exactly minimum value of the energy function is unknown beforehand. Numbers of nodes, edges, positive edges and negative edges are provided in Table 2. We use these social networks to test if Meme-SB can effectively compute global structural balance. Meanwhile, a comparison between the results obtained by our algorithm and those obtained by GA-SB is made to demonstrate that the proposed algorithm has better performances and higher efficiency. Because of the small scale of these 4 social networks, our algorithm Meme-SB and both the two compared algorithms,

Table 2
Social networks used in this study. n and m are the number of nodes and edges of the signed networks, m^+ and m^- are the number of positive and negative edges of the networks.

Network	n	m	m^+	m^-
SPP	10	45	18	27
GGS	16	58	29	29
SC	28	42	30	12
CP	119	686	528	158



(a) *SPP*. (b) *GGS*. (c) *SC*.

(d) *CP*.

Fig. 7. Comparison between Meme-SB and GA-SB on social networks.

GA-SB and HRT-SB, can obtain the same minimum value in each independent run. So in this subsection, we do not analyze the average value and standard deviation of the global structural balance.

Slovene Parliamentary Party network (*SPP*): The relation network of 10 Slovene Parliamentary parties was set up by a series of experts on Parliamentary activities in 1994 [31]. SKD, ZLSD, SDSS, LDS, ZS-ESS, ZS, DS, SLS, SPS-SNS, and SNS are the names of the 10 parties respectively in short.

Gahuku-Gama Subtribes Network (*GGS*): This network, established on the basis of Reads study about the cultures of highland New Guinea [32], describes the political alliances and oppositions among 16 Gahuku-Gama subtribes distributed in a particular area and were involved in warfare with one another in 1954.

Social Circle network (*SC*): This network is a simple signed social network [33] consisting of 28 nodes and 42 edges. Positive edges mean that two people are good friends, while negative edges mean two people are enemies.

Cooperation Performance network (*CP*): This network is a survey result which comes from a research report called "Organization structure and team performance". The report finished by members from Xi'an Jiaotong University in 2009, describes the relationships between staff of a company in Xi'an. The survey contraposes the company's 119 employees. In the network, each node represents an employee. The positive relationship between two employees means that their cooperation can improve the efficiency of work, while the negative relationship means their cooperation is an obstacle.

For all the networks, Fig. 7 displays the values of energy function obtained by Meme-SB and GA-SB in one run, when the number of generations increases from 1 to 50. In this experiment, we test the GA version algorithm on the social networks with the same parameters as in Meme-SB. As shown in Fig. 7, for each dataset, our Meme-SB algorithm can find the minimum value of the energy function within 10 generations. In particular, for *SPP*, Meme-SB finds a minimum energy function value of 2 in just 1 generation. Meanwhile, for *GGS* and *SC*, Meme-SB only takes less than 5 generations to find the minimum value. In contrast, GA-SB needs more generations. From this comparison, we can find that the local search procedure plays a very important role in Meme-SB. Without the local search procedure, it becomes harder to find the exactly optimum value. Moreover, the local search procedure also speeds up the convergence of Meme-SB. It is obviously shown in Fig. 7(d), for *CP*, Meme-SB finds the optimum value of 94 in less than 10 generations. However, without the local search procedure, the GA-SB algorithm does not find the optimum value within 50 generations.

According to the experimental results, we give the nodes classifications of networks *SPP* and *SC*, which are shown in Figs. 8 and 9. For *SPP*, Meme-SB obtains the exactly minimum value of 2 which means there are 2 edges that lead to the unbalance of the network. As shown in Fig. 8, the 2 edges are marked by circles and if we change their signs into opposite, the network

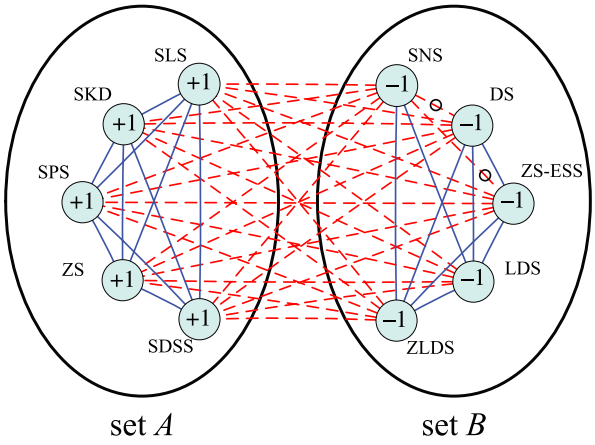


Fig. 8. Nodes classification result of *SPP*. Solid lines represent positive relationships, while dotted lines represent negative relationships.

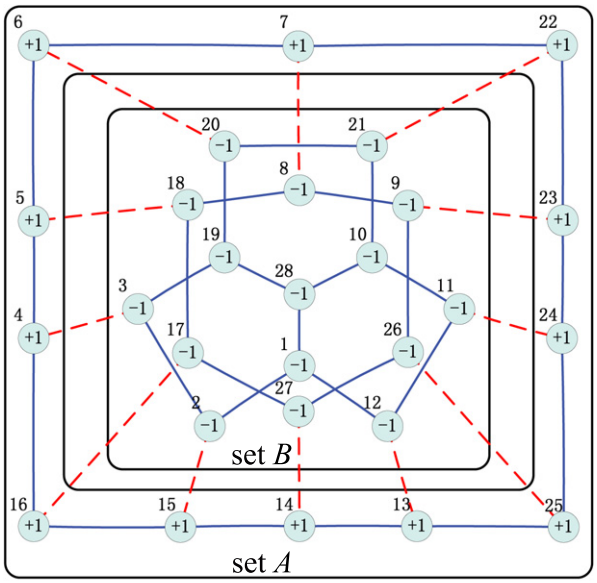


Fig. 9. Nodes classification result of *SC*.

will become exactly balanced. For *SC*, our algorithm can get the exactly minimum value which equals 0. It indicates that the network *SC* itself is exactly balanced. As shown in Fig. 9, the classification result of *SC* does not have any "stain".

This experiment shows Meme-SB is a high valid method for computing the structural balance on the small-scale social networks.

4.2. Biological networks

In this subsection, we present experimental studies on 4 biological networks which have been used in Ref. [21], the *EGFR* network, the *macrophage* network, the *yeast* network, and the *E.coli* network. All the networks are symmetric by being previously processed. Here, for each biological network, the energy function's exactly minimum value is unknown. See Table 3 for details. The compared algorithm include GA-SB and HRT-SB. Considering the complexity and the large scale of these biological networks, we increase the number of iterations to 500.

EGFR is the name of "Epidermal growth factor receptor pathway" in short, which is a network consisting of 330 nodes and 779 edges [34].

Macrophage network is a macrophage's molecular interaction map [35], which has 697 nodes and 1425 edges.

Yeast network is the gene regulatory network of *S.cerevisiae* [36], which contains 690 nodes and 1080 edges.

E.coli network is the gene regulatory network of *E.coli* [37]. Our version has 1476 nodes and 3215 edges.

For each network, Fig. 10 displays the values of energy function obtained by Meme-SB and GA-SB in one run, when the number of generations increases from 1 to 500. It clearly shows that, with the local search procedure, the results obtained

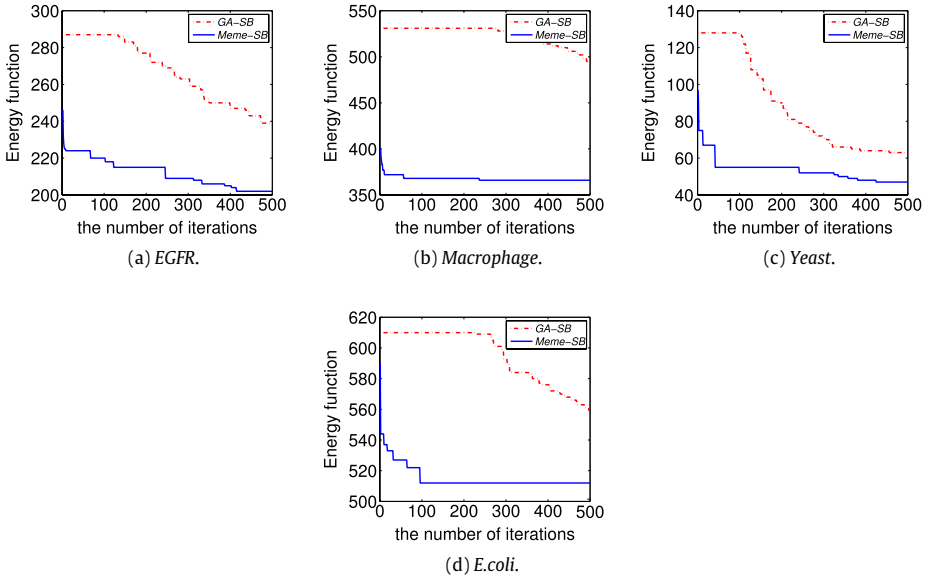


Fig. 10. Comparison between Meme-SB and GA-SB on biological networks.

Table 3
Biological networks used in this study.

Network	n	m	m^+	m^-
EGFR	330	779	515	264
macrophage	697	1425	947	478
yeast	690	1080	860	220
E.coli	1476	3215	1879	1336

Table 4
The results of 30 runs of Meme-SB, HRT-SB, and GA-SB on four biological networks.

Network	$\delta_{avg}/\delta_{std}$		
	Meme-SB	HRT-SB	GA-SB
EGFR	202.50/1.08	219.30/0.48	240.50/4.14
macrophage	364.50/1.08	374.30/2.79	502.00/12.16
yeast	48.20/1.93	53.20/2.62	62.00/3.65
E.coli	511.10/1.37	517.90/3.14	561.30/6.31

by Meme-SB are a lot better than GA-SB. Although the results obtained by Meme-SB may not be the best, at least they are satisfactory. Moreover, as shown in Fig. 10(b) and (d), the local search procedure can speed up the convergence of Meme-SB obviously. This comparison clearly illustrates the point that, with the help of the local search procedure, Meme-SB performs better than GA-SB.

In the following, for each network, we run our algorithm and the two compared algorithms 30 times, and compute the average value and standard deviation of the energy function (δ_{avg} and δ_{std}) over the 30 runs. The results are reported in Table 4. Here, it is important to note that considering the time complexity, we just run the processes Algorithms 1 and 2 of HRT-SB. Because we have known that Meme-SB has better performances than GA-SB, here, we just analyze the experimental results obtained by Meme-SB and HRT-SB. As shown in Table 4, on *EGFR* network, *macrophage* network, *yeast* network, and *E.coli* network, the average values of the energy function found by Meme-SB are 202.50, 364.50, 48.20, and 511.10, respectively, while the average values found by HRT-SB are 219.30, 374.30, 53.20 and 517.90, respectively. We can see that, on all these 4 networks the results obtained by Meme-SB are better than those by HRT-SB. In the aspect of stability, on *EGFR* network, the standard deviation obtained by Meme-SB is 1.08, while that obtained by HRT-SB is 0.48. It indicates that the stability of HRT-SB is slightly better than Meme-SB on *EGFR* network. However, on the other 3 networks, the standard deviations obtained by Meme-SB are all smaller than those by HRT-SB. This comparison clearly shows that the results obtained by the algorithm Meme-SB are not only more excellent, but also more steady than the two compared algorithms.

4.3. Complexity analysis of Meme-SB

In this part, the time complexity of the proposed algorithm Meme-SB is analyzed. Here we use n and m to denote the node and edge numbers of the network, respectively. At each generation, firstly, we need to perform the crossover operator $\lfloor S_{pool}/2 \rfloor$ times and the mutation operator S_{pool} times at most, where S_{pool} is the size of the mating pool. The time complexity of the calculation of energy function is $O(m)$. Therefore, the time complexity of the genetic operator is $O(S_{pool}(n+m))$. Secondly, we need to perform the local search procedure which requires $\log n$ steps to reach a local optimum in the search space. Moreover, at each step of the local search procedure, it needs to consider $\log n$ neighbors of each node. Therefore, the time complexity of the local search procedure is $O(n(\log n)^2)$. Finally, we need to reconstruct the population, which needs $O(S_{pop}+S_{pool})$ basic operations, where S_{pop} is the size of the population. In practical applications, we have the following relations: $O(S_{pop}+S_{pool}) \leq O(S_{pool}(n+m)) \leq O(n(\log n)^2)$. Therefore, the time complexity of the proposed Meme-SB algorithm at each generation is $O(n(\log n)^2)$.

Through the similar complexity analysis with Meme-SB, we can obtain that the computation complexities of GA-SB and HRT-SB are $O(S_{pool}(n+m))$ and $O(n+m)$, respectively. By comparing Meme-SB with GA-SB and HRT-SB, it can be seen that the time complexity of Meme-SB is higher than those of the two compared algorithms. And in the aspect of computation time, Meme-SB needs slightly more time to perform its procedure. The reason Meme-SB is defective in computation time is that it performs the local search procedure. However, with the help of the local search procedure, Meme-SB has better performances on computing global structural balance than GA-SB and HRT-SB. Its perfect performances make it possible for exactly computing global structural balance in a reasonable time.

5. Concluding remarks

In this paper, a novel algorithm Meme-SB is proposed to optimize the energy function for computing global structural balance in signed networks. The proposed algorithm combines GAs and a greedy strategy as the local search procedure. The experimental results show that Meme-SB has better performances than the two compared algorithms. The comparative results on both social and biological networks demonstrate the effectiveness and efficiency of Meme-SB in computing global structural balance. In addition, according to the minimum value of the energy function, we can get the connotative information that how many edges (or pairwise relationships) which lead to the unbalance of the network exist. By changing the signs of these edges into the opposite, we can obtain an exactly balanced network.

Acknowledgments

The authors wish to thank the editors and anonymous reviewers for their valuable comments and helpful suggestions which greatly improved the paper's quality. This work was supported by the National Natural Science Foundation of China (Grant Nos. 61273317, 61422209 and 61473215), Key Project from the National Science Foundation of China (Grant No. 12AZD110), National Top Youth Talents Program of China, the Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20130203110011) and the Fundamental Research Fund for the Central Universities.

References

[1] M. Gong, B. Fu, L. Jiao, H. Du, A memetic algorithm for community detection in networks, Phys. Rev. E 84 (5) (2011) 056101.
[2] D. Easley, J. Kleinberg, Networks, crowds, and markets: Reasoning about a highly connected world (2012).
[3] M. Szell, R. Lambiotte, S. Thurner, Multirelational organization of large-scale social networks in an online world, Proc. Natl. Acad. Sci. 107 (31) (2010) 13636–13641.
[4] J. Kunegis, A. Lommatzsch, C. Bauckhage, The slashdot zoo: mining a social network with negative edges, in: Proceedings of the 18th International Conference on World Wide Web, 2009, pp. 741–750.
[5] J. Leskovec, D. Huttenlocher, J. Kleinberg, Signed networks in social media, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2010, pp. 1361–1370.
[6] A. Srinivasan, Local balancing influences global structure in social networks, Proc. Natl. Acad. Sci. 108 (5) (2011) 1751–1752.
[7] M. Moore, An international application of Heider's balance theory, Eur. J. Soc. Psychol. 8 (3) (1978) 401–405.
[8] R. Guha, R. Kumar, P. Raghavan, A. Tomkins, Propagation of trust and distrust, in: Proceedings of the 13th International Conference on World Wide Web, 2004, pp. 403–412.
[9] F. Heider, Attitudes and cognitive organization, J. Psychol. 21 (1) (1946) 107–112.
[10] D. Cartwright, F. Harary, Structural balance: a generalization of heider's theory, Psychol. Rev. 63 (5) (1956) 277–292.
[11] F. Barahona, On the computational complexity of Ising spin glass models, J. Phys. A: Math. Gen. 15 (10) (1982) 3241.
[12] G. Facchetti, G. Iacono, C. Altafini, Computing global structural balance in large-scale signed social networks, Proc. Natl. Acad. Sci. 108 (52) (2011) 20953–20958.
[13] P. Moscato, On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms, Caltech Concurrent Computation Program, C3P Report 826 (1989) 1989.
[14] F. Harary, On the notion of balance of a signed graph, Michigan Math. J. 2 (2) (1953) 143–146.
[15] F. Harary, On the measurement of structural balance, Behav. Sci. 4 (4) (1959) 316–323.
[16] F. Harary, A matrix criterion for structural balance, Nav. Res. Logist. Q. 7 (2) (1960) 195–199.
[17] S. Galam, Fragmentation versus stability in bimodal coalitions, Physica A 230 (1) (1996) 174–188.
[18] T. Antal, P. Krapivsky, S. Redner, Dynamics of social balance on networks, Phys. Rev. E 72 (3) (2005) 036121.
[19] S.A. Marvel, S.H. Strogatz, J.M. Kleinberg, Energy landscape of social balance, Phys. Rev. Lett. 103 (19) (2009) 198701.
[20] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E.W. De Luca, S. Albayrak, Spectral analysis of signed graphs for clustering, prediction and visualization, in: Proceedings of the Siam Conference on Data Mining, vol. 10, 2010, pp. 559–570.

[21] G. Iacono, F. Ramezani, N. Soranzo, C. Altafini, Determining the distance to monotonicity of a biological network: a graph-theoretical approach, IET Syst. Biol. 4 (3) (2010) 223–235.
[22] G. Toulouse, Theory of the frustration effect in spin glasses: I, Commun. Phys. 2 (4) (1977) 115–119.
[23] R. Dawkins, The Selfish Gene, Oxford University Press, 2006.
[24] R. Berretta, C. Cotta, P. Moscato, Enhancing the performance of memetic algorithms by using a matching-based recombination algorithm, in: Metaheuristics: Computer Decision-making, Kluwer Academic Publishers, 2004, pp. 65–90.
[25] F. Daolio, M. Tomassini, S. Vérel, G. Ochoa, Communities of minima in local optima networks of combinatorial spaces, Physica A 390 (9) (2011) 1684–1694.
[26] M. Gong, L. Ma, Q. Zhang, L. Jiao, Community detection in networks by using multiobjective evolutionary algorithm with decomposition, Physica A 391 (15) (2012) 4050–4060.
[27] L. Ma, M. Gong, Q. Cai, L. Jiao, Enhancing community integrity of networks against multilevel targeted attacks, Phys. Rev. E 88 (2) (2013) 022810.
[28] M. Gong, Q. Cai, X. Chen, L. Ma, Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition, IEEE Trans. Evolut. Comput. 18 (1) (2014) 82–97.
[29] L. Ma, M. Gong, J. Liu, Q. Cai, L. Jiao, Multi-level learning based memetic algorithm for community detection, Appl. Soft Comput. 19 (2014) 121–133.
[30] G. Facchetti, G. Iacono, C. Altafini, Exploring the low-energy landscape of large-scale signed social networks, Phys. Rev. E 86 (3) (2012) 036116.
[31] S. Kropivnik, A. Mrvar, An analysis of the slovene parliamentary parties network, in: A. Ferligoj, A. Kramberger (Eds.), Developments in Statistics and Methodology, 1996, pp. 209–216.
[32] K.E. Read, Cultures of the central highlands, new guinea, Southwestern J. Anthropol. (1954) 1–43.
[33] B. Yang, W.K. Cheung, J. Liu, Community mining from signed social networks, IEEE Trans. Knowl. Data Eng. 19 (10) (2007) 1333–1348.
[34] K. Oda, Y. Matsuoka, A. Funahashi, H. Kitano, A comprehensive pathway map of epidermal growth factor receptor signaling, Mol. Syst. Biol. 1 (1) (2005) 2005.
[35] K. Oda, T. Kimura, Y. Matsuoka, A. Funahashi, M. Muramatsu, H. Kitano, Molecular interaction map of a macrophage, AfCS Res. Rep. 2 (14) (2004) 1–12.
[36] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (5594) (2002) 824–827.
[37] H. Salgado, S. Gama-Castro, M. Peralta-Gil, E. Díaz-Peredo, F. Sánchez-Solano, A. Santos-Zavaleta, I. Martínez-Flores, V. Jiménez-Jacinto, C. Bonavides-Martínez, J. Segura-Salazar, et al., Regulondb (version 5.0): Escherichia coli k-12 transcriptional regulatory network, operon organization, and growth conditions, Nucleic Acids Res. 34 (suppl 1) (2006) D394–D397.