A survey on network community detection based on evolutionary computation

Qing Cai, Lijia Ma and Maoguo Gong*

Key Laboratory of Intelligent Perception and
Image Understanding of Ministry of Education,
International Research Center
for Intelligent Perception and Computation,
Xidian University,
Xi'an, Shaanxi Province 710071, China
Fax: +86-029-88201023
Email: OMEGACaiQ@gmail.com
Email: OMEGAMaLJ@gmail.com
Email: gong@ieee.org
*Corresponding author

Dayong Tian

Center for Quantum Computation and Intelligent Systems,
University of Technology, Sydney,
Broadway, NSW 2007, Austrilia
Email: dayongt@gmail.com

Abstract: Uncovering community structures of a complex network can help us to understand how the network functions. Over the past few decades, network community detection has attracted growing research interest from many fields. Many community detection methods have been developed. Network community structure detection can be modelled as optimisation problems. Due to their inherent complexity, these problems often cannot be well solved by traditional optimisation methods. For this reason, evolutionary algorithms have been adopted as a major tool for dealing with community detection problems. This paper presents a survey on evolutionary algorithms for network community detection. The evolutionary algorithms in this survey cover both single objective and multiobjective optimisations. The network models involve weighted/unweighted, signed/unsigned, overlapping/non-overlapping and static/dynamic ones.

Keywords: complex network; community structure; community detection; evolutionary computation; multiobjective optimisation.

Reference to this paper should be made as follows: Cai, Q., Ma, L., Gong, M. and Tian, D. (xxxx) 'A survey on network community detection based on evolutionary computation', *Int. J. Bio-Inspired Computation*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Qing Cai received his BS in Electronic Information Engineering from Wuhan Textile University, Wuhan, China in 2010. Currently, he is working towards his PhD in Pattern Recognition and Intelligent Systems from the School of Electronic Engineering, Xidian University, Xi'an, China. His current research interests are in the area of computational intelligence and complex network analysis.

Lijia Ma received his BS in Communication Engineering from Hunan Normal University, Hunan, China in 2010. He is currently working toward his PhD in Pattern Recognition and Intelligent Systems at the School of Electronic Engineering, Xidian University, Xi'an, China. His research interests include evolutionary multiobjective optimisation, data mining and complex network analysis.

Maoguo Gong received his BS in Electronic Engineering (with first class honours) and PhD in Electronic Science and Technology from Xidian University, Xi'an, China, in 2003 and 2009, respectively. Since 2006, he has been a Teacher with Xidian University. In 2008 and 2010, he was promoted as an Associate Professor and as a Full Professor, respectively, both with exceptive admission. He is currently a Full Professor with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University. His research interests include computational intelligence with applications. He is a senior

member of the IEEE, an executive committee member of the Chinese Association for Artificial Intelligence, and a senior member of the Chinese Computer Federation. He was the recipient of the prestigious National Top Young Talent of China (selected by the Central Organization Department of China), the excellent young scientist foundation (selected by the National Natural Science Foundation of China), the New Century Excellent Talent in University (selected by the Ministry of Education of China), the Young Teacher Award by the Fok Ying Tung Education Foundation, the Young Scientist Award of Shaanxi Province, the New Scientific and Technological Star of Shaanxi Province, and the National Natural Science Award of China.

Dayong Tian received his BS of Electronic Information Science and Techonology and ME of Circuit and System in Xidian University, Xi'an, China. Currently, he is a PhD student in Biomedical Engineering in University of Technology, Sydney, Australia. His research interests are machine learning theory and applications in face image restoration, recognition and image retrieval.

1 Introduction

With the development of science and technology, diverse kinds of networks such as the the computer network, the power system network, the social media network and the biological network, have emerged and are changing our daily life (Newman, 2010). Network becomes the engine of scientific research activities in the 21st century, and the active topic across disciplines. Many real-world networks are rather complex, characterised by big data volume, dynamics, interactivity and heterogeneity. Network analysis provides an effective tool to probe networks' potential properties such as the small-world property (Watts and Strogatz, 1998), the scale-free property (Barabási and Albert, 1999), and improve our understanding of the complex world. In recent years, another outstanding property of networks, i.e., the community structure (Girvan and Newman, 2002), has become a hot research topic. Community detection is one of the theoretical underpinnings of network science, social science, physical science, biological science, and so forth. Several surveys on community detection in networks can be found in Orman and Labatut (2009), Fortunato (2010), Kim et al. (2011) and Xie et al. (2013).

Community detection sheds light on the functionalities of complex networks. Thus far, a large number of community detection methods have been proposed in the literature. However, it is very hard to give a definite classification for the existing avenues. Amongst the existing methods, one category is based on the evolutionary algorithms (EAs). EAs are a class of artificial intelligent optimisation metaheuristics inspired by principles from biology, ethology, etc. These metaheuristics are notable for their good local learning and global searching abilities and have been developed for successfully solving a wide range of optimisation problems. The EAs based community mining methods can automatically determine the clusters of the complex networks which makes it very convenient for practical applications. The essence of the EAs based community detection methods is to first model the network community detection task as different optimisation problems and then design suitable metaheuristics to deal with them. One of the landmark contributions is the work done by Newman (2004). Strictly speaking, the method used in that work is not a metaheuristic, but the stopping criterion, i.e., modularity, first put forward in his early work in Newman and Girvan (2004), is by far the most used and best known quality function and has rapidly become an essential element of many optimisation metaheuristics based community detection methods. Metaheuristics have found wide applications to solve diverse kinds of complex optimisation problems in many realms. Four recent surveys on optimisation metaheuristics can be found in Reyes-Sierra and Coello (2006), Guliashki et al. (2009), Ong et al. (2010), Neri and Tirronen (2010), Zhou et al. (2011), Chen et al. (2011b), Mohankrishna et al. (2012), Neri and Cotta (2012), Boussaïd et al. (2013) and Črepinšek et al. (2013).

The focus of the recent survey on network community detection in Fortunato (2010) primarily deals with methods from the scope of physics and EAs have not been taken into account. Although in recent years many EAs have been devised to solve the community detection problem, there is seldom literature that give a macroscopic view on these methods and this is the very motivation for this survey. Thus, the purpose of this paper is to survey the EAs based community detection methods. These methods can be divided into two categories, the single objective optimisation based and the multiobjective optimisation based. Since most if not all of the EAs have similar workflow, this survey lays more emphasis on the algorithm components designing and the optimisation models construction.

The rest of this paper is organised as follows. Section 2 gives some related backgrounds including the fundamentals of network and network community detection. Section 3 depicts some common characters of EAs and briefly illustrates evolutionary multiobjective optimisation. Section 4 exhausts the fitness functions that are to be optimised by the available EAs based community detection methods. Section 5 illustrates the general operators designing of EAs for community detection, including the individual representation scheme, recombination and local search strategy. Section 6 lists some useful resources for network community detection. The conclusions are finally summarised in Section 7.

2 Network related backgrounds

2.1 Graph-based network notation

In order to better analyse a complex network, one direct way is to represent a network with a graph denoted as $G = \{V, E\}$, where V representing the network objects is the aggregation of vertices, and E representing the relations between the objects is the aggregation of edges. Graph G can be denoted by an adjacency matrix A whose element a_{ij} is defined as:

$$\begin{cases} a_{ij} = \omega_{ij} \; if \; \exists L < i, j > \\ a_{ij} = 0 \quad if \; \nexists L < i, j > \end{cases} \tag{1}$$

where $L < i, j >$ represents the link between node i and j and ω_{ij} denotes the weight of $L < i, j >$.

In the field of social science, the networks that include both positive and negative edges are called signed social networks (Doreian and Mrvar, 1996) or signed networks for short. In signed networks, the so called positive links (L^+) denote positive relationships such as friendship, common interests, and negative links (L^-) may denote negative relationships such as hostility, different interests, and so forth. A signed graph is normally denoted as $G = \{V, PE, NE\}$, where PE and NE represent the aggregations of positive and negative edges, respectively, and the element a_{ij} of the corresponding adjacency matrix A is defined as:

$$\begin{cases} a_{ij} = \omega_{ij} & \text{if } \exists L^+ < i, j > \\ a_{ij} = -\omega_{ij} & \text{if } \exists L^- < i, j > \\ a_{ij} = 0 & if \; \nexists L < i, j > \end{cases} \tag{2}$$

Matrix A is symmetric with the diagonal elements 0, but, if the corresponding network is directed, like the e-mail network, A is asymmetric.

2.2 Network community detection

Network community detection plays an important role in the networked data mining field. Community detection helps to discover latent patterns in networked data and it affects the ultimate knowledge presentation (Newman, 2003).

As illustrated above, a complex network can be expressed with a graph that is composed of nodes and edges. The task for network community detection is to separate the whole network into small parts which are also called communities. There is no uniform definition for community in the literature, but in academic domain, a community, also called a cluster or a module, is normally regarded as a groups of vertices which probably share common properties and/or play similar roles within the graph. Figure 1 exhibits the community detection problem under different network scenarios.

From the figure we can notice that community detection under dynamic context is quite different from the others. In a dynamic network, the community structure is temporally changed. How to design algorithms to uncover time-varying communities is challenging.

2.3 Qualitative community definition

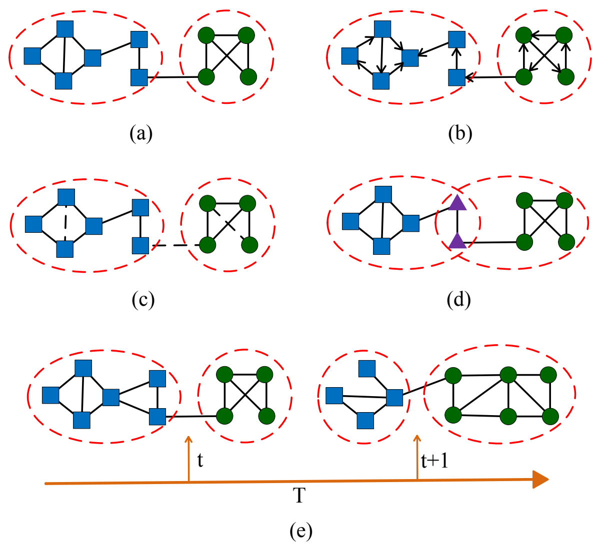
In order to formalise the qualitative community in unsigned network, Radicchi et al. (2004) gave a definition based on node degree. Given a network represented as $G = (V, E)$, where V is the set of nodes and E is the set of edges. Let k_i be the degree (the number of links that have connections with node i) of node i and A be the adjacency matrix of G. Given that $S \subset G$ is a subgraph, let $k_i^{in} = \sum_{i,j \in S} A_{ij}$ and $k_i^{out} = \sum_{i \in S, j \notin S} A_{ij}$ be the internal and external degree of node i, then S is a community in a strong sense if

$$\forall i \in S, k_i^{in} > k_i^{out} \tag{3}$$

S is a community in a weak sense if

$$\textstyle\sum_{i \in S} k_i^{in} > \sum_{i \in S} k_i^{out} \tag{4}$$

Figure 1 Graphical illustration of community detection, (a) common model, (b) directed model, (c) signed model, (d) overlapping model and (d) dynamic model (see online version for colours)



The above community definition only fits for unsigned networks. In Gong et al. (2014) the authors give a definition under signed context. Given a signed network modelled as $G = (V, PE, NE)$, where PE and NE are the set of positive and negative links, respectively. Given that $S \subset G$ is a subgraph, let $(k_i^+)^{in} = \sum_{j \in S, L_{ij} \in PE} A_{ij}$ and $(k_i^-)^{in} = \sum_{j \in S, L_{ij} \in NE} |A_{ij}|$ be the positive and negative internal degree of node i, respectively. Then S is a community in a strong sense if

$$\forall i \in S, (k_i^+)^{in} > (k_i^-)^{in} \tag{5}$$

Let

$$(k_i^-)^{out} = \sum_{j \notin S, L_{ij} \in NE} |A_{ij}| and$$

$$(k_i^+)^{out} = \sum_{j \notin S, L_{ij} \in PE} A_{ij}$$

be the negative and positive external degree of node i, respectively. Then S is a community in a weak sense if

$$\begin{cases} \sum_{i \in S}(k_i^+)^{in} > \sum_{i \in S}(k_i^+)^{out} \\ \sum_{i \in S}(k_i^-)^{out} > \sum_{i \in S}(k_i^-)^{in} \end{cases} \tag{6}$$

The above definitions only give the conditions that a community should satisfy, but they have not told how good on earth a community is. Therefore, there should have quantitative indexes that can measure the quality of a community. These indexes will be illustrated in Section 4.

3 EAs and multiobjective optimisation

3.1 General characters of EAs

EAs are originated from the evolution principles and behaviour of living things and have been recognised as effective approaches for solving many NP-Hard optimisation problems. Most if not all of the EAs share the following commom properties:

1. They are population based stochastic searching methods. A population consists of a set of individuals, each individual represents a solution to the optimisation problem. An EA optimises the problem by having a population of initialised solutions and then apply stochastic components to generate new solutions in the decision space.
2. They are recursively iterative methods. These methods iteratively search for optimal solutions in the search space. The search process will not stop untill the maximum iteration number or a prescribed threshold is reached.
3. They have some inherent parameters, like the population size and the maximum iteration number, etc. These parameters are normally set empirically.

A general framework of the optimisation metaheuristics is shown in Algorithm 1. The main difference between diverse metaheuristics lies in the strategy for generating new individuals, for instances, genetic algorithms (GAs) (Goldberg et al., 1989; Davis et al., 1991; Mitchell, 1998; Haupt and Haupt, 2004) generate new individuals by using genetic operations; swarm intelligence (SI) (Engelbrecht, 2005; Kennedy, 2006) produces solutions through learning tactics.

3.2 Multiobjective optimisation

EAs are effective tools for solving optimisation problems. However, in reality, many optimisation problems involve multiple objectives, i.e., there are more than one objectives to be optimised. For the community detection problems, it is natural to separately model the inter-community and the intra-community properties as two different optimisation objectives. So it is necessary to give a brief introduction to multiobjective optimisation.

Algorithm 1 General framework of EAs

Input: algorithm parameters, problem instance
Output: optimal solutions to the optimisation problem

1 **Begin**
2 population initialisation
3 store optimal solutions
4 **for** i=1 to max_iteration **do**
  a **for** each individual in the population, **do**
    1 generate a new individual through stochastic components
    2 evaluate the fitness of the new individual
  b **end for**
  c update optimal solutions
5 **end for**
6 **End**

A multiobjective optimisation problem can be mathematically formulated as[1]:

$$\min \; F(x) = (f_1(x), f_2(x), ..., f_k(x))^T \tag{7}$$

which subjects to $x = (x_1, x_2, \cdots, x_n) \in \Phi$. Where x is called the decision vector, and Φ is the feasible region in decision space. The objectives in equation (7) often conflict with each other. Improvement of one objective may lead to deterioration of another. Thus, a single solution, which can optimise all objectives simultaneously, does not exist. For multi-objective optimisation problems, the aim is to find good compromises (trade-offs) which are also called Pareto optimal solutions. The Pareto optimality concept was first proposed by Edgeworth and Pareto (Stadler, 1979). To understand the concept, here are some related definitions.

Definition 1 (Pareto optimality): A point $x^* \in \Phi$ is Pareto optimal if for every $x \in \Phi$ and $I = \{1, 2, ..., k\}$ either $\forall i \in I$, $f_i(x) = f_i(x^*)$ or, there is at least one $i \in I$ such that $f_i(x) > f_i(x^*)$.

Definition 2 (Pareto dominance): Given two vectors $x = (x_1, x_2, ..., x_n) \in \Phi$, $y = (y_1, y_2, ..., y_n) \in \Phi$, we say that x dominates y (denoted as $x \prec y$), if $x_i \leqq y_i$ for $i = 1, 2, ..., n$, and $x \neq y$. x is non-dominated with respect to Φ, if there does not exist another $x' \in \Phi$ such that $F(x') \prec F(x)$.

Definition 3 (Pareto optimal set): The set of all Pareto optimal solutions is called Pareto optimal set which is defined as:

$$PS = \{x \in \Phi | \neg \exists x^* \in \Phi, \; F(x^*) \prec F(x)\} \tag{8}$$

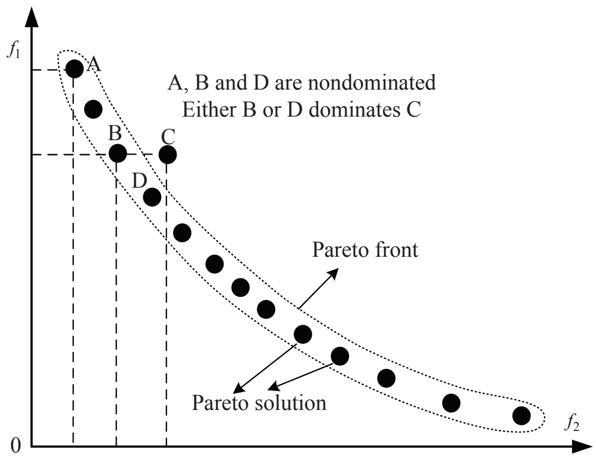
Definition 4 (Pareto front): The image of the Pareto set (PS) in the objective space is called the Pareto front (PF) which is defined as:

$$PF = \{F(x) | x \in PS\} \tag{9}$$

Figure 2 gives an example of the above mentioned definitions. Each dot except that labelled by C in the figure represents a non-dominated solution to the optimisation problem. The aim of a multiobjective optimisation algorithm is to find the set of those non-dominated solutions approximating the true PF.

In the last few years, many efforts have been devoted to the application of EAs to the development of multiobjective optimisation algorithms. A lot of multiobjective EAs have been proposed, e.g., Knowles and Corne (2000), Deb et al. (2002), Zitzler et al. (2002), Coello et al. (2004), Zhang and Li (2007), Gong et al. (2008), Bandyopadhyay et al. (2008), Bader and Zitzler (2011) and Yang (2013).

Figure 2 Graphical illustration of Pareto optimal solution and Pareto front



4 Fitness function

In the literature, many community evaluation criteria have been proposed and quantities of methods that combine either single objective or multiobjective EAs with community detection have emerged. Most if not all of these methods share the common feature that they model the community detection problem as different optimisation problems and then design different metaheuristics to solve them. This section is dedicated to summarising the fitness functions that are to be optimised by metaheuristics.

4.1 Single objective optimisation

4.1.1 Modularity-based model

The most popular evaluation criterion for community detection is the modularity (normally denoted as Q) proposed by Newman and Girvan (2004). The modularity index can be given in the following form:

$$Q = \frac{1}{2m}\sum_{i,j}^{n}\Big(A_{ij} - \frac{k_i \cdot k_j}{2m}\Big)\delta(i,j) \tag{10}$$

where n and m are the number of nodes and edges of a network, respectively. $\delta(i,j) = 1$, if node *i* and *j* are in the same group, otherwise, 0. By assumption, higher values of Q indicate better partitions.

Q is very popular, a lof of bio-inspired metaheuristics have been utilised to optimise Q to find the community structure with biggest Q value (Gog et al., 2007; Tasgin et al., 2007; Liu et al., 2007; Xiaodong et al., 2008; Lipczak and Milios, 2009; He et al., 2009; Shi et al., 2009a,b; Jin et al., 2010; Li et al., 2010; Jia et al., 2012; Gach and Hao, 2012; Huang et al., 2012; Shang et al., 2013; Li and Song, 2013; Li and Gao, 2013; Li et al., 2013b; Cai et al., 2014b). However, Q has several drawbacks. First, to maximise Q is proved to be NP-hard (Brandes et al., 2006). Second, large Q value does not always make sense. Random networks with no community structures can also possess high Q values (Guimerò et al., 2004; Reichardt and Bornholdt, 2006). Third, which is also the most important, Q has the resolution limitation (Fortunato and Barthelemy, 2007), i.e., maximising Q cannot discover communities whose sizes are smaller than a scale which depends on the total size of the network and on the degree of inter connectedness of the modules, even in the case scenario where modules are unambiguously defined.

To overcome these demerits, many researchers have devoted themselves to designing efficient operators for the optimisation algorithms to enhance the exploration and exploitation; some scholars make efforts to design new evaluation criteria, such as extended modularity (Reichardt and Bornholdt, 2006; Arenas et al., 2008a; Pons and Latapy, 2011), multi-resolution index (Li et al., 2008), and so forth. Because Q is originally designed for unsigned, unweighted, undirected, nonoverlapped and static networks, thus, many creative jobs have been done to extend Q to handle other types of networks.

Gómez et al. (2009) presented a reformulation of Q that allows the analysis of weighted, signed, and netwoks that have self-loops. The presented Q is formulised as:

$$Q_{sw} = \frac{1}{2(w^+ + w^-)} \sum_{i,j}\Big[w_{ij} - \Big(\frac{w_i^+ w_j^+}{2w^+} - \frac{w_i^- w_j^-}{2w^-}\Big)\Big]\delta(i,j) \tag{11}$$

where w_{ij} is the weight of the signed adjacency matrix, $w_i^+(w_i^-)$ denotes the sum of all positive (negative) weights of node *i*. Based on the Q_{sw} metric, the authors in Cai et al. (2014c) suggested a discrete particle swarm optimisation (DPSO) algorithm to detect communities from signed networks.

Q_{sw} can be easily changed to handle directed, weighted graphs (Arenas et al., 2007; Leicht and Newman, 2008; Rosvall and Bergstrom, 2008b), and the expression of directed and weighted Q reads:

$$Q_{dw} = \frac{1}{w}\sum_{i,j}\Big(A_{ij} - \frac{w_i^{out} \cdot w_j^{in}}{w}\Big)\delta(i,j) \tag{12}$$

where w_i^{out} (w_i^{in}) denotes the out-degrees (in-degrees) of node *i*. It can be noticed that the factor 2 is removed because the sum of the in-degrees (outdegrees), the number of non-vanishing elements of the asymmetric adjacency matrix, all equal w.

In the case when a node may belong to more than one community, Q has been modified to fit overlapping communities (Zhang et al., 2007; Nicosia et al., 2009; Shen et al., 2009), and a general expression reads:

$$Q_{ov}(U_k) = \sum_{c=1}^{k}\left[\frac{A(\overline{V}_c,\overline{V}_c)}{A(V,V)} - \left(\frac{A(\overline{V}_c,V)}{A(V,V)}\right)^2\right] \tag{13}$$

where $U_k = [u_1, u_2, \ldots, u_k]$ is a fuzzy parition of the nodes of the network into k clusters.

$$A(\overline{V}_c,\overline{V}_c) = \sum_{i\in\overline{V}_c}\sum_{j\in\overline{V}_c}((u_{ic}+u_{jc})/2)w_{ij},$$

where $\overline{V}_c$ is the set of vertices in community c,

$$A(\overline{V}_c,V) = A(\overline{V}_c,\overline{V}_c) + \sum_{i\in\overline{V}_c}\sum_{j\in V-\overline{V}_c}((u_{ic}+(1-u_{jc}))/2)w_{ij}$$

and

$$A(V,V) = \sum_{i\in V}\sum_{j\in V} w_{ij}.$$

u_{ic} is the membership value that node i belongs to community c.

The existing overlapping community detection methods can be roughly divided into two categories, the node-based (directly cluster nodes) and the link-based (cluster links and then map link communities to node communities) ones, but the the mainstream for single solution based overlapping community detection is to first utilise soft clustering technique such as fuzzy K-means to find a fuzzy partition of the nodes of a network into k clusters, and then apply a criterion to choose the best overlapping network partition (Zhang et al., 2007; Lancichinetti et al., 2009; Lázár et al., 2010; Rees and Gallagher, 2012). The key technique lies in the evaluation of an overlapped community. As long as an evaluation criterion is decided, bio-inspired metaheuristics can be easily utilised to solve this problem (Pizzuti, 2009b; Liu et al., 2010; Cai et al., 2011; Lin et al., 2013c). For more information about the fitness evaluation for overlapping communities, please refer to Chira and Gog (2011) and Xie et al. (2011).

Other extended criteria such as the local modularity can be found in Massen and Doye (2005) and Muff et al. (2005), the triangle modularity in Arenas et al. (2008b) and the bipartite modularity in Guimerà et al. (2007).

4.1.2 Multi-resolution model

To overcome the resolution limitation of modularity, many multi-resolution models have been developed. Pizzuti (2008) proposed a genetic algorithm for community detection. The highlight of the work is the suggested community score (*CS*) evaluation metric. Let $\mu_i = \frac{1}{|S|}k_i^{in}$ be the fraction of edges connecting node i to the other nodes in S and $M(S) = \frac{\sum_{i\in S}(\mu_i)^r}{|S|}$ be the power mean of S of order r. $|S|$ is the cardinality of S, i.e., the number of nodes in S. We further define $v_S = \frac{1}{2}\sum_i k_i^{in}$ be the volume of S, i.e., the number of edges connecting vertices inside S, then the score of S is defined as $score(S) = M(S) \times v_S$. Assume that G has a partition of k subgraphs, i.e., $\Omega = \{S_1, S_2, ..., S_k\}$, then *CS* can be written as:

$$CS = \textstyle\sum_{i=1}^{k} score(S_i) \tag{14}$$

The *CS* metric takes one parameter r which is hard to tune. The author claims that higher values of the exponent r bias the *CS* towards matrices containing a low number of zeroes, i.e., higher values of r help in detecting communities.

Li et al. (2008) put forward the modularity density (*D*) index. D can break the resolution limitation brought by Q. For an unsigned network, let us define

$$L(S_a,S_b) = \sum_{i\in S_a, j\in S_b} A_{ij} \text{ and}$$
$$L(S_a,\overline{S_a}) = \sum_{i\in S_a, j\in\overline{S_a}} A_{ij},$$

where $\overline{S_a} = \Omega - S_a$. Then D is defined as:

$$D_\alpha = \sum_{i=1}^{k}\frac{2\alpha L(S_i,S_i) - 2(1-\alpha)L(S_i,\overline{S_i})}{|S_i|} \tag{15}$$

where $\alpha \in [0,1]$ is a resolution control parameter. D_α can be viewed as a combination of the ratio association and the ratio cut (Dhillon et al., 2004). Generally, optimise the ratio association algorithm often divides a network into small communities, while optimise the ratio cut often divides a network into large communities. By tuning the α value, we can use this general function to uncover more detailed and hierarchical organisation of a complex network. Based on modularity density, many algorithms have emerged (Guoqiang and Xiaofang, 2010; Chen et al., 2011a; Gong et al., 2011a, 2012b; Li et al., 2013a; Cai et al., 2014a).

4.2 Multiobjective optimisation

Many real-world optimisation problems involve multiple objectives. From the statement of the community detection problem discussed earlier we can notice that, community detection can also be modelled as multiobjective optimisation problems. Many multiobjective optimisation based community detection methods have been developed in this respect. Each run of these methods can yield a set of community partitions for the decision maker to choose. The most important point for these methods should own to their abilities for breaking through the resolution limit of modularity. As stated earlier, components used in single objective optimisation models, such as the individual representation, recombination, etc., serve multiobjective optimisation models as well. This subsection primarily deals with the multiobjective community detection models.

4.2.1 General model

As stated earlier, for an unsigned network, the links within a community should be dense while the links between communities should be sparse, as for a signed network, the inter and intra links should all be dense. On the basis of this property, many multiobjective community models are established.

Pizzuti (2012) and Pizzuti (2009a) proposed a multiobjective genetic algorithm-based method called MOGA-Net. In this method, the author modelled the community detection task as a multiobjective optimisation problem and then applied the fast elitist non-dominated sorting genetic algorithm (NSGA-II) (Deb et al., 2002) framework to solve it. The two objectives introduced are the *CS* and the *CF*. Thus, the proposed optimisation model is:

$$\max \begin{Bmatrix} f_1 = CS \\ f_2 = -CF \end{Bmatrix} \tag{16}$$

CF (community fitness) is a criterion put forward by Lancichinetti et al. (2008). *CF* is formulated as:

$$CF = \textstyle\sum_{S \in \Omega} \sum_{i \in S} \frac{k_i^{in}}{k_i} \tag{17}$$

From the formulation of *CF* and *CS* we may notice that, *CF* to some extent measures the link density within communities, while *CS* can be regarded as an index to measure the averaged degrees within communities.

An improved version of MOGA-Net can be found in Butun and Kaya (2013). To optimise the above model, other metaheuristics, such as the multi–objective enhanced firefly algorithm (Amiri et al., 2013), hybrid EA based on harmony search algorithm (HSA, Geem et al., 2001) and chaotic local search (CLS) (Amiri et al., 2011, 2012a, 2012b), non-dominated neighbour immune algorithm (Gong et al., 2011b), have all find their niche in community detection.

In paper (Gong et al., 2012c) the authors presented a multiobjective evolutionary algorithm based on decomposition (MOEA/D) based method. MOEA/D is proposed by Zhang and Li (2007). The highlight of this work is the newly cranked out multiobjective community optimisation model which optimises two objectives termed as *NRA* (negative ratio association) and ratio cut (*RC*). The optimisation model is:

$$\min \begin{Bmatrix} NRA = -\sum_{i=1}^{k} \frac{L(S_i, S_i)}{|S_i|} \\ RC = \sum_{i=1}^{k} \frac{L(S_i, \overline{S_i})}{|S_i|} \end{Bmatrix} \tag{18}$$

It can be noticed that equation (18) is the decomposition of equation (15). *RC* measures the link density between two communities and *RA* calculates the link density within a community. To minimise *NRA* and *RC* we can ensure that the connections within a community is dense and the links between communities are sparse. A similar optimisation model can be found in Gong et al. (2013).

Other optimisation models such as maximising the combinations of Q and *CS* can be found in Agrawal (2011), and maximising the two parts of the Q index, i.e., Q is decomposed into two objectives, can be found in Shi et al. (2012). A three objectives model can be found in Shelokar et al. (2014). Small surveys on the selection of objective functions in multiobjective community detection can be found in Shi et al. (2011a, 2013).

4.2.2 Signed model

Many social networks involve friendly and hostile relations between the objects that compose the networks. These networks are called signed networks. In Gong et al. (2014) the authors put forward a novel discrete multiobjective PSO framework for community detection. To handle signed networks, the authors have suggested a signed optimisation model which optimises two objectives named as signed ratio association (*SRA*) and signed ratio cut (*SRC*). The optimisation model reads:

$$\min \begin{Bmatrix} SRA = -\sum_{i=1}^{k} \frac{L^+(S_i, S_i) - L^-(S_i, S_i)}{|S_i|} \\ SRC = \sum_{i=1}^{k} \frac{L^+(S_i, \overline{S_i}) - L^-(S_i, \overline{S_i})}{|S_i|} \end{Bmatrix} \tag{19}$$

where $L^+(S_i, S_j) = \sum_{i \in S_i, j \in S_j} A_{ij}, (A_{ij} > 0)$ and $L^-(S_i, S_j) = \sum_{i \in S_i, j \in S_j} |A_{ij}|, (A_{ij} < 0)$. To minimise *SRA* and *SRC* we can make sure that the positive links within a community are dense while the negative links between communities are also dense, which is in accordance with the feature of signed community.

In Amelio and Pizzuti (2013) the authors put forward another signed optimisation model which uses the NSGA-II framework to optimise it. The model reads:

$$\min \begin{Bmatrix} f_1 = -Q_{sw} \\ f_2 = frustration \end{Bmatrix} \tag{20}$$

where $frustration = \sum_{i,j}^{n} (A_{ij}^{+}(1 - \delta(i,j)) - A_{ij}^{-}\delta(i,j))$. The first objective Q_{sw} measures how good a signed community is and to minimise $frustration$ we will ensure that the sum of the negative links within a community and the positive links between difference communities are minimum.

Recently, to detect communities from signed networks, the authors in Liu et al. (2014) put forward a signed optimisation model based on node similarity. The optimisation model is as follows:

$$\max \begin{Bmatrix} f_{pos-in}(\Omega) = \frac{1}{m} \sum_{i=1}^{m} \frac{P_{in}^{S_i}}{P_{in}^{S_i} + P_{out}^{S_i}} \\ f_{neg-out}(\Omega) = \frac{1}{m} \sum_{i=1}^{m} \frac{N_{out}^{S_i}}{N_{in}^{S_i} + N_{out}^{S_i}} \end{Bmatrix} \tag{21}$$

where $P_{in}^{S_i}$ (or $P_{out}^{S_i}$) is the internal (or external) positive similarity of community S_i, and $N_{in}^{S_i}$ (or $N_{out}^{S_i}$) is the internal (or external) negative similarity of community S_i. See reference Liu et al. (2014) for more information about the similarity of a community. To maximise f_{pos-in} we can ensure high positive similarities within communities, and to maximise $f_{neg-out}$ we can guarantee high negative similarities between different communities.

4.2.3 Overlapping model

In real world, a node of a network may belong to more than one community, just like the friendship network. From the perspective of finding overlapping communities, intuitively, the nodes that connect multiple communities with similar strength are more likely to be overlapping nodes. For instance, if node i has both l links with community a and b, then we can regard i as an overlapping node. From the viewpoint of finding nonoverlapping or separated communities, the less the number of overlapping nodes, the more the separated communities.

Based on the above principle, the authors in Liu et al. (2010) put forward a three objectives optimisation model reads:

$$\max \left\{ \begin{array}{l} f_1 = f_{\text{quality}}(\Omega) = \dfrac{CF}{k} \\ f_2 = f_{\text{separated}}(\Omega) = - \mid V_{overlap} \mid \\ f_3 = f_{\text{overlapping}}(\Omega) = \displaystyle\sum_{i \in V_{overlap}} \min_{s \in \Omega} \left\{ \frac{k_i^s}{k_i} \right\} \end{array} \right\} \quad (22)$$

where k_i^s denotes the number of edges connect node i and community s, $V_{overlap}$ is the set of the overlapping nodes. To maximise f_2 and f_3 one can get a tradeoff between nonoverlapping and overlapping communities.

4.2.4 Dynamical model

In reality, networks may evolve with the time, the nodes and the links may disappear or new nodes may just come out, therefore, the community structures are also changing according to the time. However, traditional approaches mostly focuse on static networks for small groups. As the technologies move forward, in the presence of big data, how to design methods and tools for modelling and analysing big dynamic networks is a challenging research topic in the years to come. To analyse the community structures of dynamical networks will help to predict the change tendency which may give support to the analysis of other network or networked scientific issues. Community detection in dynamic networks is challenging.

Dynamic community detection is normally based on a temporal smoothness framework which assumes that the variants of community division in a short time period are not desirable (Folino and Pizzuti, 2010a). According to the temporal smoothness framework, the community detection in dynamic networks can be naturally modelled as a bi-objective optimisation problem. The optimisation of one objective is to reveal a community structure with high quality at this moment, and the optimisation of the other objective is to uncover a community structure at the next moment which is highly similar with that at the previous time (Folino and Pizzuti, 2010b,a; Gong et al., 2012a; Chen et al., 2013; Folino and Pizzuti, 2013). The commonly used dynamical optimisation model can be written as:

$$\max \left\{ \begin{array}{l} f_1 = CS \text{ or } Q \text{ or } D_\alpha \\ f_2 = NMI \end{array} \right. \quad (23)$$

NMI, normalised mutual information (Danon et al., 2005), comes from the field of information theory. *NMI* can be regarded as a similarity index. For the community detection problem, given that A and B are two partitions of a network, respectively, C is a confusion matrix, C_{ij} equals to the number of nodes shared in common by community i in partition A and by community j in partition B. Then $NMI(A, B)$ is written as:

$$NMI = \frac{-2 \sum_{i=1}^{C_A} \sum_{j=1}^{C_B} C_{ij} log(C_{ij} \cdot n / C_{i.} C_{.j})}{\sum_{i=1}^{C_A} C_{i.} log(C_{i.}/n) + \sum_{j=1}^{C_B} C_{.j} log(C_{.j}/n)} \quad (24)$$

where C_A (or C_B) is the number of clusters in partition A(or B), $C_{i.}$ (or $C_{.j}$) is the sum of elements of C in row i(or column j). $NMI(A, B) = 1$ means that A and B are identical and $NMI(A, B) = 0$ indicates that A and B are completely different.

The first objective in equation (23) is the snapshot cost which measures how well a community structure A is at time t and the second objective is the temporal cost which measures how similar the community structure B is at time $t + 1$ with the previous community structure A.

Another dynamical model which maximises the Min-max cut and global silhouette index can be found in Kim et al. (2010).

4.3 Model selection

It should be pointed out that multiobjective optimisation returns a set of equally good solutions. Each solution is a tradeoff between the objective functions.

For the multiobjective community detection problem, each Pareto-optimal solution represents a network partition which provides great convenience to analyse the structures of a network at different hierarchical levels. However, it is usually required that there exist a criterion so as to automatically select one solution with respect to another.

For the community detection problem, to this end, the modularity index seems to be the most used model selection mechanism. That is to say, choose the very solution from the PF that has the biggest modularity value. If this very solution is more than one (since different network partitions may have the same modularity value), then we use the D_α, the CF, the CS, etc., indexes as the secondary selection standard. Some other selection models can be found in Shi et al. (2011b).

5 Operator designing

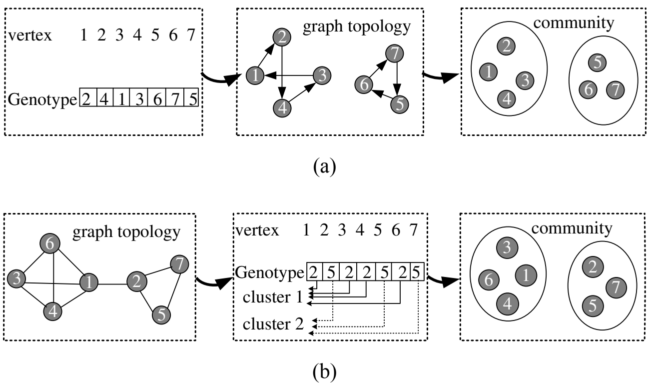
The above section exhausts the fitness functions that are to be optimised by EAs based community detection methods.

This section is devoted to summarising the fundamental operators designing of these methods.

5.1 Individual representation

Individual representation bridges the gap between metaheuristics and optimisation problems. For the community detection problem, an individual represents a network partition. In the existing research findings, there are two representative individual representation schemas: the locus-based and the string-based. Figure 3 clearly shows the principles of the two schemas.

Figure 3 (a) Locus-based individual representation scheme
(b) String-based individual representation schema



Notes: (a) Left: one possible genotype.
Middle: translate the genotype into the graph structure (shown as directed only to help the understanding).
Right: the final clusters (every connected component is interpreted as an individual cluster).
(b) Middle: one possible genotype.
Right: decoded network communities

It can be seen from the above figure, locus-based coding scheme takes the advantage of network linkage relations, but the corresponding diversity promotion strategies such as the crossover and mutation operations are hard to design. By constrast, the string-based schema is easier to code and decode and it can provide great convenience to the diversity promotion strategies. It should be pointed out that when initialising individuals, one may make full use of network prior knowledge to speed the convergence.

5.2 Individual reproduction

Whatever kind of individual representation schema we use, our goal is to handle network problems using optimisation metaheuristics. As regard the optimisation metaheuristics, one key factor is the reproduction operation. Reproduction acts as the backbone for preserving population diversity. Generally, different representation schema could lead to different designing of the individual reproduction operators. However, it is strongly recommended that network priori knowledge should be utilised to design efficient individual reproduction operators.

With regard to the reproduction, the crossover and mutation operations are two widely used mechanisms to generate offsprings. In the following we will take the string-based individual representation schema as an example to show the crossover and the mutation operations for the community detection problem. Tables 1 and 2 give examples of the two-way crossover and the one-point mutation operators, respectively.

Table 1 Illustration of two-way crossover operator

v		x_a		x_b		x_c	x_d		x_a		x_b		v
1		⑤	→	2	→	⑤	5		5		2		1
2		3		6		6	⑥	←	3	←	⑥		2
③	→	⑤	→	6	→	⑤	⑥	←	5	←	⑥	←	③
4		7		5		5	7		7		5		4
5		2		6		6	⑥	←	2	←	⑥		5
6		⑤	→	3	→	⑤	5		5		3		6
7		3		2		2	3		3		2		7

Note: The position of node 3 is selected to carry out the operation.

Table 2 Illustration of one-point mutation operator

v		x_a		x_c	x_d		x_b		v
1		5		5	2		2		1
2		3		3	6		6		2
③	→	⑤	→	⑦	③	←	⑥	←	③
4		7		7	5		5		4
5		2		2	6		6		5
6		5		5	3		3		6
7		3		3	3		2		7

Note: The position of node 3 is selected to carry out the operation.

Given that x_a = (5, 3, 5, 7, 2, 5, 3) and x_b = (2, 6, 6, 5, 6, 3, 2) are two individuals. Here we take the position of node 3 as an example. The two-way crossover operation works in the following way: first we determine the 3rd element of x_a, i.e., 5; afterward we find out all the positions in x_a that have the value 5, these positions are the 1st, 3rd and 6th; then we change the 1st, 3rd and 6th elements of x_b into 5, and an offspring x_c = (5, 6, 5, 5, 6, 5, 2) is generated. The changes to x_a work in the same way, and an offspring x_d = (5, 6, 6, 7, 6, 5, 3) is generated. For the one-point mutation operation, we randomly change the element of the permutation in the 3rd position as shown in Table 2.

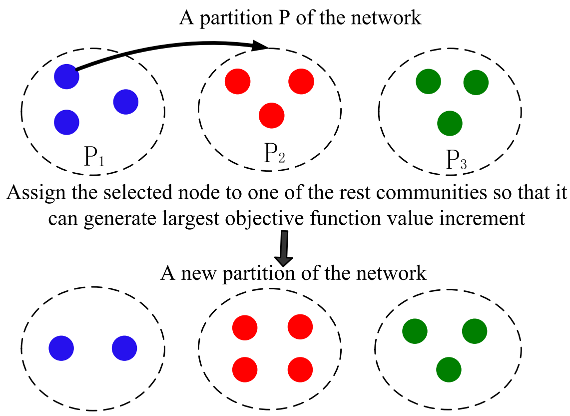
The two-way crossover and the one-point mutation operators shown above are very representative for the community detection problem. Many approaches either adopt these methods or just make some minor changes to them. For different types of networks, these two mechanisms may also be redesigned due to the specific network properties, but the main idea is the same, i.e., recombine two individuals and/or modify some information of an individual so as to produce new solutions.

5.3 Individual local search

Since metaheuristics are stochastic approaches, they cannot ensure global convergence. Hence, much creative effort has been done to enhance the search abilities of the heuristics based community detection methods, and amongst which, the designing of local search operators has gained much attention.

The local search operator aims to search for promising region around a given solution. In Gong et al. (2011a) the authors have designed a hill climbing based local search mechanism illustrated in Figure 4.

Figure 4 Graphical illustration of the hill climbing based local search strategy (see online version for colours)



It can be seen from the figure that the hill climbing based local search is actually a greedy algorithm since each step of the operator assigns a node to the very community that can generate biggest fitness increment. It should be pointed out that the fitness increment may be negative. The authors argue that to accept this kind of deteriorate solution with certain probability may assist exploitation. Based on this idea, they suggested a simulated annealing (Kirkpatrick et al., 1983) based local search in Gong et al. (2012b).

However, for a real-life network, the number of communities may be very big, thus, the local search operators in Gong et al. (2011a, 2012b) are time consuming and can hardly be applied to real applications. To save the complexity, in Cai et al. (2014a) they put forward a new local search which is based on the common phenomenon that two objects with no connections are less likely to be in the same community. Therefore, the key component of the newly suggested local search assign a node to one of the neighbour communities (communities that the neighbour nodes belong to) that can bring about largest fitness increment.

For a big network, an optimisation metaheuristic based community detection method is actually facing a large scale global optimisation (LSGO) problem. To further enhance the search ability, multi-level learning strategies have been invented. In Ma et al. (2014) the authors have put forward a three-level learning based Memetic algorithm for community detection. The proposed three-level learning strategy takes the advantages of label propagation (Bagrow and Bollt, 2005) and BGLL (Blondel et al., 2008) to design local search.

One point we should keep in mind is that, different networks have different properties, and only by taking advantages of these properties can we design metaheuristics with high efficiency.

6 Network relevant resource

For the network community detection problem, one may need some instances for testing purpose. Also, one may need to visualise maybe the topology or the dynamic changes of a particular network. Besides, one may need to compare with some state-of-the-art community detection methods. Based on such considerations, several useful websites are listed as follows:

- http://www-personal.umich.edu/˜mejn/ (mark Newman website)
- http://deim.urv.cat/˜aarenas/data/welcome.htm (Alex Arenas website)
- http://snap.stanford.edu/index.html (Stanford network analysis project. Diverse kinds of network data and graphical visualisation softwares and tools and useful codes are available)
- http://www.correlatesofwar.org/ (the correlates of war project, a large amount of signed networks mainly related to war are free to access)
- http://www.gmw.rug.nl/˜huisman/sna/software.html (a collection of softwares for social network analysis)
- http://tuvalu.santafe.edu/˜aaronc/hierarchy/ (hierarchical random graphs)
- http://www.cs.unm.edu/˜aaron/research/fastmodularity.htm [the code of the FM algorithm (Clauset et al., 2004)]
- http://perso.uclouvain.be/vincent.blondel/research/louvain.html [the code of the BGLL algorithm (Blondel et al., 2008)]
- http://www.tp.umu.se/˜rosvall/code.html [the code of the Infomod (Rosvall and Bergstrom, 2007) and the Infomap (Rosvall and Bergstrom, 2008a)algorithms]
- http://see.xidian.edu.cn/faculty/mggong/publication.htm [the code of the MODPSO algorithm (Gong et al., 2014)].

7 Concluding remarks

Network analysis is one of the theoretical underpinnings of social computing and big data. The network community detection serves as the backbone of network analysis. The past decades have witnessed the prosperity of the research on community detection. This paper tries to survey the community detection based on EAs. The main idea of these methods is the same, i.e., they model the community detection as either a single objective or a multiobjective optimisation problem and then design optimisation metaheuristics to solve it. In this survey, the constructed optimisation models and the key components of these methods are exhausted.

Although a lot of EAs have been successfully developed to mine the community structures in complex networks, according to the no free lunch (NFL) theory (Wolpert and Macready, 1997), there is no one-for-all method that can deal with all kinds of networks. Different networks have different space-time properties. When designing algorithms to solve community detection problem, one should take into account the special characters of the networks. In addition, one thing should be kept in mind that due to the huge volume of data set, metaheuristics based community detection is an LSGO problem. LSGO problems refer to optimisation problems that involve a large number of decision variables. LSGO challenges existing optimisation techniques. Researches on scaling up EAs to solve LSGO problems have attracted much attention (Zhao et al., 2008; Yang et al., 2008; Wang and Li, 2010; Li and Yao, 2012). For the LSGO community detection problem, in the presence of big data, how to design fast and effective algorithms is worth thinking.

We expect that complex network analysis's scope will continue to expand and its applications to multiply. We are positive that methods and theories that work for community detection are helpful for other network issues such as link prediction, network recommender, network robustness, network resource allocation, anomaly detection, network core detection, network ranking, network compression, network security, network sentiment computing, and so forth. From both theoretical and technological perspectives, network community detection technology will move beyond network structure analysis toward emphasising network intelligence. We do hope that this survey can benefit scholars who set foot in this field. Our future work will focus on more in-depth analysis of network issues. Such analysis is expected to shed light on how networks change the real world.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61273317 and 61422209), the National Top Youth Talents Programme of China, the Specialised Research Fund for the Doctoral Programme of Higher Education (Grant No. 20130203110011) and the Fundamental Research Fund for the Central Universities (Grant Nos. K50510020001, K5051202053 and JB142001-8).

References

Agrawal, R. (2011) 'Bi-objective community detection (BOCD) in networks using genetic algorithm', in *Contemporary Computing*, pp.5–15.

Amelio, A. and Pizzuti, C. (2013) 'Community mining in signed networks: a multiobjective approach', in *Proc. IEEE/ACM Int. Conf. Adv. Social Networks Analysis and Mining*, pp.95–99.

Amiri, B., Hossain, L. and Crawford, J.W. (2011) 'An efficient multiobjective evolutionary algorithm for community detection in social networks', in *Proc. IEEE Congr. Evol. Comput.*, pp.2193–2199.

Amiri, B., Hossain, L. and Crawford, J.W. (2012a) 'A hybrid evolutionary algorithm based on hsa and cls for multi-objective community detection in complex networks', in *Proc. IEEE/ACM Int. Conf. Adv. in Social Networks Analysis and Mining*, pp.243–247.

Amiri, B., Hossain, L. and Crawford, J.W. (2012b) 'A multiobjective hybrid evolutionary algorithm for clustering in social networks', in *Proc. 14th Int. Conf. Genet. Evol. Comput.*, pp.1445–1446.

Amiri, B., Hossain, L., Crawford, J.W. and Wigand, R.T. (2013) 'Community detection in complex networks: multi–objective enhanced firefly algorithm', *Knowledge-Based Systems*, Vol. 46, pp.1–11.

Arenas, A., Duch, J., Fernández, A. and Gómez, S. (2007) 'Size reduction of complex networks preserving modularity', *New J. Phys.*, Vol. 9, No. 6, pp.176.

Arenas, A., Fernández, A. and Gómez, S. (2008a) 'Analysis of the structure of complex networks at different resolution levels', *New J. Phys.*, Vol. 10, No. 5, p.053039.

Arenas, A., Fernandez, A., Fortunato, S. and Gómez, S. (2008b) 'Motif-based communities in complex networks', *Journal of Physics A: Mathematical and Theoretical*, Vol. 41, No. 22, p.224001.

Bader, J. and Zitzler, E. (2011) 'Hype: an algorithm for fast hypervolume-based many-objective optimization', *Evol. Comput.*, Vol. 19, No. 1, pp.45–76.

Bagrow, J.P. and Bollt, E.M. (2005) 'A local method for detecting communities', *Phys. Rev. E*, October, Vol. 72, p.046108.

Bandyopadhyay, S., Saha, S., Maulik, U. and Deb, K. (2008) 'A simulated annealing-based multiobjective optimization algorithm: AMOSA', *IEEE Trans. Evol. Comput.*, Vol. 12, No. 3, pp.269–283.

Barabási, A-L. and Albert, R. (1999) 'Emergence of scaling in random networks', *Science*, Vol. 286, No. 5439, pp.509–512.

Blondel, V.D., Guillaume, J-L., Lambiotte, R. and Lefebvre, E. (2008) 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, p.10008.

Boussaïd, I., Lepagnot, J. and Siarry, P. (2013) 'A survey on optimization metaheuristics', *Inf. Sci.*, Vol. 237, pp.82–117.

Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z. and Wagner, D. (2006) 'Maximizing modularity is hard', arXiv preprint physics/0608255.

Butun, E. and Kaya, M. (2013) 'A multi-objective genetic algorithm for community discovery', in *Proc. IEEE 7th Int. Conf. Intell. Data Acquisition and Adv. Comput. Syst.*, Vol. 1, pp.287–292.

Cai, Y., Shi, C., Dong, Y., Ke, Q. and Wu, B. (2011) 'A novel genetic algorithm for overlapping community detection', in *Advanced Data Mining and Applications*, pp.97–108.

Cai, Q., Gong, M., Ma, L. and Jiao, L. (2014a) 'A novel clonal selection algorithm for community detection in complex networks', *Comput. Intell.*, pp.1–24.

Cai, Q., Gong, M., Ma, L., Ruan, S., Yuan, F. and Jiao, L. (2014b) 'Greedy discrete particle swarm optimization for large-scale social network clustering', *Inf. Sci.*

Cai, Q., Gong, M., Shen, B., Ma, L. and Jiao, L. (2014c) 'Discrete particle swarm optimization for identifying community structures in signed social networks', *Neural Networks*, Vol. 58, pp.4–13.

Chen, G., Wang, Y. and Yang, Y. (2011a) 'Community detection in complex networks using immune clone selection algorithm', *International Journal of Digital Content Technology and its Applications*, Vol. 5, pp.182–189.

Chen, X., Ong, Y-S., Lim, M-H. and Tan, K.C. (2011b) 'A multi-facet survey on memetic computation', *IEEE Trans. Evol. Comput.*, Vol. 15, No. 5, pp.591–607.

Chen, G., Wang, Y. and Wei, J. (2013) 'A new multiobjective evolutionary algorithm for community detection in dynamic complex networks', *Mathematical Problems in Engineering*, Vol. 2013.

Chira, C. and Gog, A. (2011) 'Fitness evaluation for overlapping community detection in complex networks', in *Proc. IEEE Congr. Evol. Comput.*, pp.2200–2206.

Clauset, A., Newman, M.E.J. and Moore, C. (2004) 'Finding community structure in very large networks', *Phys. Rev. E*, Vol. 70, No. 6, p.066111.

Coello, C.A.C., Pulido, G.T. and Lechuga, M.S. (2004) 'Handling multiple objectives with particle swarm optimization', *IEEE Trans. Evol. Comput.*, Vol. 8, No. 3, pp.256–279.

Črepinšek, M., Liu, S-H. and Mernik, M. (2013) 'Exploration and exploitation in evolutionary algorithms: a survey', *ACM Computing Surveys (CSUR)*, Vol. 45, No. 3, p.35.

Danon, L., Díaz-Guilera, A., Duch, J. and Arenas, A. (2005) 'Comparing community structure identification', *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2005, No. 09, p.09008.

Davis, L. et al. (1991) *Handbook of Genetic Algorithms*, Vol. 115, Van Nostrand Reinhold New York.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002) 'A fast and elitist multiobjective genetic algorithm: NSGA-II', *IEEE Trans. Evol. Comput.*, Vol. 6, No. 2, pp.182–197.

Dhillon, I.S., Guan, Y. and Kulis, B. (2004) 'Kernel k-means: spectral clustering and normalized cuts', in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery and Data Mining*, pp.551–556.

Doreian, P. and Mrvar, A. (1996) 'A partitioning approach to structural balance', *Social Networks*, Vol. 18, No. 2, pp.149–168.

Engelbrecht', A.P. (2005) *Fundamentals of Computational Swarm Intelligence*, Vol. 1, Wiley Chichester.

Folino, F. and Pizzuti, C. (2010a) 'Multiobjective evolutionary community detection for dynamic networks', in *Proc. 12th Annu. Conf. Genet. Evol. Comput.*, pp.535–536.

Folino, F. and Pizzuti, C. (2010b) 'A multiobjective and evolutionary clustering method for dynamic networks', in *Proc. Int. Conf. Advances in Social Networks Analysis and Mining*, pp.256–263.

Folino, F. and Pizzuti, C. (2013) 'An evolutionary multiobjective approach for community discovery in dynamic networks', *IEEE Trans. Knowl. Data Eng.*, Vol. 99, pp.1–1.

Fortunato, S. and Barthelemy, M. (2007) 'Resolution limit in community detection', *PNAS*, Vol. 104, No. 1, pp.36–41.

Fortunato, S. (2010) 'Community detection in graphs', *Phys. Rep.*, Vol. 486, Nos. 3–5, pp.75–174.

Gómez, S., Jensen, P. and Arenas, A. (2009) 'Analysis of community structure in networks of correlated data', *Phys. Rev. E*, July, Vol. 80, p.016114

Gach, O. and Hao, J-K. (2012) 'A memetic algorithm for community detection in complex networks', in *Parallel Problem Solving from Nature-PPSN XII*, pp.327–336.

Geem, Z.W., Kim, J.H. and Loganathan, G.V. (2001) 'A new heuristic optimization algorithm: harmony search', *Simulation*, Vol. 76, No. 2, pp.60–68.

Girvan, M. and Newman, M.E.J. (2002) 'Community structure in social and biological networks', *PNAS*, Vol. 99, No. 12, pp.7821–7826.

Gog, A., Dumitrescu, D. and Hirsbrunner, B. (2007) 'Community detection in complex networks using collaborative evolutionary algorithms', in *Advances in Artificial Life*, pp.886–894.

Goldberg, D.E. et al. (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*, Vol. 412, Addison-wesley Reading Menlo Park.

Gong, M., Jiao, L., Du, H. and Bo, L. (2008) 'Multiobjective immune algorithm with nondominated neighbor-based selection', *Evol. Comput.*, Vol. 16, No. 2, pp.225–255.

Gong, M., Fu, B., Jiao, L. and Du, H. (2011a) 'Memetic algorithm for community detection in networks', *Phys. Rev. E*, Vol. 84:056101.

Gong, M., Hou, T., Fu, B. and Jiao, L. (2011b) 'A non-dominated neighbor immune algorithm for community detection in networks', in *Proc. 13th Annu. Conf. Genet. Evol. Comput.*, pp.1627–1634.

Gong, M., Zhang, L., Ma, J. and Jiao, L. (2012a) 'Community detection in dynamic social networks based on multiobjective immune algorithm', *Journal of Computer Science and Technology*, Vol. 27, No. 3, pp.455–467.

Gong, M., Cai, Q., Li, Y. and Ma, J. (2012b) 'An improved memetic algorithm for community detection in complex networks', in *Proc. IEEE Congr. Evol. Comput.*, pp.1–8.

Gong, M., Ma, L., Zhang, Q. and Jiao, L. (2012c) 'Community detection in networks by using multiobjective evolutionary algorithm with decomposition', *Phys. A*, Vol. 391, No. 15, pp.4050–4060.

Gong, M., Chen, X., Ma, L., Zhang, Q. and Jiao, L. (2013) 'Identification of multi-resolution network structures with multi-objective immune algorithm', *Appl. Soft Comput.*, Vol. 13, No. 4, pp.1705–1717.

Gong, M., Cai, Q., Chen, X. and Ma, L. (2014) 'Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition', *IEEE Trans. Evol. Comput.*, Vol. 18, No. 1), pp.82–97.

Guimerò, R., Sales-Pardo, M. and Anaral, L.A.N. (2004) 'Modularity from fluctuations in random graphs and complex networks', *Phys. Rev. E*, Vol. 70, p.025101.

Guimerà, R., Sales-Pardo, M. and Amaral, L.A.N. (2007) 'Module identification in bipartite and directed networks', *Phys. Rev. E*, Vol. 76, No. 3, p.036102.

Guliashki, V., Toshev, H. and Korsemov, C. (2009) 'Survey of evolutionary algorithms used in multiobjective optimization', *Problems of Engineering Cybernetics and Robotics*, Bulgarian Academy of Sciences.

Guoqiang, C. and Xiaofang, G. (2010) 'A genetic algorithm based on modularity density for detecting community structure in complex networks', in *Proc. 2010 Int. Conf. Comput. Intell. Secur.*, pp.151–154.

Haupt, R.L. and Haupt, S.E. (2004) *Practical Genetic Algorithms*, John Wiley & Sons.

He, D., Wang, Z., Yang, B. and Zhou, C. (2009) 'Genetic algorithm with ensemble learning for detecting community structure in complex networks', in *Proc. 4th Int. Conf. Comput. Sci. Converg. Inf. Tech.*, pp.702–707.

Huang, Q., White, T., Jia, G., Musolesi, M., Turan, N., Tang, K., He, S., Heath, J.K. and Yao, X. (2012) 'Community detection using cooperative co-evolutionary differential evolution', in *Parallel Problem Solving from Nature-PPSN XII*, pp.235–244.

Jia, G., Cai, Z., Musolesi, M., Wang, Y., Tennant, D.A., Weber, R.J.M., Heath, J.K. and He, S. (2012) 'Community detection in social and biological networks using differential evolution', in *Learning and Intelligent Optimization*, pp.71–85.

Jin, D., He, D., Liu, D. and Baquero, C. (2010) 'Genetic algorithm with local search for community mining in complex networks', in *Proc. 22nd IEEE Int. Conf. Tools with Artif. Intell.*, volume 1, pp.105–112.

Kennedy, J. (2006) *Swarm intelligence*, Springer.

Kim, K., McKay, R.I. and Moon, B-R. (2010) 'Multiobjective evolutionary algorithms for dynamic social network clustering', in *Proc. 12th Annu. Conf. Genet. Evol. Comput.*, pp.1179–1186.

Kim, J., Hwang, I., Kim, Y-H. and Moon, B-R. (2011) 'Genetic approaches for graph partitioning: a survey', in *Proc. 13th Annual Conf. Genetic Evol. Comput.*, pp.473–480.

Kirkpatrick, S., Vecchi, M.P. et al. (1983) 'Optimization by simmulated annealing', *Science*, Vol. 220, No. 4598, pp.671–680.

Knowles, J.D. and Corne, D.W. (2000) 'Approxmating the nondominated front using the pareto archived evolution strategy', *Evol. Comput.*, Vol. 8, No. 2, pp.149–172.

Lázár, A., Ábel, D. and Vicsek, T. (2010) 'Modularity measure of networks with overlapping communities', *Europhys. Lett.*, Vol. 90, No. 1, pp.18001.

Lancichinetti, A., Fortunato, S. and Kertesz, J. (2008) 'Detecting the overlapping and hierarchical community structure of complex networks', *New J. Phys.*, Vol. 11, p.033015.

Lancichinetti, A., Fortunato, S. and Kertész, J. (2009) 'Detecting the overlapping and hierarchical community structure in complex networks', *New J. Phys.*, Vol. 11, No. 3, p.033015.

Leicht, E.A. and Newman, M.E.J. (2008) 'Community structure in directed networks', *Phys. Rev. Lett.*, Vol. 100, No. 11, p.118703.

Li, X. and Gao, X. (2013) 'A novel community detection algorithm based on clonal selection', *Journal of Computational Information Systems*, Vol. 9, No. 5, pp.1899–1906.

Li, J. and Song, Y. (2013) 'Community detection in complex networks using extended compact genetic algorithm', *Soft Comput.*, pp.1–13.

Li, X. and Yao, X. (2012) 'Cooperatively coevolving particle swarms for large scale optimization', *IEEE Trans. Evol. Comput.*, Vol. 16, No. 2, pp.210–224.

Li, Z., Zhang, S., Wang, R-S., Zhang, X-S. and Chen, L.N. (2008) 'Quantitative function for community detection', *Phys. Rev. E*, Vol. 77, No. 3, p.036109.

Li, S., Chen, Y., Du, H. and Feldman, M.W. (2010) 'A genetic algorithm with local search strategy for improved detection of community structure', *Complexity*, Vol. 15, No. 4, pp.53–60.

Li, Y., Liu, J. and Liu, C. (2013a) 'A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks', *Soft Comput.*, pp.1–20.

Li, Y., Liu, G. and Lao, S-y. (2013b) 'Complex network community detection algorithm based on genetic algorithm', in *Proc. 19th Int. Conf. Ind. Eng. and Eng. Manage.*, pp.257–267.

Lin, C-C., Liu, W-Y. and Deng, D-J. (2013c) 'A genetic algorithm approach for detecting hierarchical and overlapping community structure in dynamic social networks', in *Proc. IEEE Conf. Wirel. Commun. Net.*, pp.4469–4474.

Lipczak, M. and Milios, E. (2009) 'Agglomerative genetic algorithm for clustering in social networks', in *Proc. 11th Annual Conf. Genetic and Evol. Comput.*, pp.1243–1250.

Liu, X., Li, D., Wang, S. and Tao, Z. (2007) 'Effective algorithm for detecting community structure in complex networks based on ga and clustering', in *Proc. Int. Conf. Comput. Sci.*, pp.657–664.

Liu, J., Zhong, W., Abbass, H.A. and Green, D.G. (2010) 'Separated and overlapping community detection in complex networks using multiobjective evolutionary algorithms', in *Proc. IEEE Congr. Evol. Comput.*, pp.1–7.

Liu, C., Liu, J. and Jiang, Z. (2014) 'A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks', *IEEE Trans. Cybern.*, PP(99, pp.1–14.

Ma, L., Gong, M., Liu, J., Cai, Q. and Jiao, L. 'Multi-level learning based memetic algorithm for community detection', *Appl. Soft Comput.*, Vol. 19, pp.121–133.

Massen, C.P. and Doye, J.P.K. (2005) 'Identifying communities within energy landscapes', *Phys. Rev. E*, Vol. 71, No. 4, p.046101.

Mitchell, M. (1998) *An Introduction to Genetic Algorithms*, MIT Press.

Mohankrishna, S., Maheshwari, D., Satyanarayana, P. and Satapathy, S. (2012) 'A comprehensive study of particle swarm based multi-objective optimization', in *Proceedings of the International Conference on Information Systems Design and Intelligent Applications*, pp.689–701.

Muff, S., Rao, F. and Caflisch, A. (2005) 'Local modularity measure for network clusterizations', *Phys. Rev. E*, Vol. 72, No. 5, p.056107.

Neri, F. and Cotta, C. (2012) 'Memetic algorithms and memetic computing optimization: a literature review', *Swarm and Evolutionary Computation*, Vol. 2, pp.1–14.

Neri, F. and Tirronen, V. (2010) 'Recent advances in differential evolution: a survey and experimental analysis', *Artificial Intelligence Review*, Vol. 33, Nos. 1–2, pp.61–106.

Newman, M.E.J. and Girvan, M. (2004) 'Finding and evaluating community structure in networks', *Phys. Rev. E*, Vol. 69, No. 2, p.026113.

Newman, M.E.J. (2003) 'The structure and function of complex networks', *SIAM Review*, Vol. 45, No. 2, pp.167–256.

Newman, M.E.J. (2004) 'Fast algorithm for detecting community structure in networks', *Phys. Rev. E*, Vol. 69, p.066133.

Newman, M.E.J. (2010) *Networks: An Introduction*, Oxford University Press.

Nicosia, V., Mangioni, G., Carchiolo, V. and Malgeri, M. (2009) 'Extending the definition of modularity to directed graphs with overlapping communities', *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2009, No. 03, p.03024.

Ong, Y-S., Lim, M. and Chen, X. (2010) 'Memetic computation – past, present & future [research frontier]', *IEEE Comput. Intell. Magazine*, Vol. 5, No. 2, pp.24–31.

Orman, G.K. and Labatut, V. (2009) 'A comparison of community detection algorithms on artificial networks', in *Discovery Science*, pp.242–256.

Pizzuti, C. (2008) 'GA-Net: a genetic algorithm for community detection in social networks', in *Parallel Problem Solving from Nature (PPSN)*, Vol. 5199, pp.1081–1090.

Pizzuti, C. (2009a) 'A multi-objective genetic algorithm for community detection in networks', in *Proc. 21st Int. Conf. Tools with Artif. Intell.*, pp.379–386.

Pizzuti, C. (2009b) 'Overlapped community detection in complex networks', in *GECCO*, Vol. 9, pp.859–866.

Pizzuti, C. (2012) 'A multiobjective genetic algorithm to find communities in complex networks', *IEEE Trans. Evol. Comput.*, Vol. 16, No. 3, pp.418–430.

Pons, P. and Latapy, M. (2011) 'Post-processing hierarchical community structures: Quality improvements and multi-scale view', *Theoretical Computer Science*, Vol. 412, Nos. 8–10, pp.892–900.

Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. and Parisi, D. (2004) 'Defining and identifying communities in networks', *PNAS*, Vol. 101, No. 9, pp.2658–2663.

Rees, B.S. and Gallagher, K.B. (2012) 'Overlapping community detection using a community optimized graph swarm', *Social Network Analysis and Mining*, Vol. 2, No. 4, pp.405–417.

Reichardt, J. and Bornholdt, S. (2006) 'Statistical mechanics of community detection', *Phys. Rev. E*, Vol. 74, No. 1, p.016110.

Reyes-Sierra, M. and Coello, C.A.C. (2006) 'Multi-objective particle swarm optimizers: a survey of the state-of-the-art', *Int. J. Comput. Intell. Res.*, Vol. 2, No. 3, pp.287–308.

Rosvall, M. and Bergstrom, C.T. (2007) 'An information-theoretic framework for resolving community structure in complex networks', *PNAS*, Vol. 104, No. 18, pp.7327–7331.

Rosvall, M. and Bergstrom, C.T. (2008a) 'Maps of random walks on complex networks reveal community structure', *PNAS*, Vol. 105, No. 4, pp.1118–1123.

Rosvall, M. and Bergstrom, C.T. (2008b) 'Maps of random walks on complex networks reveal community structure', *PNAS*, Vol. 105, No. 4, pp.1118–1123.

Shang, R., Bai, J., Jiao, L. and Jin, C. (2013) 'Community detection based on modularity and an improved genetic algorithm', *Phys. A*, Vol. 392, No. 5, pp.1215–1231.

Shelokar, P., Quirin, A. and Cordón, Ó (2014) 'Three-objective subgraph mining using multiobjective evolutionary programming', *Journal of Computer and System Sciences*, Vol. 80, No. 1, pp.16–26.

Shen, H., Cheng, X., Cai, K. and Hu, M-B. (2009) 'Detect overlapping and hierarchical community structure in networks', *Phys. A*, Vol. 388, No. 8, pp.1706–1712.

Shi, Z., Liu, Y. and Liang, J. (2009a) 'Pso-based community detection in complex networks', in *Proc. 2nd Int. Sympos. Knowl. Acquisition Model.*, Vol. 3, pp.114–119.

Shi, C., Wang, Y., Wu, B. and Zhong, C. (2009b) 'A new genetic algorithm for community detection', in *Complex Sciences*, pp.1298–1309.

Shi, C., Yu, P.S., Cai, Y., Yan, Z. and Wu, B. (2011a) 'On selection of objective functions in multi-objective community detection', in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, pp.2301–2304.

Shi, C., Yan, Z., Pan, X., Cai, Y. and Wu, B. (2011b) 'Multi-objective decisionmaking in the detection of comprehensive community structures', in *Proc. IEEE Congr. Evol. Comput.*, pp.1489–1495.

Shi, C., Yan, Z., Pan, X., Cai, Y. and Wu, B. (2012) 'Multi-objective community detection in complex networks', *Appl. Soft Comput.*, Vol. 12, No. 2, pp.850–859.

Shi, C., Yu, P.S., Yan, Z., Huang, Y. and Wang, B. (2013) 'Comparison and selection of objective functions in multiobjective community detection', *Comput. Intell.*

Stadler, W. (1979) 'A survey of multicriteria optimization or the vector maximum problem, Part I: 1776–1960', *Journal of Optimization Theory and Applications*, Vol. 29, No. 1, pp.1–52.

Tasgin, M., Herdagdelen, A. and Bingol, H. (2007) 'Community detection in complex networks using genetic algorithms', arXiv preprint arXiv:0711.0491.

Wang, Y. and Li, B. (2010) 'Two-stage based ensemble optimization for large-scale global optimization', in *Proc. IEEE Congr. Evol. Comput.*, pp.1–8.

Watts, D.J. and Strogatz, S.H. (1998) 'Collective dynamics of 'small-world' networks', *Nature*, Vol. 393, No. 6684, pp.440–442.

Wolpert, D.H. and Macready, W.G. (1997) 'No free lunch theorems for optimization', *IEEE Trans. Evol. Comput.*, Vol. 1, No. 1, pp.67–82.

Xiaodong, D., Cunrui, W., Xiangdong, L. and Yanping, L. (2008) 'Web community detection model using particle swarm optimization', in *Proc. IEEE Congr. Evol. Comput.*, pp.1074–1079.

Xie, J., Kelley, S. and Szymanski, B.K. (2011) 'Overlapping community detection in networks: the state of the art and comparative study', arXiv preprint arXiv, pp.1110.5813.

Xie, J., Kelley, S. and Szymanski, B.K. (2013) 'Overlapping community detection in networks: The state-of-the-art and comparative study', *ACM Computing Surveys (CSUR)*, Vol. 45, No. 4, pp.43.

Yang, Z.Y., Tang, K. and Yao, X. (2008) 'Large scale evolutionary optimization using cooperative coevolution', *Inf. Sci.*, Vol. 178, No. 15, pp.2985–2999.

Yang, X-S. (2013) 'Multiobjective firefly algorithm for continuous optimization', *Engineering with Computers*, Vol. 29, No. 2, pp.175–184.

Zhang, Q. and Li, H. (2007) 'MOEA/D: a multiobjective evolutionary algorithm based on decomposition', *IEEE Trans. Evol. Comput.*, Vol. 11, No. 6, pp.712–731.

Zhang, S., Wang, R-S. and Zhang, X-S. (2007) 'Identification of overlapping community structure in complex networks using fuzzy c-means clustering', *Phys. A*, Vol. 374, No. 1, pp.483–490.

Zhao, S.Z., Liang, J.J., Suganthan, P.N. and Tasgetiren, M.F. (2008) 'Dynamic multi-swarm particle swarm optimizer with local search for large scale global optimization', in *Proc. IEEE Congr. Evol. Comput.*, pp.3845–3852.

Zhou, A., Qu, B-Y., Li, H., Zhao, S-Z., Suganthan, P.N. and Zhang, Q. (2011) 'Multiobjective evolutionary algorithms: a survey of the state of the art', *Swarm and Evol. Comput.*, Vol. 1, No. 1, pp.32–49.

Zitzler, E., Laumanns, M. and Thiele, L. (2002) 'SPEA2: improving the strength pareto evolutionary algorithm', in *Proc. Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems*, pp.95–100.

Notes

1 Without loss of generality, we will assume only minimisation problems.