European Journal of Operational Research 254 (2016) 780–800

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

Iterated maxima search for the maximally diverse grouping problem

Xiangjing Lai[a], Jin-Kao Hao[a,b,*]

[a] LERIA, Université d'Angers, 2 Bd Lavoisier, Angers 49045, France
[b] Institut Universitaire de France, Paris, France

ARTICLE INFO

Article history:
Received 10 May 2015
Accepted 10 May 2016
Available online 18 May 2016

Keywords:
Graph grouping and clustering problems
Iterated search
Heuristics

ABSTRACT

The maximally diverse grouping problem (MDGP) is to partition the vertices of an edge-weighted and undirected complete graph into m groups such that the total weight of the groups is maximized subject to some group size constraints. MDGP is a NP-hard combinatorial problem with a number of relevant applications. In this paper, we present an innovative heuristic algorithm called iterated maxima search (IMS) algorithm for solving MDGP. The proposed approach employs a maxima search procedure that integrates organically an efficient local optimization method and a weak perturbation operator to reinforce the intensification of the search and a strong perturbation operator to diversify the search. Extensive experiments on five sets of 500 MDGP benchmark instances of the literature show that IMS competes favorably with the state-of-the-art algorithms. We provide additional experiments to shed light on the rationality of the proposed algorithm and investigate the role of the key ingredients.

1. Introduction

Given an edge-weighted and undirected complete graph $G = (V, E, D)$, where $V = \{1, 2, \ldots, n\}$ is the set of n vertices, $E = \{\{i, j\} : i, j \in V, i \neq j\}$ is the set of $n \times (n-1)/2$ edges, and $D = \{d_{ij} \geq 0 : \{i, j\} \in E\}$ represents the set of non-negative edge weights, the maximally diverse grouping problem (MDGP for short) is to partition the vertex set V into m disjoint subsets or groups such that the size of each group g lies in a given interval $[a_g, b_g]$ $(g = 1, 2, \ldots, m)$ while maximizing the sum of the edge weights of the m groups. Here, a vertex $v \in V$ is usually called an element, an edge weight $d_{ij} \in D$ is called the diversity between elements i and j, while a_g and b_g are respectively called the lower and upper limits (or bounds) of the size of group g.

MDGP can be expressed as a quadratic integer program with binary variables x_{ig} that take the value of 1 if element i is in group g and 0 otherwise (Gallego, Laguna, Martí, & Duarte, 2013; Rodriguez, Lozano, García-Martínez, & González-Barrera, 2013).

$$\text{Maximize} \quad \sum_{g=1}^{m}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{ij}x_{ig}x_{jg} \tag{1}$$

$$\text{Subject to} \quad \sum_{g=1}^{m} x_{ig} = 1, i = 1, 2, \ldots, n \tag{2}$$

$$a_g \leq \sum_{i=1}^{n} x_{ig} \leq b_g, g = 1, 2, \ldots, m \tag{3}$$

$$x_{ig} \in \{0, 1\}, i = 1, 2, \ldots, n; g = 1, 2, \ldots, m \tag{4}$$

where the set of constraints (2) guarantees that each element is located in exactly one group and the set of constraints (3) forces the size of group g is at least a_g and at most b_g.

MDGP belongs to the category of vertex-capacitated clustering problems which are a type of extensively studied combinatorial search problems and can further be divided into the max-clustering problem and min-clustering problem (Ferreira, Martin, Souza, Weismantel, & Wolsey, 1996, 1998; Johnson, Mehrotra, & Nemhauser, 1993; Özsoy & Labbé, 2010; Wang, Alidaee, Glover, & Kochenberger, 2006). In short, the max-clustering (min-clustering) problem is to partition the vertices of an undirect graph $G = (V, E)$ with edge and vertex weights into m mutually disjoint subsets (groups or clusters) such that the sum of the vertex weights of the subsets is bounded from below by a and from above by b while maximizing (minimizing) the sum of the weights of the edges inside the subsets (Johnson et al., 1993).

In addition to its theoretical signification as a typical NP-hard problem, MDGP has a variety of real-world applications, like assignment of students to groups (Johnes, 2015; Krass & Ovchinnikov, 2010; Yeoh & Nor, 2011), creation of peer review groups

* Corresponding author at: LERIA, Université d'Angers, 2 Bd Lavoisier, 49045 Angers, France. Tel.: +33241735076; fax: +33241735073.

E-mail addresses: laixiangjing@gmail.com (X. Lai), hao@info.univ-angers.fr, jin-kao.hao@univ-angers.fr (J.-K. Hao).

http://dx.doi.org/10.1016/j.ejor.2016.05.018
0377-2217/

(Chen, Fan, Ma, & Zeng, 2011), VLSI design (Weitz & Lakshminarayan, 1997), storage allocation of large programs onto paged memory, and creation of diverse groups in companies so that people from different backgrounds work together (Bhadury, Mighty, & Damar, 2000). For a review of possible applications of MDGP, the readers are referred to recent papers like Gallego et al. (2013), Palubeckis, Ostreika, and Rubliauskas (2015), Rodriguez et al. (2013), Urošević (2014).

Given the practical importance and high computational complexity of MDGP, a number of exact and heuristic algorithms have been proposed in the literature. One of the most representative exact algorithm for MDGP is the column generation approach presented in Johnson et al. (1993). Nevertheless, local search or evolutionary heuristics remain the dominant approach in the literature to find high-quality sub-optimal solutions for large graphs. Examples of local search heuristics includes multistart algorithm (Arani & Lofti, 1989), Weitz-Jelassi (WJ) algorithm (Weitz & Jelassi, 1992), Lotfi-Cerveny-Weitz (LCW) algorithm (Weitz & Lakshminarayan, 1998), T-LCW method based on LCW and tabu search (Gallego et al., 2013), tabu search with strategic oscillation (TS-SO) (Gallego et al., 2013), multistart simulated annealing (MSA) (Palubeckis, Karčiauskas, & Riškus, 2011), variable neighborhood search (VNS) (Palubeckis et al., 2011), general variable neighborhood search (GVNS) (Urošević, 2014), skewed general variable neighborhood search (SGVNS) (Brimberg, Mladenović, & Urošević, 2015), iterated tabu search (ITS) (Palubeckis et al., 2015) and several graph theoretical heuristics (Feo & Khellaf, 1990). The population-based evolutionary approach includes hybrid genetic algorithm (LSGA) (Fan, Chen, Ma, & Zeng, 2010), hybrid grouping genetic algorithm (Chen et al., 2011), hybrid steady-state genetic algorithm (HGA) (Palubeckis et al., 2011), artificial bee colony (ABC) algorithm (Rodriguez et al., 2013), and constructive genetic algorithm (Lorena & Antonio, 2001). According to the computational results reported on the MDGP benchmarks, the heuristics T-LCW, TS-SO, HGA, MSA, VNS, ABC, GVNS, ITS, and SGVNS achieved high performances at the time they were published.

In this paper, we propose an effective heuristic called the iterated maxima search (IMS) algorithm for solving MDGP. IMS follows and extends the iterated local search framework (Lourenco, Martin, & Stützle, 2003). Though IMS shares ideas from breakout local search (Benlic & Hao, 2013a; 2013b) and three-phase local search (Fu & Hao, 2015), it distinguishes itself from these approaches by three specific features: its local search procedure (to improve the incumbent solution), the maxima search scheme (to locate other local optima within a limited region of the search space) and its perturbation operator (to modify greatly the incumbent solution). In addition, IMS employs a randomized procedure for initial solution generation. When the proposed algorithm was assessed on five sets of 500 benchmark instances ($120 \le n \le 3000$) commonly used in the literature, the computational results showed that the algorithm achieves highly competitive results compared to the state-of-the-art algorithms especially on the large-sized instances.

In Section 2, we describe the components of the proposed algorithm. Section 3 is dedicated to extensive computational assessments based on the commonly used benchmarks with respect to the top performing algorithms from the literature. In Section 4, the important components of the proposed algorithm are analyzed and discussed. Conclusions are provided in the last Section.

2. Iterated maxima search algorithm for MDGP

The proposed iterated maxima search (IMS) algorithm can be considered as an extended iterated local search algorithm (Lourenco et al., 2003) and shares ideas from breakout local search (Benlic & Hao, 2013a; 2013b) and three-phase local search (Fu & Hao, 2015) (see Section 2.7 for more discussions). IMS is composed of four basic procedures: solution initialization, local search, weak perturbation, and strong perturbation. The basic idea of the IMS algorithm is to provide the search algorithm with a desirable tradeoff between intensification and diversification. This is achieved by iterating the maxima search procedure (local search combined with weak perturbation to explore a limited region around a locally optimal solution) followed by the strong perturbation procedure (to displace the search to a distant region by changing strongly the attained local optimum).

Algorithm 1: Main framework of iterated maxima search method for MDGP.

Input: Instance I, the depth of maxima search (α), the strength of strong perturbation (η_s), the cutoff time (t_{max})
Output: The best solution s^* found

begin
  $s \leftarrow InitialSolution(I)$ /* Sections 2.3 and 2.4 */
  $s^* \leftarrow s$
  while $Time() \le t_{max}$ **do**
    $s \leftarrow MaximaSearch(s, \alpha)$ /* Section 2.5 */
    if $f(s) > f(s^*)$ **then**
      $s^* \leftarrow s$
    end
    $s \leftarrow StrongPerturbation(s, \eta_s)$ /* Section 2.6 */
  end
  return s^*
end

2.1. General procedure

The IMS algorithm (Algorithm 1) starts from a feasible initial solution (Section 2.2) that is obtained with a randomized construction procedure (Section 2.3). Then it repeats a number of iterations until a cutoff time is reached. At each iteration, the incumbent solution s is improved by the maxima search procedure (Sections 2.4 and 2.5) and then perturbed by the strong perturbation procedure (Sections 2.6). The best solution found (s^*) is updated whenever it is needed and finally returned as the output at the end of the IMS algorithm. In the rest of this section, we describe the different components of the proposed algorithm.

2.2. Search space, fitness function and solution representation

For a given MDGP instance $G = (V, E)$ with its edge diversity matrix $D = [d_{ij}]_{n\times n}$ and the number m of groups, the search space Ω explored by the IMS algorithm contains all partitions of the elements of V into m groups such that the size of each group lies between its lower and upper limits. In other words, our IMS algorithm visits only feasible solutions.

Formally, let $P : V \to \{1, \dots, m\}$ be a partition function of the n elements of V to m groups. For each group $g \in \{1, \dots, m\}$ with lower and upper limits a_g and b_g, let $P_g = \{i \in V : P(i) = g\}$ (i.e., P_g is the set of elements of group g). Then the search space is given by:

$$\Omega = \{P : \forall g \in \{1, \dots, m\}, a_g \le |P_g| \le b_g\}.$$

For any candidate solution $s = \{P_1, P_2, \dots, P_m\}$ of Ω, its quality or fitness is given by the objective value $f(s)$:

$$f(s) = \sum_{g=1}^{m} \sum_{i,j\in P_g, i<j} d_{ij} \quad (5)$$