European Journal of Operational Research 254 (2016) 769–779

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

On solving manufacturing cell formation via Biclister Editing

Rian G. S. Pinheiro[a,b], Ivan C. Martins[a], Fábio Protti[a,*], Luiz S. Ochi[a], Luidi G. Simonetti[a], Anand Subramanian[c]

[a] Fluminense Federal University Niterói, RJ, Brazil
[b] Federal Rural University of Pernambuco Garanhuns, PE, Brazil
[c] Federal University of Paraíba João Pessoa, PB, Brazil

ARTICLE INFO

Article history:
Received 11 July 2014
Accepted 6 May 2016
Available online 14 May 2016

Keywords:
Combinatorial optimization
Biclusterization
Graph partitioning
Manufacturing cell formation

ABSTRACT

This work investigates the Bicluster Graph Editing Problem (BGEP) and how it can be applied to solve the Manufacturing Cell Formation Problem (MCFP). We develop an exact method for the BGEP with a new separation algorithm. We also describe a new preprocessing procedure for the BGEP derived from theoretical results on vertex distances in the input graph. Computational experiments performed on randomly generated instances with various levels of difficulty show that our separation algorithm accelerates the convergence speed, and our preprocessing procedure is effective for low density instances. Another contribution of this work is to take advantage of the fact that the BGEP and the MCFP share the same solution space. This leads to the proposal of two new exact approaches for the MCFP that are based on mathematical formulations for the BGEP. Both approaches use the grouping efficacy measure as the objective function. Up to the authors' knowledge, these are the first exact methods that employ such a measure to optimally solve instances of the MCFP. The first approach is based on a new ILP formulation for the MCFP, and the second consists of iteratively running several calls to a parameterized version of the BGEP. Computational experiments performed on instances of the MCFP found in the literature show that our exact methods for the MCFP are able to prove several previously unknown optima.

1. Introduction

The Bicluster Graph Editing Problem (BGEP) is described as follows: given a bipartite graph $G = (U, V, E)$, where U and V are nonempty stable sets (called the "parts" of the bipartition (U, V)) and E is a set of unweighted edges linking vertices in U and a vertices in V, the goal is to transform G into a disjoint union of complete bipartite graphs (or *bicliques*) by performing a minimum number of *edge editing operations*. Each edge editing operation consists of either removing an existing edge in E or adding to E a new edge between a vertex in U and a vertex in V. Note that, in a feasible solution, each edge and each vertex must belong to exactly one biclique.

A *bicluster* is a subgraph of G isomorphic to a biclique. In some graph theoretical models used in Computational Biology and other areas, the existence of biclusters indicates a high degree of similarity between the data (vertices). In particular, a perfectly clustered bipartite graph is called a *bicluster graph*, i.e., a bipartite graph in which each of its connected components is a bicluster. Hence, we can alternatively define the goal of the BGEP, as stated by Amit (2004), as follows: "find a minimum number of edge editing operations in order to transform an input bipartite graph into a bicluster graph".

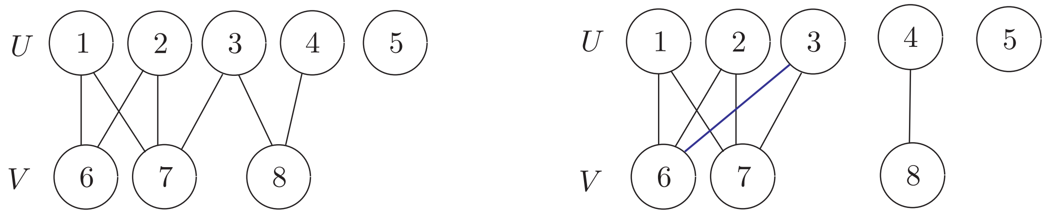
Fig. 1 shows an example where adding an edge between vertices 3, 6 and deleting the edge between vertices 3, 8 transforms G into a bicluster graph. Note that this does not correspond to an optimal solution, since G can also be transformed into a bicluster graph by simply removing the edge between 3 and 7. We remark that a single vertex is considered as a bicluster (e.g., vertex 5 in Fig. 1).

The Cluster Graph Editing Problem (CGEP) is a clustering problem similar to the BGEP. The CGEP was first studied by Gupta and Palit (1979) and its goal is to transform a (not necessarily bipartite) graph G into a disjoint union of complete graphs (cliques). The CGEP and the BGEP can also be viewed as important examples of partition problems in graphs.

The concept of biclustering was introduced in the mid-70s by Hartigan (1975). In 2000, Cheng and Church (2000) use biclustering within the context of Computational Biology. Since then, algorithms for biclustering have been proposed and used in various applications, such as multicast network design

* Corresponding author. Tel.: +55 2126295669; fax: +55 2126295669.
E-mail addresses: rgpinheiro@ic.uff.br (R.G.S. Pinheiro), imartins@ic.uff.br (I.C. Martins), fabio@ic.uff.br (F. Protti), satoru@ic.uff.br (L.S. Ochi), luidi@ic.uff.br (L.G. Simonetti), anand@ct.ufpb.br (A. Subramanian).

http://dx.doi.org/10.1016/j.ejor.2016.05.010
0377-2217/



(a) Instance. (b) Solution.

Fig. 1. BGEP example.

$$
\begin{array}{c|ccccc|}
 & M_1 & M_2 & M_3 & M_4 & M_5 \\
\hline
P_1 & 0 & 1 & 1 & 0 & 1 \\
P_2 & 1 & 0 & 0 & 1 & 0 \\
P_3 & 0 & 1 & 1 & 0 & 0 \\
P_4 & 1 & 0 & 0 & 1 & 0 \\
P_5 & 1 & 0 & 0 & 0 & 1 \\
P_6 & 1 & 0 & 1 & 1 & 0 \\
P_7 & 0 & 0 & 1 & 0 & 1 \\
\hline
\end{array}
\quad \Rightarrow \quad
\begin{array}{c|ccc|cc|}
 & M_2 & M_3 & M_5 & M_1 & M_4 \\
\cline{2-4}
P_1 & 1 & 1 & 1 & 0 & 0 \\
P_3 & 1 & 1 & 0 & 0 & 0 \\
P_7 & 0 & 1 & 1 & 0 & 0 \\
\hline
P_2 & 0 & 0 & 0 & 1 & 1 \\
P_4 & 0 & 0 & 0 & 1 & 1 \\
P_5 & 0 & 0 & 1 & 1 & 0 \\
P_6 & 0 & 1 & 0 & 1 & 1 \\
\cline{5-6}
\end{array}
$$

Fig. 2. MCFP example. The matrix on the right represents a feasible solution, viewed as a permutation of rows/columns of the input matrix on the left. Cells are shown on the right as submatrices delimited by rectangles.

(Faure, Chretienne, Gourdin, & Sourd, 2007) and analysis of biological data (Abdullah & Hussain, 2006; Bisson & Hussain, 2008).

In Biology, concepts such as co-clustering, two-way clustering, among others, are often used in the literature to refer to the same problem. Matrices are used instead of graphs to represent relationships between genes and characteristics, and their rows/columns represent graph bipartitions; in this case, the goal is to find significant submatrices having certain patterns. The BGEP can be used to solve any problem whose goal is to obtain a biclusterization with *exclusive* rows and columns, i.e., each gene (characteristic) must be associated with only one submatrix.

Amit (2004) proved the $\mathcal{NP}$-hardness of the BGEP via a polynomial reduction from the 3-Exact 3-Cover Problem; in the same work, a binary integer programming formulation and an 11-approximation algorithm based on the relaxation of a linear program are described. The work by Protti, Dantas da Silva, and Szwarcfiter (2006) describes an algorithm for the parameterized version of the BGEP that uses a strategy based on modular decomposition techniques. Guo, Hffner, Komusiewicz, and Zhang (2008) developed a randomized 4-approximation algorithm for the BGEP. More recently, Sousa Filho, dos Anjos F. Cabral, Ochi, and Protti (2012) proposed a GRASP-based heuristic for the BGEP.

A new application of the BGEP, introduced in this work, is related to the Manufacturing Cell Formation Problem (MCFP). The input of the MCFP is given as a binary product-machine matrix M such that each entry $M(i, j)$ has value 1 if product i is manufactured by machine j, and 0 otherwise. Any feasible solution of the MCFP consists of a product-machine cell assignment, i.e., a collection of product-machine cells where every product (or machine) is allocated to exactly one cell. Hence, for each cell C, machines allocated to C are exclusively dedicated to manufacture products also allocated to C. In an ideal solution of the MCFP, for each cell C there must be a high similarity between products and machines allocated to it. Fig. 2 shows an example of the MCFP solved as a block diagonalization problem. Note that a solution of the MCFP can be viewed as a permutation of rows/columns of the input matrix yielding a new matrix M' where diagonal block submatrices represent cells. Of course, the permutation is not needed to obtain a solution, it only helps to make it more visual. In the figure, products P_1, P_3, P_7 and machines M_2, M_3, M_5 are gathered to form a cell, while the remaining products/machines form another cell.

Among several measures of performance used as objective functions for the MCFP, Sarker and Khan (2001) evaluated the quality of a solution using different measures reported in the literature as: *Grouping Efficiency* (Chandrasekharan & Rajagopalan, 1986); *Grouping Efficacy* (Kumar & Chandrasekharan, 1990); *Grouping Capability Index* (Hsu, 1990); and *Grouping Measure* (Miltenburg & Zhang, 1991). The *grouping efficacy* μ is considered in the literature as a standard measure to represent the quality of solutions. It is defined as:

$$\mu = \frac{N_1 - N_1^{out}}{N_1 + N_0^{in}} , \tag{1}$$

where N_1 is the total number of 1's in the input matrix, and N_1^{out} (N_0^{in}) is the total number of 1's outside (respectively, 0's inside) diagonal blocks in the solution matrix. In Fig. 2, $\mu = \frac{16-2}{16+3} = 0.7368$.

Some works define a minimum value for the size of the cells. For instance, in (Chandrasekharan & Rajagopalan, 1987; Gonçalves & Resende, 2004; Srinivasan & Narendran, 1991), cells with less than two products or machines are not allowed; such cells are called *singletons*. However, there is no consensus with respect to the size of the cells. Other studies do not consider any size constraint, allowing the existence of empty cells, such as the work by Pailla, Trindade, Parada, and Ochi (2010). An example of a solution with an empty cell is shown in Fig. 3.

In the present work, we deal with two versions of the MCFP found in the literature:

1. unrestricted version, allowing singletons and empty cells;
2. with cell size constraints (the minimum size of each cell is 2×2).

Cellular manufacturing is an application of the Group Technology concept (Goldengorin, Krushinsky, & Pardalos, 2013). The goal is to identify and cluster similar parts in order to optimize the manufacturing process. Such a concept was originally introduced by Flanders (1924) and formally described by Mitrofanov (1966) in 1966. In the early 70s, Burbidge (1971) presented one of the first techniques for creating a system of cellular manufacturing. Since this work, several approaches have been proposed to the MCFP,