International Journal of Production Research

ISSN: 0020-7543 (Print) 1366-588X (Online) Journal homepage: https://www.tandfonline.com/loi/tprs20

NP-completeness of cell formation problem with grouping efficacy objective

Mikhail V. Batsyn, Ekaterina K. Batsyna & Ilya S. Bychkov

International Journal of Production Research, 2019
https://doi.org/10.1080/00207543.2019.1668072

NP-completeness of cell formation problem with grouping efficacy objective

Mikhail V. Batsyn [a*], Ekaterina K. Batsyna [b] and Ilya S. Bychkov [a]

[a] *Laboratory of Algorithms and Technologies for Network Analysis, National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation;* [b] *Department of Applied Mathematics and Informatics, National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation*

(*Received 31 January 2019; accepted 6 September 2019*)

In the current paper we provide a proof of NP-completeness for the Cell Formation Problem (CFP) with the fractional grouping efficacy objective function. First the CFP with a linear objective function is considered. Following the ideas of Pinheiro et al. (2016) we show that it is equivalent to the Bicluster Graph Editing Problem (BGEP), which is known to be NP-complete due to the reduction from the 3-Exact 3-Cover Problem – 3E3CP (Amit, 2004). Then we suggest a polynomial reduction of the CFP with the linear objective to the CFP with the grouping efficacy objective. It proves the NP-completeness of this fractional CFP formulation. Along with the NP-status our paper presents important connections of the CFP with the BGEP and 3E3CP. Such connections could be used for "transferring" of known theoretical properties, efficient algorithms, polynomial cases, and other features of well-studied graph editing and exact covering problems to the CFP.

Keywords: cell formation problem; bicluster graph editing problem; grouping efficacy; np-complete

1. Introduction

The Cell Formation Problem (CFP) consists in optimal grouping of machines together with parts processed on them into manufacturing cells. The goal of such a bi-clustering (clustering of both machines and parts) is to minimise the inter-cell movement of parts between different cells during the manufacturing process and to maximise the loading of machines with parts processing inside their cells. The input to the classical CFP problem is given by a binary machine-part matrix defining for every machine what parts are processed on it. In terms of input matrix the objective of the CFP is to partition rows (machines) and columns (parts) of the input matrix into rectangular cells minimizing the number of ones outside cells, called exceptions, and the number of zeroes inside cells, called voids. Exceptions represent inter-cell movements of parts and voids reflect underloading of machines in their production cells (Won and Logendran 2015). An example of the input matrix is shown in Table 1 and a feasible solution for this instance is shown in Table 2. There are two voids in this solution, because machine 4 (row 4) does not process part 4 (row 4) in cell 1, and machine 3 (row 3) does not process part 3 in cell 2. And there are ten exceptions due to the operations performed by machines on the parts from other cells. For example, machine 4 in cell 1 has to process parts 3 and 5 from cell 2 and part 7 from cell 3.

The CFP is the main problem of the Group Technology (GT) concept in manufacturing (Mitrofanov 1959; Burbidge 1975). As noted by Gunasekaran et al. (1994) the GT concept plays an important role in many engineering practices and technologies including Just-In-Time manufacturing – JIT (Golhar and Stamm 1991), Design For eXcellence – DFX (Battaïa et al. 2018), X-Manufacturing Systems – XMS (Dolgui and Proth. 2010; Battaïa, Dolgui, and Guschinsky 2017). The research devoted to real-life CFP formulations is very active (Alhourani 2016; Eguia et al. 2017; Rezaei-Malek et al. 2017). Along with CFP it also covers the closely related GT problems: cell part scheduling problem (Liu et al. 2016; Zeng, Tang, and Fan 2019) and machine operator assignment problem (Norman et al. 2002; Egilmez, Erenay, and Suer 2014; Kuo, Chen, and Wang 2018).

The main issue with real-life CFP is that there are no efficient exact approaches or approximate methods with guaranteed accuracy in literature for such problems. Most of the papers present heuristic algorithms or inefficient mathematical programming models. The research in the field of efficient exact methods currently deals with classical formulations of the CFP (Elbenani and Ferland 2012; Bychkov, Batsyn, and Pardalos 2014; Žilinskas, Goldengorin, and Pardalos 2015; Brusco 2015; Pinheiro et al. 2016; Utkina, Batsyn, and Batsyna 2018; Bychkov and Batsyn 2018). At the same time exact approaches are

*Corresponding author. Email: mbatsyn@hse.ru

Table 1. CFP instance.

	1	2	3	4	5	6	7
1	1	0	0	1	0	1	1
2	1	1	0	1	0	0	0
3	1	0	0	1	1	0	1
4	1	1	1	0	1	0	1
5	1	1	1	1	1	0	0

Table 2. CFP solution.

	1	2	4	3	5	6	7
2	1	1	1	0	0	0	0
4	1	1	0	1	1	0	1
5	1	1	1	1	1	0	0
3	1	0	1	0	1	0	1
1	1	0	1	0	0	1	1

important for any optimisation problem since only such methods are able to provide guarantees of solutions optimality to decision-makers. This research also requires theoretical study of the CFP starting from its classical formulations.

A number of papers on the classical CFP are devoted to its simplest formulation, called Machine Partitioning Problem (MPP), in which only machines are clustered into cells and the objective is computed as an explicit function from this partition and machine-part matrix (Kusiak, Boe, and Cheng 1993; Spiliopoulos and Sofianopoulou 1998; Arkat, Abdollahzadeh, and Ghahve 2012). Though we are not aware of the proof of NP-completeness for the MPP, we believe it exists in literature. It is probably present in the PhD thesis of Ballakur (1985), judging by the references to this work. Unfortunately we have failed to find it in electronic databases. Besides Ghosh et al. (1996) states that the NP-hardness of the MPP can be proved 'by a straightforward reduction of the clustering problem (Garey and Johnson. 1979)' to it.

The CFP problem becomes much harder when we want to cluster machines and parts together into biclusters. In spite of the fact that most papers in last decades consider the CFP in its biclustering formulation, there are no papers providing the proof of its NP-status to the best of our knowledge. There is a big number of papers, where authors just write that the problem is NP-hard (Mak, Wong, and Wang 2000; Goncalves and Resende 2004; Chan et al. 2008). Other authors including Tunnukij and Hicks (2009), Elbenani and Ferland (2012) state that the CFP is NP-hard citing the paper of Dimopoulos and Zalzala (2000). But Dimopoulos and Zalzala (2000) only mention that 'the cell-formation problem is a difficult optimization problem'.

Many papers including James, Brown, and Keeling (2007), Chung, Wu, and Chang (2011), Paydar and Saidi-Mehrabad (2013), Solimanpur, Saeedi, and Mahdavi (2010), Utkina, Batsyn, and Batsyna (2016) refer to Ballakur and Steudel (1987) when writing about the NP-hardness of the CFP. However Ballakur and Steudel (1987) present a heuristic for the CFP with different objective functions and do not state anything about the NP status of these CFP formulations. Finally there are some papers citing Ballakur (1985) PhD thesis where several CFP formulations are considered. However this paper is not available in any electronic publication databases. According to the existing references to this thesis and other papers of Ballakur we can only conclude that he considers the machine partitioning and machine-part partitioning problems with some objective functions, but not with the grouping efficacy function introduced later by Kumar and Chandrasekharan (1990). At the same time the grouping efficacy is currently widely accepted and considered as the best function successfully joining the both objectives of inter-cell part movement minimisation and intra-cell machine loading maximization.

In the current paper we provide a proof of NP-completeness for the CFP problem with the fractional grouping efficacy objective. For this purpose we first consider the CFP with the linear objective minimizing the total number of exceptions and voids. Following the ideas of Pinheiro et al. (2016) we show that it is equivalent to the Bicluster Graph Editing Problem (BGEP), which is known to be NP-complete. Amit (2004) proved it by the polynomial reduction of the 3-Exact 3-Cover Problem (3E3CP) to the BGEP. Then we suggest a reduction of the CFP problem with the linear objective function to the CFP with the grouping efficacy objective.

Our reduction and the reduction of Amit (2004) show that the CFP, BGEP, and 3E3CP are closely connected with each other. This means that efficient algorithms for the BGEP and 3E3CP could potentially be adapted for the CFP. While exact cover problems (Bonnet, Paschos, and Sikora 2016; Nishino et al. 2017) are not so popular, the BGEP is a well-studied problem due to its application in the analysis of gene expression data in biomedicine and other applications (Busygin, Prokopyev, and Pardalos 2008). The algorithms for the BGEP (also called Bicluster Editing and Bipartite Correlation Clustering problem) include approximation algorithms (Amit 2004; Guo et al. 2008; Ailon et al. 2011), fixed-parameter tractable algorithms (Guo et al. 2008; Protti, Dantas da Silva, and Luiz Szwarcfiter 2009; Sun, Guo, and Baumbach 2014; Drange et al. 2015), and heuristic algorithms (Sun et al. 2014; De Sousa Filho et al. 2017). Due to the close relations between the CFP and the BGEP such algorithms are of interest for future CFP research.

The paper is organised as follows. In the next section we describe the CFP problem and give its mathematical programming model with linear and fractional objective functions. In Section 3 the proof of NP-completeness for the CFP with the linear objective is provided. Section 4 is the main section of our paper containing the NP-completeness proof for the CFP with the grouping efficacy objective. Section 5 presents the conclusion of the paper together with future work and perspectives.

2. Problem formulation

In the CFP we are given m machines, p parts processed on these machines, and $m \times p$ Boolean matrix A in which $a_{ij} = 1$, if machine i processes part j during the production process, and $a_{ij} = 0$ otherwise. We should cluster both machines and parts into biclusters, called cells, so that for every part we minimise simultaneously the number of processing operations of this part on machines from other cells and the number of machines from the same cell which do not process this part. Thus we minimise the movement of parts to other cells (inter-cell operations) and maximise the loading of machines with processing operations inside cells (intra-cell operations) during the production process.

In other words we need to choose machine-part cells in matrix A, such that the number of ones outside these cells (called exceptions) is minimal possible and at the same time the number of zeroes inside these cells (called voids) is also minimal possible. For example in the solution shown in Table 2 machines 2, 4, 5 together with parts 1, 2, 4 are joined into cell 1; machine 3 with parts 3, 5 – into cell 2; and machine 1 with parts 6, 7 – into cell 3. In terms of machine-part matrix we need to transform it to a matrix maximally close to a block-diagonal one by moving its rows and columns.

One of the simplest objective functions for the CFP is the following linear function:

$$f_1 = e + v \rightarrow \min. \tag{1}$$

Here e is the number of exceptions (ones outside cells) and v is the number of voids (zeroes inside cells). However this is not the best way to combine the two contradictory CFP objectives of minimizing exceptions and minimizing voids.

The objective function which provides a good combination of these two goals and is widely accepted in literature is the grouping efficacy suggested by Kumar and Chandrasekharan (1990):

$$f = \frac{n_1 - e}{n_1 + v} \rightarrow \max. \tag{2}$$

Here n_1 is the number of ones in the input matrix. This function has several properties which make it a good measure of the CFP solution quality. First, it belongs to segment $[0, 1]$ and it is equal to zero only for the worst solutions with no ones inside cells and equal to one only for ideal solutions with all ones and no zeroes inside cells. Second, the grouping efficacy provides a good balance when taking into account both exceptions and voids in a solution. Third, this function guarantees that a solution with less exceptions and voids will always have higher efficacy. Finally, according to the computational results on real-life instances it is the best in measuring how close a solution to a block-diagonal matrix (an ideal solution). That is why the grouping efficacy is the most widely used objective function in CFP papers since it has been introduced.

Below we present a straightforward fractional programming model for the CFP (Bychkov, Batsyn, and Pardalos 2014; Utkina, Batsyn, and Batsyna 2018). Since the number of cells cannot be greater than the number of machines and the number of parts, then the maximal possible number of cells is equal to $\min(m, p)$. We denote this value as $c = \min(m, p)$.

Decision variables:

$$x_{ik} = \begin{cases} 1 & \text{if machine } i \text{ is assigned to cell } k \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$y_{jk} = \begin{cases} 1 & \text{if part } j \text{ is assigned to cell } k \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$e = n_1 - \sum_{k=1}^{c}\sum_{i=1}^{m}\sum_{j=1}^{p} a_{ij}x_{ik}y_{jk} \tag{5}$$

$$v = \sum_{k=1}^{c}\sum_{i=1}^{m}\sum_{j=1}^{p} (1 - a_{ij})x_{ik}y_{jk} \tag{6}$$

Objective functions:

$$f_1 = e + v \to \min \tag{7a}$$

$$f_2 = \frac{n_1 - e}{n_1 + v} \to \max \tag{7b}$$

Constraints:

$$\sum_{k=1}^{c} x_{ik} = 1 \quad \forall i = 1, \dots, m \tag{8}$$

$$\sum_{k=1}^{c} y_{jk} = 1 \quad \forall j = 1, \dots, p \tag{9}$$

Objective function (7a) minimises the number of exceptions and voids and objective function (7b) maximises the grouping efficacy. Assignment constraints (8) and (9) provide that all machines and parts are partitioned into disjoint cells. In the next sections we prove the NP-completeness of the CFP formulations with both objective functions (7a) and (7b). In spite of almost thirty years passed after Kumar and Chandrasekharan (1990) had introduced the grouping efficacy objective and a big number of papers devoted to this CFP model, to the best of our knowledge we were the first to prove its NP-completeness.

3. NP-completeness of CFP with linear objective

To prove the NP-completeness of the CFP with linear objective function (7a) we use the Bicluster Graph Editing Problem (BGEP). The first authors who have noticed the closeness of the CFP and BGEP problems are Pinheiro et al. (2016). They applied it in their exact algorithm for the CFP with the grouping efficacy objective.

The BGEP problem consists in determining the minimum number of edges which should be added to/removed from the given bipartite graph so that it transforms to a set of isolated bicliques. An example of a BGEP instance is presented in Figure 1 and its solution – in Figure 2. Here dashed thick lines show the added edges and red thin lines – the removed edges. The BGEP problem (in its decision version) is NP-complete. The decision version of an optimization problem with objective function $f \to \max$ ($f \to \min$) is a problem with the same constraints, which only answers the question, whether there exists a feasible solution with $f \geq c$ ($f \leq c$) for a given constant c. Since the theory of NP-completeness is applicable only for decision problems in all the propositions and theorems below we will talk about the decision versions of the problems.

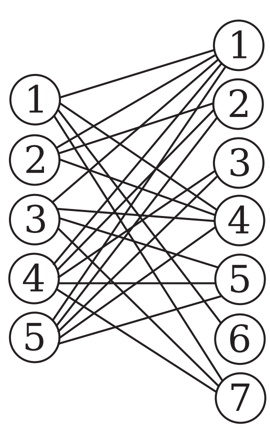

Figure 1. BGEP instance.

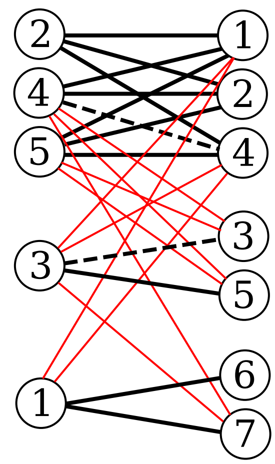

Figure 2. BGEP solution.

THEOREM 1 Amit (2004) *The BGEP problem is NP-complete because the NP-complete 3-exact 3-cover problem can be polynomially reduced to BGEP.*

The 3-exact 3-cover problem is defined as follows. Given a set of elements $U = \{1, 2, \ldots, 3n\}$ and a collection C of triplets of these elements, such that each element can belong to at most 3 triplets, determine if there exists a subcollection of C with size n which covers U.

Hereafter we will call as CFP 1 the CFP problem with the linear objective function $f_1 = e + v$ (7a), and as CFP 2 – the CFP problem with the grouping efficacy objective $f_2 = (n_1 - e)/(n_1 + v)$ (7b).

THEOREM 2 *The CFP with linear objective $f_1 = e + v$ (CFP 1) is NP-complete since it is equivalent to the BGEP problem.*

Proof There is a one-to-one correspondence between these two problems. Every machine in the CFP corresponds to a vertex in one part of the bipartite graph in the BGEP, and every part in the CFP corresponds to a vertex in another part of this graph. The machine-part matrix in the CFP coincides with the bipartite graph biadjacency matrix in the BGEP. Every exception in a solution of the CFP corresponds to an edge which should be removed from the bipartite graph in the BGEP in order to transform it to a set of isolated bicliques. And every void in a CFP solution corresponds to an edge which should be added to the bipartite graph in the BGEP.

It is clear that $f_1 = e + v \rightarrow$ min objective is equivalent to the BGEP objective of minimizing the number of added / removed edges needed to transform the input bipartite graph to a set of isolated bicliques. Every biclique corresponds to a rectangular cell in the CFP. If we remove the added edges and return back the removed edges then every isolated clique will become a non-isolated quasi-biclique completely coinciding with a rectangular cell in a CFP solution. Thus the CFP 1 problem is equivalent to the BGEP problem and it is NP-complete. ∎

For example, rows (machines) 2, 4, 5, 3, 1 in Table 2 correspond to vertices 2, 4, 5, 3, 1 in the left part of the bipartite graph in Figure 2 and columns (parts) $1, \ldots, 7$ correspond to vertices $1, \ldots, 7$ in the right part of this graph. The solution of this BGEP instance contains 3 bicliques shown with thick lines in Figure 2. Here two dashed lines represent two edges which should be added to the graph to form bicliques. Red thin lines show the edges which should be removed from the graph to isolate the bicliques from each other.

4. NP-completeness of CFP with grouping efficacy objective

To prove the NP-completeness of the CFP 2 problem we suggest the reduction of CFP 1 problem to it. The CFP 2 objective can be written in the following way.

$$f_2 = \frac{n_1 - e}{n_1 + v} = 1 - \frac{e + v}{n_1 + v} \rightarrow \max \quad \Leftrightarrow \quad \frac{e + v}{n_1 + v} \rightarrow \min$$

This expression is almost equivalent to the linear objective of the CFP 1, except the value of v in the denominator. Our idea is to nullify the influence of this value by significant increasing of the number of ones n_1. We reduce the CFP 1 problem to

Table 3. Extended matrix $\tilde{A}$.

A	0		
0	1	...	1
	⋮		⋮
	1	...	1

Table 4. Extended matrix example.

	1	2	3	4	5	6	7	8	...	42
1	1	0	1	0	0	1	1	0	...	0
2	1	1	1	0	0	0	0	0	...	0
3	1	0	1	0	1	0	1	0	...	0
4	1	1	0	1	1	0	1	0	...	0
5	1	1	1	1	1	0	0	0	...	0
6	0	0	0	0	0	0	0	1	...	1
⋮				⋮				⋮		⋮
40	0	0	0	0	0	0	0	1	...	1

CFP 2 by extending the original machine-part matrix A with a big block of ones as it is shown in Table 3. For example, for the CFP 1 instance shown in Table 1 the extended matrix $\tilde{A}$ will be as shown in Table 4. Before the main theorem using the suggested reduction and stating the NP-completeness of CFP 2 we will need to prove two propositions.

PROPOSITION 1 *If the machine-part matrix for the CFP 2 problem has identical rows/columns then there will be optimal solutions in which these rows/columns belong to the same cell.*

Proof Let us assume that there are two identical rows/columns which belong to different cells in an optimal solution, the first of these rows/columns has e_1 exceptions (ones outside its cell) and v_1 voids (zeroes inside its cell), the second row/column has e_2 exceptions and v_2 voids, and the objective function value for this solution is the following:

$$f = \frac{n_1 - e}{n_1 + v} = \frac{n_1^{in}}{n_1^{v}},$$

where n_1^v denotes $n_1 + v$. Let us also denote $\Delta e = e_2 - e_1, \Delta v = v_2 - v_1$. Our two identical rows/columns together have $e_1 + e_2$ exceptions and $v_1 + v_2$ voids. If we move the second of the identical rows/columns to the cell of the first one then these two rows/columns will have $e_1 + e_2 - \Delta e$ exceptions and $v_1 + v_2 - \Delta v$ voids, and the objective function value will be $f_1 = (n_1^{in} + \Delta e)/(n_1^v - \Delta v)$. Otherwise, if we move the first row/column to the cell of the second one, we will get $e_1 + e_2 + \Delta e$ exceptions and $v_1 + v_2 + \Delta v$ voids, and $f_2 = (n_1^{in} - \Delta e)/(n_1^v + \Delta v)$. Now we will prove that either $f_1 \geq f$ or $f_2 \geq f$:

$$f_1 \geq f \quad \Leftrightarrow \quad \frac{n_1^{in} + \Delta e}{n_1^v - \Delta v} \geq \frac{n_1^{in}}{n_1^v} \quad \Leftrightarrow \quad n_1^v \Delta e \geq -n_1^{in} \Delta v,$$

$$f_2 \geq f \quad \Leftrightarrow \quad \frac{n_1^{in} - \Delta e}{n_1^v + \Delta v} \geq \frac{n_1^{in}}{n_1^v} \quad \Leftrightarrow \quad -n_1^v \Delta e \geq n_1^{in} \Delta v.$$

From these expressions it is now obvious that $f_1 \geq f$ or $f_2 \geq f$. Without loss of generality we can assume that $f_1 \geq f$, which means that joining the identical rows/columns in the first cell will not decrease the objective function value:

$$\frac{n_1^{in} + \Delta e}{n_1^v - \Delta v} \geq \frac{n_1^{in}}{n_1^v}.$$

Let us assume that there are k identical rows/columns in the second cell. We will prove that moving all of them to the first cell can only increase f:

$$\frac{n_1^{in} + k\Delta e}{n_1^v - k\Delta v} \geq \frac{n_1^{in}}{n_1^v} \quad \Leftrightarrow \quad kn_1^v \Delta e \geq -kn_1^{in} \Delta v.$$

This is true, because we have $f_1 \geq f \;\Leftrightarrow\; n_1^v \Delta e \geq -n_1^{in} \Delta v$. Thus we have proved that any number of identical rows/columns can be moved from one cell to another, depending of the fact, in which cell they give a better value of the objective.

Now we have some identical rows/columns joined in the first cell, but there could be other identical rows/columns in some third cell. In the same way we can move all the identical rows/columns from the first cell to the third one if moving one of them gives the same or greater objective value. Otherwise, as we have proved, the opposite is true and we can move all the identical rows/columns from the third cell to the first one. Continuing in this way we could gather all identical rows/columns from all cells into one cell providing the best value of the grouping efficacy, which is not less than the initial value. So the solution with all identical rows/columns joined in one cell is also optimal. ■

The next proposition determines how much ones it is enough to add in the extended matrix in order to nullify the influence of f_2 denominator.

Proposition 2 *If the number of added ones Δn_1 in the extended matrix $\tilde{A}$ is equal to $(mp)^2$ then the maximum of f_2 on $\tilde{A}$ is obtained at the same solution (extended with the cell of added ones) at which f_1 has its minimum on matrix A.*

Proof According to Proposition 1 there exists an optimal solution for CFP 2 on the extended matrix in which the added rows as well as the added columns belong to the same cell. If the cell of the added rows differs from the cell of the added columns, we will get many voids, because these rows contain zeroes in all other columns. So in an optimal solution it has to be the same cell. No other rows or columns can belong to it, because this only increases the number of voids and exceptions due to the zeroes in all other positions of the added rows and columns. That is why the added block of ones forms a separate cell in the optimal solution we are considering. This block adds no voids or exceptions to the solution and thus a CFP 1 solution and the corresponding CFP 2 solution (obtained by adding the block of ones as an additional cell) have the same number of voids v and exceptions e. We will now prove that if $\Delta n_1 = (mp)^2$ then for any two CFP 1 solutions with objective function values f_1 and f_1' and the corresponding CFP 2 solutions with objective function values f_2 and f_2' from $f_1' < f_1$ it follows that $f_2' > f_2$.

$$f_2' = \frac{\tilde{n}_1 - e'}{\tilde{n}_1 + v'}, \quad f_2 = \frac{\tilde{n}_1 - e}{\tilde{n}_1 + v}, \quad f_1' = e' + v', \quad f_1 = e + v$$

$$f_2' > f_2 \quad \Leftrightarrow \quad \frac{\tilde{n}_1 - e'}{\tilde{n}_1 + v'} > \frac{\tilde{n}_1 - e}{\tilde{n}_1 + v} \quad \Leftrightarrow \quad \frac{e' + v'}{\tilde{n}_1 + v'} < \frac{e + v}{\tilde{n}_1 + v}$$

$$\Leftrightarrow \quad \frac{f_1'}{\tilde{n}_1 + v'} < \frac{f_1}{\tilde{n}_1 + v} \quad \Leftrightarrow \quad f_1' < f_1 \frac{\tilde{n}_1 + v'}{\tilde{n}_1 + v} \quad \Leftrightarrow \quad f_1' < f_1 + f_1 \frac{v' - v}{\tilde{n}_1 + v}.$$

So we have:

$$f_1' < f_1 + f_1 \frac{v' - v}{\tilde{n}_1 + v} \quad \Leftrightarrow \quad f_2' > f_2. \tag{*}$$

Note that in case $n_1 = 0$ the CFP 1 problem becomes trivial and so we consider only the case $n_1 \geq 1$. Since $f_1 \leq mp$ and $v - v' \leq mp$, then for $\Delta n_1 = (mp)^2$ we have:

$$f_1 \frac{v - v'}{\tilde{n}_1 + v} = f_1 \frac{v - v'}{n_1 + \Delta n_1 + v} \leq \frac{(mp)^2}{(mp)^2 + 1} \quad \Leftrightarrow \quad f_1 \frac{v' - v}{\tilde{n}_1 + v} \geq \frac{-(mp)^2}{(mp)^2 + 1}.$$

From this it follows that:

$$f_1 + f_1 \frac{v' - v}{\tilde{n}_1 + v} \geq f_1 - \frac{(mp)^2}{(mp)^2 + 1} > f_1 - 1$$

Since f_1' and f_1 are integer, then $f_1' < f_1$ is equivalent to $f_1' \leq f_1 - 1$. Thus we have:

$$f_1' < f_1 \quad \Rightarrow \quad f_1 + f_1 \frac{v' - v}{\tilde{n}_1 + v} > f_1 - 1 \geq f_1' \quad \Rightarrow \quad f_1' < f_1 + f_1 \frac{v' - v}{\tilde{n}_1 + v}$$

And according to (*) we actually get that from $f_1' < f_1$ it follows that $f_2' > f_2$. This means that the minimum value of f_1 gives the maximum of f_2 on the 'extended' solution. ■

Theorem 3 *The CFP with grouping efficacy objective* $f_2 = (n_1 - e)/(n_1 + v)$ *(CFP 2) is NP-complete because CFP 1 can be polynomially reduced to it.*

Proof We will prove that CFP 1, which answers the question, whether there exists a solution with $f_1 = e + v \leq c$ with the input matrix A can be polynomially reduced to problem CFP 2 on the extended matrix $\tilde{A}$ (see Table 3), which answers the question, whether there exists a solution with $f_2 = (\tilde{n_1} - e)/(\tilde{n_1} + v) \geq \tilde{c}$. Here constant $\tilde{c}$ can depend on constant c and other input parameters.

According to Proposition 2 to get a better solution (solution with a greater f_2 value) for CFP 2 on the extended matrix $\tilde{A}$ we should extend the CFP 1 solution with a block of $(mp)^2$ ones as shown in Table 3. So the solution for CFP 1 on matrix A and the corresponding solution for CFP 2 on matrix $\tilde{A}$ are connected in the following way.

$$\tilde{e} = e, \quad \tilde{v} = v, \quad \tilde{n}_1 = n_1 + (mp)^2,$$
$$f_2 = \frac{\tilde{n}_1 - \tilde{e}}{\tilde{n}_1 + \tilde{v}} = \frac{\tilde{n}_1 - e}{\tilde{n}_1 + v} = 1 - \frac{e + v}{\tilde{n}_1 + v} = 1 - \frac{f_1}{\tilde{n}_1 + v}.$$

Let us find the value of $\tilde{c}$ such, that for a CFP 1 solution with $f_1 \leq c$ the corresponding solution for CFP 2 will have $f_2 \geq \tilde{c}$.

$$f_1 \leq c \quad \Rightarrow \quad f_2 = 1 - \frac{f_1}{\tilde{n}_1 + v} \geq 1 - \frac{c}{\tilde{n}_1}.$$

So for $\tilde{c} = 1 - c/\tilde{n}_1$ we have $f_1 \leq c \Rightarrow f_2 \geq \tilde{c}$. This guarantees that if there are no solutions for CFP 2 with $f_2 \geq \tilde{c}$ then there exist no solutions for CFP 1 with $f_1 \leq c$.

Now let us prove that for this $\tilde{c}$ from $f_2 \geq \tilde{c}$ for CFP 2 solution it follows that the original CFP 1 solution has $f_1 \leq c$. We have:

$$f_2 = \frac{\tilde{n}_1 - e}{\tilde{n}_1 + v} = 1 - \frac{e + v}{\tilde{n}_1 + v} \geq 1 - c/\tilde{n}_1$$
$$\Leftrightarrow \quad \frac{e + v}{\tilde{n}_1 + v} \leq c/\tilde{n}_1 \quad \Leftrightarrow \quad e + v \leq c + \frac{cv}{\tilde{n}_1}$$

Note that the cases when $n_1 = 0$ or $c \geq mp$ are trivial, because in such cases we do not need to construct any CFP 2 instance and can immediately answer the CFP 1 question. So we consider $n_1 \geq 1$ and $c < mp$. Since also $v \leq mp$ we have:

$$e + v \leq c + \frac{cv}{\tilde{n}_1} < c + \frac{(mp)^2}{n_1 + (mp)^2} < c + 1.$$

Taking into account that $e + v$ is integer we can conclude that $f_1 = e + v \leq c$.

Thus we have found the value of $\tilde{c} = 1 - c/\tilde{n}_1$ such that the answer for any CFP 1 instance on the question, whether there exists a solution with $f_1 \leq c$, is 'yes', if and only if the answer for the corresponding CFP 2 instance on the question, whether there exists a solution with $f_2 \geq \tilde{c}$, is also 'yes'. Consequently, the answer to the CFP 1 question is 'no', if and only if, the answer to the CFP 2 question is also 'no'. This means that we have found a polynomial reduction by Karp of CFP 1 to CFP 2. It is also clear that CFP 2 belongs to class NP, because any 'yes'-solution can be verified in polynomial time. This proves that CFP 2 is an NP-complete problem. ■

5. Conclusion

The Cell Formation Problem is studied for more than 50 years already. It was introduced in the Grouping Technology framework by Mitrofanov (1959) and Burbidge (1975). Its most popular formulation has the grouping efficacy objective function suggested by Kumar and Chandrasekharan (1990) more than 25 years ago. At the same time the NP-status of this CFP formulation has been unknown. In this paper we fill this gap and provide the proof on NP-completeness for the classical CFP problem with the grouping efficacy objective function.

As it is noticed by Pinheiro et al. (2016) the closest problem to the CFP is the Bicluster Graph Editing Problem. It is easy to see that the CFP with the linear objective function is equivalent to the BGEP, which NP-completeness is proved by Amit (2004). The main contribution of the current paper is the suggested polynomial reduction of the CFP with the linear objective to the CFP with the grouping efficacy objective. The NP-completeness of this fractional CFP formulation follows from our reduction.

One of the perspectives of our work is to adopt the existing efficient algorithms for the BGEP to the CFP with the grouping efficacy objective. As we have shown the maximization of the grouping efficacy is equivalent to the minimization of $(e + v)/(n_1 + v)$ fraction. If we now fix the value of the denominator $n_1 + v$, we will come exactly to the BGEP problem. We have successfully applied such an approach in our exact model for the fractional CFP (Bychkov, Batsyn, and Pardalos 2014). The idea is to enumerate all possible values of the denominator, which is not so much if using good bounds on its value.

Another observation we have is that the BGEP can be also formulated as the problem of partitioning a graph into a set of quasi-biclques. The simplest graph partitioning problem is the Vertex Colouring Problem, which partitions the complementary graph into a set of cliques. Another well-known graph partitioning problem is the Minimum Cut Problem. The connection with it has been successfully used in the CFP exact algorithm by Krushinsky and Goldengorin (2012). So graph partitioning algorithms also present an interesting research direction along with graph editing ones.

One more perspective of future research is the development of branch-and-cut and branch-and-price algorithms for the CFP, which are well known in combinatorial optimization for their high performance even on real-life problems. Every possible bicluster (cell) in the CFP can be modeled with a binary variable taking value of 1 if this bicluster belongs to the solution. Thus we come to a model with an exponential number of variables which could be solved with a branch-and-price algorithm. Then to strengthen the existing models for the CFP it is necessary to find families of strong inequalities and apply branch-and-cut as well as branch-and-cut-and-price approaches to solve these models efficiently. The properties of graph editing and graph partitioning problems could be used for this purpose.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Sections 1, 2, 5 and Theorems 1, 2 in Section 3 were prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics (NRU HSE). Propositions 1, 2 and Theorem 3 in Section 4 were formulated and proved with the support of Russian Science Foundation grant RSF [grant number 14-41-00039].

ORCID

Mikhail V. Batsyn http://orcid.org/0000-0002-0302-2673
Ekaterina K. Batsyna http://orcid.org/0000-0003-2999-0338
Ilya S. Bychkov http://orcid.org/0000-0003-1681-4861

References

Ailon, Nir, Noa Avigdor-Elgrabli, Edo Liberty, and Anke van Zuylen. 2011. “Improved Approximation Algorithms for Bipartite Correlation Clustering.” In *Algorithms – ESA 2011. 19th Annual European Symposium*, edited by Camil Demetrescu, and Magnús M. Halldórsson, Lecture Notes in Computer Science 6942, 25–36. Cham: Springer International.

Alhourani, Farouq. 2016. “Cellular Manufacturing System Design Considering Machines Reliability and Parts Alternative Process Routings.” *International Journal of Production Research* 54 (3): 846–863.

Amit, Noga. 2004. “The Bicluster Graph Editing Problem.” MS diss., Tel Aviv University.

Arkat, J., H. Abdollahzadeh, and H. Ghahve. 2012. “A New Branch-and-bound Algorithm for Cell Formation Problem.” *Applied Mathematical Modelling* 36: 5091–5100.

Ballakur, A. 1985. “An Investigation of Part Family/Machine Group Formation for Designing Cellular Manufacturing Systems.” PhD diss., University of Wisconsin-Madison.

Ballakur, A., and H. J. Steudel. 1987. “A Within Cell Utilization Based Heuristic for Designing Cellular Manufacturing Systems.” *International Journal of Production Research* 25: 639–665.

Battaïa, Olga, Alexandre Dolgui, and Nikolai Guschinsky. 2017. “Decision Support for Design of Reconfigurable Rotary Machining Systems for Family Part Production.” *International Journal of Production Research* 55 (5): 1368–1385.

Battaïa, Olga, Alexandre Dolgui, Sunderesh S. Heragu, Semyon M. Meerkov, and Manoj Kumar Tiwari. 2018. “Design for Manufacturing and Assembly/disassembly: Joint Design of Products and Production Systems.” *International Journal of Production Research* 56 (24): 7181–7189.

Bonnet, Edouard, Vangelis Paschos, and Florian Sikora. 2016. “Parameterized Exact and Approximation Algorithms for Maximum K -set Cover and Related Satisfiability Problems.” *RAIRO – Theoretical Informatics and Applications* 50 (3): 227–240.

Brusco, Michael J. 2015. “An Exact Algorithm for Maximizing Grouping Efficacy in Part-machine Clustering.” *IIE Transactions* 47 (6): 653–671.

Burbidge, J. L. 1975. *The Introduction of Group Technology*. New York: John Wiley & Sons.

Busygin, S., O. Prokopyev, and P. M. Pardalos. 2008. "Biclustering in Data Mining." *Computers & Operations Research* 35 (9): 2964–2987.

Bychkov, Ilya, and Mikhail Batsyn. 2018. "An Efficient Exact Model for the Cell Formation Problem with a Variable Number of Production Cells." *Computers & Operations Research* 91: 112–120.

Bychkov, I., M. Batsyn, and P. M. Pardalos. 2014. "Exact Model for the Cell Formation Problem." *Optimization Letters* 8 (8): 2203–2210.

Chan, F. T. S., K. W. Lau, L. Y. Chan, and V. H. Y. Lo. 2008. "Cell Formation Problem with Consideration of Both Intracellular and Intercellular Movements." *International Journal of Production Research* 46 (10): 2589–2620.

Chung, S.-H., T.-H. Wu, and C.-C. Chang. 2011. "An Efficient Tabu Search Algorithm to the Cell Formation Problem with Alternative Routings and Machine Reliability Considerations." *Computers & Industrial Engineering* 60 (1): 7–15.

De Sousa Filho, Gilberto F., Teobaldo L. Bulhões Júnior, Lucidio A. F. Cabral, Luiz Satoru Ochi, and Fábio Protti. 2017. "New Heuristics for the Bicluster Editing Problem." *Annals of Operations Research* 258 (2): 781–814.

Dimopoulos, C., and A. Zalzala. 2000. "Recent Developments in Evolutionary Computation for Manufacturing Optimization: Problems, Solutions and Comparisons." *IEEE Transactions on Evolutionary Computation* 4 (2): 93–113.

Dolgui, Alexandre, and Jean-Marie Proth. 2010. *Supply Chain Engineering: Useful Methods and Techniques*. London: Springer-Verlag.

Drange, Pål Grønås, Felix Reidl, Fernando Sánchez Villaamil, and Somnath Sikdar. 2015. "Fast Biclustering by Dual Parameterization." In *10th International Symposium on Parameterized and Exact Computation (IPEC 2015)*, edited by Thore Husfeldt and Iyad Kanj, Leibniz International Proceedings in Informatics, 402–413. Saarbrücken: Dagstuhl.

Egilmez, G., B. Erenay, and G. A. Suer. 2014. "Stochastic Skill-based Manpower Allocation in a Cellular Manufacturing System." *Journal of Manufacturing Systems* 33 (4): 578–588.

Eguia, Ignacio, Jose Carlos Molina, Sebastian Lozano, and Jesus Racero. 2017. "Cell Design and Multi-period Machine Loading in Cellular Reconfigurable Manufacturing Systems with Alternative Routing." *International Journal of Production Research* 55 (10): 2775–2790.

Elbenani, B., and J. A. Ferland. 2012. "An Exact Method for Solving the Manufacturing Cell Formation Problem." *International Journal of Production Research* 50 (15): 4038–4045.

Garey, M. R., and D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: Freeman.

Ghosh, S., A. Mahanti, R. Nagi, and D. S. Nau. 1996. "Manufacturing Cell Formation by State-space Search." *Annals of Operations Research* 65 (1): 35–54.

Golhar, D. Y., and C. L. Stamm. 1991. "The Just-in-time Philosophy: A Literature Review." *International Journal of Production Research* 29 (4): 657–676.

Goncalves, J. F., and M. G. C. Resende. 2004. "An Evolutionary Algorithm for Manufacturing Cell Formation." *Computers & Industrial Engineering* 47: 247–273.

Gunasekaran, A., S. K. Goyal, I. Virtanen, and P. Yli-Olli. 1994. "An Investigation Into the Application of Group Technology in Advanced Manufacturing Systems." *International Journal of Computer Integrated Manufacturing* 7 (4): 215–228.

Guo, Jiong, Falk Hüffner, Christian Komusiewicz, and Yong Zhang. 2008. "Improved Algorithms for Bicluster Editing." In *Theory and Applications of Models of Computation. 5th International Conference TAMC 2008*, edited by Manindra Agrawal, Dingzhu Du, Zhenhua Duan, and Angsheng Li, Lecture Notes in Computer Science 4978, 445–456, Cham: Springer International.

James, T. L., E. C. Brown, and K. B. Keeling. 2007. "A Hybrid Grouping Genetic Algorithm for the Cell Formation Problem." *Computers & Operations Research* 34 (7): 2059–2079.

Krushinsky, Dmitry, and Boris Goldengorin. 2012. "An Exact Model for Cell Formation in Group Technology." *Computational Management Science* 9 (3): 323–338.

Kumar, K. R., and M. P. Chandrasekharan. 1990. "Grouping Efficacy: A Quantitative Criterion for Goodness of Block Diagonal Forms of Binary Matrices in Group Technology." *International Journal of Production Research* 28 (2): 233–243.

Kuo, Yiyo, Yen-Po Chen, and Yu-Cheng Wang. 2018. "Operator Assignment with Cell Loading and Product Sequencing in Labour-intensive Assembly Cells – A Case Study of a Bicycle Assembly Company." *International Journal of Production Research* 56 (16): 5495–5510.

Kusiak, A., J. W. Boe, and C. Cheng. 1993. "Designing Cellular Manufacturing Systems: Branch-and-bound and A* Approaches." *IIE Transactions* 25 (4): 46–56.

Liu, Chunfeng, Jufeng Wang, Joseph Y.-T. Leung, and Kai Li. 2016. "Solving Cell Formation and Task Scheduling in Cellular Manufacturing System by Discrete Bacteria Foraging Algorithm." *International Journal of Production Research* 54 (3): 923–944.

Mak, K. L., Y. S. Wong, and X. X. Wang. 2000. "An Adaptive Genetic Algorithm for Manufacturing Cell Formation." *The International Journal of Advanced Manufacturing Technology* 16 (7): 491–497.

Mitrofanov, S. P. 1959. *Nauchnye Osnovy Gruppovoy Tekhnologii [Scientific Foundations of Group Technology]*. Leningrad: Lenizdat.

Nishino, Masaaki, Norihito Yasuda, Shin-ichi Minato, and Masaaki Nagata. 2017. "Efficient Algorithm for Enumerating All Solutions to An Exact Cover Problem." *NTT Technical Review* 15 (11): 1–5.

Norman, Bryan A., Wipawee Tharmmaphornphilas, Kim Lascola Needy, Bopaya Bidanda, and Rona Colosimo Warner. 2002. "Worker Assignment in Cellular Manufacturing Considering Technical and Human Skills." *International Journal of Production Research* 40 (6): 1479–1492.

Paydar, M. M., and M. Saidi-Mehrabad. 2013. "A Hybrid Genetic-variable Neighborhood Search Algorithm for the Cell Formation Problem Based on Grouping Efficacy." *Computers & Operations Research* 40 (4): 980–990.

Pinheiro, R. G. S., I. C. Martins, F. Protti, L. S. Ochi, L. G. Simonetti, and A. Subramanian. 2016. "On Solving Manufacturing Cell Formation Via Bicluster Editing." *European Journal of Operational Research* 254 (3): 769–779.

Protti, Fábio, Maise Dantas da Silva, and Jayme Luiz Szwarcfiter. 2009. "Applying Modular Decomposition to Parameterized Cluster Editing Problems." *Theory of Computing Systems* 44 (1): 91–104.

Rezaei-Malek, Mohammad, Jafar Razmi, Reza Tavakkoli-Moghaddam, and Alireza Taheri-Moghaddam. 2017. "Towards a Psychologically Consistent Cellular Manufacturing System." *International Journal of Production Research* 55 (2): 492–518.

Solimanpur, M., S. Saeedi, and I. Mahdavi. 2010. "Solving Cell Formation Problem in Cellular Manufacturing Using Ant-colony-based Optimization." *The International Journal of Advanced Manufacturing Technology* 50: 1135–1144.

Spiliopoulos, K., and S. Sofianopoulou. 1998. "An Optimal Tree Search Method for the Manufacturing Systems Cell Formation Problem." *European Journal of Operational Research* 105: 537–551.

Sun, Peng, Jiong Guo, and Jan Baumbach. 2014. "Complexity of Dense Bicluster Editing Problems." In *Computing and Combinatorics. 20th International Conference COCOON 2014*, edited by Zhipeng Cai, Alex Zelikovsky, Anu Bourgeois, Lecture Notes in Computer Science 8591, 154–165. Cham: Springer International.

Sun, Peng, Nora K. Speicher, Richard Röttger, Jiong Guo, and Jan Baumbach. 2014. "Bi-Force: Large-scale Bicluster Editing and Its Application to Gene Expression Data Biclustering." *Nucleic Acids Research* 42 (9): e78.

Tunnukij, T., and C. Hicks. 2009. "An Enhanced Grouping Genetic Algorithm for Solving the Cell Formation Problem." *International Journal of Production Research* 47 (7): 1989–2007.

Utkina, I., M. Batsyn, and E. Batsyna. 2016. "A Branch and Bound Algorithm for a Fractional 0-1 Programming Problem." In *Discrete Optimization and Operations Research. 9th International Conference DOOR 2016*, edited by Y. Kochetov, M. Khachay, V. Beresnev, E. Nurminski, and P. Pardalos, Lecture Notes in Computer Science 9869, 244–255, Cham: Springer International.

Utkina, I. E., M. V. Batsyn, and E. K. Batsyna. 2018. "A Branch-and-bound Algorithm for the Cell Formation Problem." *International Journal of Production Research* 56 (9): 3262–3273.

Won, Youkyung, and Rasaratnam Logendran. 2015. "Effective Two-phase P-median Approach for the Balanced Cell Formation in the Design of Cellular Manufacturing System." *International Journal of Production Research* 53 (9): 2730–2750.

Zeng, Chengkuan, Jiafu Tang, and Zhi-Ping Fan. 2019. "Auction-based Cooperation Mechanism for Cell Part Scheduling with Transportation Capacity Constraint." *International Journal of Production Research* 57 (12): 3831–3846.

Žilinskas, Julius, Boris Goldengorin, and Panos M. Pardalos. 2015. "Pareto-optimal Front of Cell Formation Problem in Group Technology." *Journal of Global Optimization* 61 (1): 91–108.