

# О балансной кластеризации (индексы, модели, примеры) <sup>1</sup>

Марк Ш. Левин

Институт проблем передачи информации, Российская академия наук, Москва, Россия  
email: mslevin@acm.org

Поступила в редколлегию 13.06.2017

**Аннотация**—В статье описывается подход к задачам балансной кластеризации. Список базовых балансных структур включает следующее: балансное разбиение множества, балансные деревья, балансное разбиение графа, балансные многоуровневые структуры. Индексы баланса (балансные индексы, характеристики) для балансных структур (решений кластеризации) основаны на различии между параметрами кластеров: мощность кластера, общий вес кластера, общий вес ребер/дуг кластера, структура кластера по типам элементов. Предложенные индексы баланса используются как компоненты оптимизационных моделей балансной кластеризации: целевые функции, ограничения. Приведены три численных примера: (1) вычисление индексов баланса для решения кластеризации на основе структуры кластера по типам элементов; (2) вычисление индексов баланса решения кластеризации для сетевого примера; (3) балансная кластеризация для формирования нескольких бригад студентов.

**КЛЮЧЕВЫЕ СЛОВА:** балансная кластеризация, индексы баланса, комбинаторная оптимизация, эвристика

## 1. ВВЕДЕНИЕ

Задачи балансировки имеют большое значение во многих теоретических и практических областях [7, 8, 14, 16, 17, 18, 23, 26, 37, 38, 48, 54, 55, 57, 58, 63]. Основная рассматриваемая задача балансировки заключается в разбиении исходного множества элементов (с учетом параметров элементов, некоторых структур над элементами) на группы взаимосвязанных элементов (кластеров), которые сбалансированы, например: по числу элементов в кластере, по общему весу элементов в кластере, по общему весу связей между элементами внутри кластера, по структуре кластера по типам входящих в него элементов. Общая схема области задач балансировки приведена на Рис. 1.



Рис. 1. Общая схема области задач балансировки

<sup>1</sup> Исследование выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект 14-50-00150).

В таблице 1 приведены основные виды приложений балансных задач и структур.

**Таблица 1.** Приложений балансных задач и структур

№.	Прикладная область	Источники
1.	Иерархическая организация памяти и поисковых процессов	[2, 9, 36, 38, 43, 60]
2.	Сбалансированные параллельные расписания для вычислительных процессов	[8, 14, 19, 61]
3.	Декомпозиция сервисов в средах обработки в виде решеток (grid)	[53, 54]
4.	Сбалансированная декомпозиция сетей связи (например, модуляризация), сбалансированное размещение центров в сетях, управление сетями связи	[1, 10, 23, 37, 48, 59, 64]
5.	Сбалансированная декомпозиция алгоритмов	[58]
6.	Сбалансированные производственные расписания (в том числе задача балансировки сборочной линии)	[5, 16, 17, 26, 27, 28, 33]
7.	Построение сбалансированных иерархических организационных структур	[7, 55]
8.	Построение распределенных систем безопасности	[30, 63]

Основная задача кластеризации имеет следующий вид. Пусть  $A = \{a_1, \dots, a_j, \dots, a_n\}$  - исходное множество элементов (объектов). Обычно рассматриваются параметры:

(1) параметры каждого элемента  $a_j \in A$  как вектор  $\bar{p}(a_j) = (p_1(a_j), \dots, p_i(a_j), \dots, p_m(a_j))$ ;

(2) структура (бинарное отношение или бинарные отношения) над  $A$ :  $G = (A, E)$  где  $E$  - множество ребер/дуг (или взвешенных ребер/дуг).

Решение задачи кластеризации (решение кластеризации) обычно представляет собой множество кластеров/подмножеств (без пересечений) [44, 45]:  $\tilde{X} = \{X_1, \dots, X_\lambda\}$ , т.е., разбиение множества  $A$  на кластеры:  $X_\iota \subseteq A \quad \forall \iota = \overline{1, \lambda}$ ,  $\eta_\iota = |X_\iota|$  - размер кластера (мощность кластера  $X_\iota$ ,  $\iota \in \overline{1, \lambda}$ ). Основной индекс баланса для решения кластеризации имеет вид разницы между максимальным и минимальным размерами кластеров в рассматриваемом решении:

$$B^c(\tilde{X}) = \max_{\iota=1, \lambda} \eta_\iota - \min_{\iota=1, \lambda} \eta_\iota.$$

Базовая задача балансной кластеризации направлена на поиск решений кластеризации с минимизацией указанного индекса баланса. Дополнительно в постановку задачи можно включать следующее: (1) ограничения: (1.1) на число кластеров в решении, (1.2) на размер кластеров; (2) целевые функции (например, минимизация общего веса связей между кластерами, максимизация общего веса связей между элементами внутри кластеров).

Данная статья содержит описание предварительного подхода автора к задачам балансной кластеризации. Указаны основные балансные структуры. Предложены новые индексы баланса для решений кластеризации. Описаны оптимизационные модели балансной кластеризации, включая многокритериальные постановки. Примеры иллюстрируют вычисление индексов баланса для решений кластеризации. Специальный пример посвящен формированию групп студентов на базе задачи балансной кластеризации с балансировкой по структуре кластеров на основе типов входящих элементов. Данный материал может быть рассмотрен как продолжение статей [44, 45, 46]. Статья базируется на предварительной публикации [47].

## 2. ОСНОВНЫЕ ТИПЫ БАЛАНСНЫХ СТРУКТУР

Основные типы балансных структур приведены в таблице 2.

**Таблица 2.** Основные типы балансных структур

№.	Тип структуры	Источники
1.	Балансное разбиение множества элементов (Рис. 2)	[31, 34, 35, 39, 50]
2.	Балансированная упаковка в контейнеры (включая сбалансированные многопроцессорные расписания)	[19, 21, 31, 24, 25]
3.	Балансные деревья	[3, 9, 22, 29, 38]
4.	Балансная декомпозиция графа (Рис. 3)	[6, 8, 11, 13, 31, 40]
5.	Балансный знаковый граф	[4, 31, 32, 56, 57]
6.	Балансные структуры кластеров в сетях	[1, 41, 48, 49, 51, 59]
7.	Балансные матрицы	[12, 20, 52]

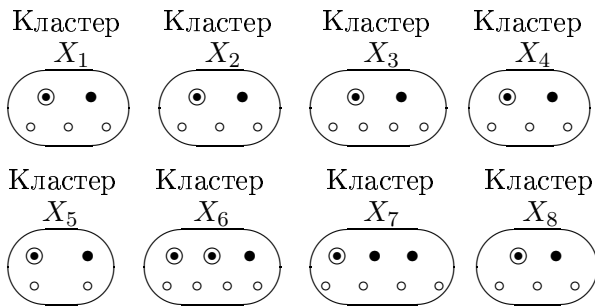


Рис. 2. Решение кластеризации

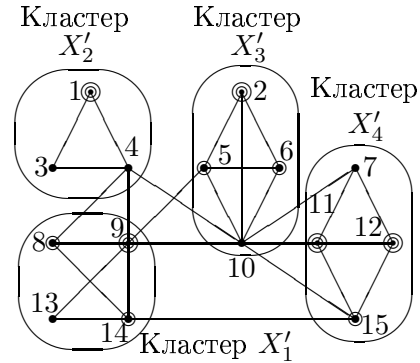


Рис. 3. Декомпозиция графа

### 3. ИНДЕКСЫ БАЛАНСА ДЛЯ РЕШЕНИЙ КЛАСТЕРИЗАЦИИ

Индексы баланса рассматриваются для решения кластеризации  $\tilde{X} = \{X_1, \dots, X_l, \dots, X_\lambda\}$ . Пусть  $p^{bal}(X_l)$  будет общий параметр кластера  $X_l$  ( $l = \overline{1, \lambda}$ ) (например, число элементов, общий вес элементов). Индексы баланса для решений кластеризации  $\tilde{X}$  можно определить на основе двух методов:

**Метод 1:** Индекс баланса определяется так:  $B(\tilde{X}) = \max_{l=\overline{1, \lambda}} p^{bal}(X_l) - \min_{l=\overline{1, \lambda}} p^{bal}(X_l)$ .

Следует отметить, что оценивание  $p^{bal}(X_l)$  можно проводить на основе различных шкал: количественная, порядковая, шкала на основе мультимножества [42, 43]. Следует рассмотреть дополнительные параметры элементов:

1. Вес элемента (вершины)  $w(a_j) \forall a_j \in A$  (т.е.,  $w(a_j) \geq 0$ ).
2. Вес связи между элементами (т.е., ребра/дуги между вершинами)  $v(a_{j_1}, a_{j_2}) \forall a_{j_1}, a_{j_2} \in A, (a_{j_1}, a_{j_2}) \in E$  (т.е.,  $v(a_{j_1}, a_{j_2}) \geq 0$ ).
3. Структура кластера по элементам определяется как специальная оценка в виде мультимножества [42, 43]:  $e(X_l), l = \overline{1, \lambda}, e(X_l) = (\theta_1, \dots, \theta_\gamma, \dots, \theta_{\bar{\gamma}})$ , где  $\theta_\gamma$  равен числу элементов типа  $\gamma$  ( $\gamma = \overline{1, \bar{\gamma}}$ ) в кластере  $X_l$ , множество типов элементов:  $\{1, \dots, \gamma, \dots, \bar{\gamma}\}, \sum_{\gamma=1}^{\bar{\gamma}} \theta_\gamma = |X_l|$ .

Пример на основе элементов 3-х типов приведен на Рис. 2 (решение кластеризации  $\tilde{X} = \{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}$ ).

Дополнительно рассматривается специальный “пустой” элемент с номером  $\bar{\gamma} + 1$ . Введение такого элемента позволяет рассматривать и сравнивать кластеры на основе выравненного размера (т.е., по числу элементов):

Для оценки кластера в виде мультимножества добавляется компонент:  $\theta_{\bar{\gamma}+1}$ , для кластера максимального размера  $\theta_{\bar{\gamma}+1} = 0$ , иначе  $\theta_{\bar{\gamma}+1} = |X|^{max} - |X_l|$ , где  $|X|^{max}$  - максимальный размер кластера.

Таким образом, число элементов в каждой оценке в виде мультимножества для структуры кластера получается равным. Близость между оценками структур двух кластеров (по типам элементов) рассматривается как путь (число шагов, расстояние по редактированию) от одной оценки к другой [42, 43] (это упрощенное определение близости оценок в виде мультимножеств). Следует отметить, что обычно типы элементов упорядочены по “важности” (в примере - четыре типа)  $1 \succeq 2 \succeq 3 \succeq 4$ . Такое упорядочение может рассматриваться как сложность коммуникационного узла: точка доступа (1), узел ретрансляции (2), узел конечного пользователя (3), “пустой” элемент (4). В результате получается шкала на основе решетки (типа частично упорядоченного множества) [42, 43].

Индексы баланса, основанные на методе 1, представлены в таблице 3.

**Метод 2.** Здесь индексы баланса определяются так:  $\widehat{B}(\tilde{X}) = \max_{\iota=1,\lambda} |p^{bal}(X_\iota) - p^0(X^0)|$ , где  $p^0$  - специально заданный опорный параметр (например, средний) для специального опорного кластера  $X^0$  (размер, вес, вес связей, оценка структуры по типам элементов).

Индексы баланса, основанные на методе 2, представлены в таблице 3.

**Таблица 3.** Индексы баланса для решения кластеризации  $\tilde{X}$

Ном.	Тип индекса	Определение индекса
I.	Метод 1:	
1.1.	По размеру кластера	$B^c(\tilde{X}) = \max_{\iota=1,\lambda}  X_\iota  - \min_{\iota=1,\lambda}  X_\iota $
1.2.	По весу элементов	$B^w(\tilde{X}) = \max_{\iota=1,\lambda} \sum_{a_j \in X_\iota} w(a_j) - \min_{\iota=1,\lambda} \sum_{a_j \in X_\iota} w(a_j)$
1.3.	По весу связей	$B^v(\tilde{X}) = \max_{\iota=1,\lambda} \sum_{a_{j_1}, a_{j_2} \in X_\iota} v(a_{j_1}, a_{j_2}) - \min_{\iota=1,\lambda} \sum_{a_{j_1}, a_{j_2} \in X_\iota} v(a_{j_1}, a_{j_2})$
1.4.	По типам элементов	$B^s(\tilde{X}) = \max_{\iota=1,\lambda} e(X_\iota) - \min_{\iota=1,\lambda} e(X_\iota)$
II.	Метод 2:	
2.1.	По размеру кластера	$\widehat{B}^c(\tilde{X}) = \max_{\iota=1,\lambda}    X_\iota  - p_{ X^0 }  $ , где $p_{ X^0 }$ - размер опорного кластера
2.2.	По весу элементов	$\widehat{B}^w(\tilde{X}) = \max_{\iota=1,\lambda}   \sum_{a_j \in X_\iota} w(a_j) - p_{w^0}  $ , где $p_{w^0}$ - общий вес опорного кластера
2.3.	По весу связей	$\widehat{B}^v(\tilde{X}) = \max_{\iota=1,\lambda}   \sum_{a_{j_1}, a_{j_2} \in X_\iota} v(a_{j_1}, a_{j_2}) - p_{v^0}  $ , где $p_{v^0}$ - общий вес связей в опорном кластере
2.4.	По типам элементов	$\widehat{B}^s(\tilde{X}) = \max_{\iota=1,\lambda}   e(X_\iota) - e_{p^0}  $ , где $e_{p^0}$ - оценка структуры опорного кластера

#### 4. ОПТИМИЗАЦИОННЫЕ МОДЕЛИ БАЛАНСНОЙ КЛАСТЕРИЗАЦИИ

Задачи балансной кластеризации направлены на построение решений кластеризации, в которых обеспечивается минимальное значение (или минимальные значения) указанных выше индексов баланса. Очевидно, что постановка задач балансной кластеризации может включать дополнительно следующее: (а) некоторые ограничения (например, фиксированное число получаемых кластеров, ограничение на размер кластеров и др.); (б) целевые функции (например, минимизация общего веса взаимосвязей между кластерами, максимизация общего веса связей внутри кластеров). В оптимизационных моделях балансной кластеризации предложенные индексы баланса могут быть используются как целевые функции и/или ограничения, например:

**Задача 1:**  $\min B^c(\tilde{X}) \quad s.t. \quad B^w(\tilde{X}) \leq w^0, \quad B^s(\tilde{X}) \leq e^0$ , где  $w^0$  - ограничение для разницы весов кластеров,  $e^0$  - ограничение на различие структур кластеров.

**Задача 2:**  $\min B^w(\tilde{X}) \quad s.t. \quad B^c(\tilde{X}) \leq c^0, \quad B^s(\tilde{X}) \leq e^0$ , где  $c^0$  - ограничение на разницу размера (мощности) кластеров,  $e^0$  - ограничение на различие структур кластеров.

**Задача 3:**  $\min B^w(\tilde{X}) \quad s.t. \quad B^c(\tilde{X}) \leq c^0, \quad |\tilde{X}| \leq \lambda^0$ , где  $c^0$  - ограничение на разницу размера (мощности) кластеров,  $\lambda^0$  - ограничение на число кластеров в решении кластеризации.

Могут рассматриваться и многокритериальные модели, например:

**Задача 4:**  $\min \tilde{B}^c(\tilde{X}), \min \tilde{B}^v(\tilde{X}) \quad s.t. \quad B^w(\tilde{X}) \leq w^0, \quad \tilde{B}^s(\tilde{X}) \leq \tilde{e}^0$ , где  $w^0$  - ограничение для разницы весов кластеров,  $\tilde{e}^0$  - ограничение на структуру кластера на основе значения структуры опорного кластера.

**Задача 5:**  $\min \tilde{B}^c(\tilde{X}), \min \tilde{B}^w(\tilde{X}), \min \tilde{B}^v(\tilde{X}) \quad s.t. \quad \tilde{B}^s(\tilde{X}) \leq \tilde{e}^0, \quad |\tilde{X}| \leq \lambda^0$ , где  $\tilde{e}^0$  - ограничение на структуру кластера на основе значения структуры опорного кластера,  $\lambda^0$  - ограничение на число кластеров в решении кластеризации.

**Задача 6:**  $\min B^w(\tilde{X}), \min \tilde{B}^v(\tilde{X}) \quad s.t. \quad B^c(\tilde{X}) \leq c^0, \quad \tilde{B}^s(\tilde{X}) \leq \tilde{e}^0$ , где  $c^0$  - ограничение на разницу размера (мощности) кластеров,  $\tilde{e}^0$  - ограничение на структуру кластера на основе значения структуры опорного кластера.

Как правило, рассматриваемые комбинаторные оптимизационные задачи являются NP-трудными. Для их решения могут применяться переборные методы или эвристики.

## 5. ПРИМЕРЫ

### 5.1. Балансные индексы для кластеров с многотипными элементами

Рассматривается пример, представленный на Рис. 2 (8 кластеров, 3 типа элементов и дополнительный “пустой” элемент). В таблице 4 содержатся описания структур кластеров и соответствующие оценки в виде мультимножеств. Таблица 5 содержит упрощенные оценки близости между оценкам структур кластеров:  $\delta(e(X_{l_1}), e(X_{l_2}))$ ,  $l_1, l_2 \in \{1, \dots, 8\}$ .

**Таблица 4.** Структуры кластеров (пример на Рис. 2)

Кластер $X_l$	Тип элемента: $\theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4$ (“дополнительный”)				Оценка в виде мультимножества $e(X_l)$
$X_1$	1	1	3	2	(1, 1, 3, 2)
$X_2$	1	1	3	2	(1, 1, 3, 2)
$X_3$	1	1	4	1	(1, 1, 4, 1)
$X_4$	1	1	3	2	(1, 1, 3, 2)
$X_5$	1	1	2	3	(1, 1, 2, 3)
$X_6$	2	1	4	0	(2, 1, 4, 0)
$X_7$	1	2	4	0	(1, 2, 4, 0)
$X_8$	1	1	3	2	(1, 1, 3, 2)

**Таблица 5.** Близость структур кластеров (пример на Рис. 2):  $\delta(e(X_{l_1}), e(X_{l_2}))$

Кластер $X_{l_1}$	Кластер $X_{l_2}$ : $X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7$						
$X_1$	0	0	1	0	1	4	3
$X_2$	0	0	1	0	1	4	3
$X_3$	1	1	0	1	2	3	2
$X_4$	0	0	1	0	1	4	3
$X_5$	1	1	1	1	0	5	4
$X_6$	4	4	3	4	5	0	1
$X_7$	3	3	2	3	4	1	0
$X_8$	0	0	1	0	1	4	3

Для решения кластеризации, представленного на Рис. 2, получаются следующие индексы баланса (метод 1):

$$B^c(\tilde{X}) = 7 - 4 = 3, \quad B^s(\tilde{X}) = e(X_6) - e(X_5) = (2, 1, 4, 0) - (1, 1, 2, 3) = \delta(e(X_6), e(X_5)) = 5.$$

Далее, кластер  $X_1$  рассматривается как опорный с соответствующими опорными параметрами. В результате, получаются следующие индексы баланса (метод 2):

$$\hat{B}^c(\tilde{X}) = |X_6| - |X_1| = 2, \quad \hat{B}^s(\tilde{X}) = \delta(e(X_6), e(X_1)) = \delta((2, 1, 4, 0), (1, 1, 3, 2)) = 4.$$

### 5.2. Балансные индексы для сетевого примера

Беспроводные сенсорные сети могут рассматриваться как много-уровневые системы (Рис. 4) [15, 62]: (1) сенсоры, (2) кластеры, (3) головной узел кластера, и т.д. Проектирование архитектуры таких систем может быть основано на кластеризации с учетом элементов разных типов.

Здесь рассматривается пример, представленный на Рис. 3 (3 типа элементов: обычные узлы в виде конечных пользователей или сенсоров, узлы-ретрансляторы, головные узлы кластера):  $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$ .

Типы элементов и их веса указаны в таблице 6, веса связи элементов (симметричное бинарное отношение) указаны в таблице 7. Рассматривается решение кластеризации из 4-х кластеров:  $\tilde{X}' = \{X'_1, X'_2, X'_3, X'_4\}$  где  $X'_1 = \{8, 9, 13, 14\}$ ,  $X'_2 = \{1, 3, 4\}$ ,  $X'_3 = \{2, 5, 6, 10\}$ ,  $X'_4 = \{7, 11, 12, 15\}$ . В таблице 8 представлена близость между структурами кластеров. Структура кластеров по типам элементов представлена как оценка на основе мультимножества  $e(X_l) = (\alpha_{\theta_1}, \alpha_{\theta_2}, \alpha_{\theta_3}, \alpha_{\theta_4})$ .

Итоговые параметры кластеров приведены в таблице 9.

Окончательно, следующие индексы баланса по методу 1 получены (решение кластеризации  $\tilde{X}' = \{X'_1, X'_2, X'_3, X'_4\}$ ):  $B^c(\tilde{X}') = 1$ ,  $B^w(\tilde{X}') = 6.7$ ,  $B^v(\tilde{X}') = 13.9$ ,  $B^s(\tilde{X}') = 4$ .

Далее, следующие опорные параметры (для опорного кластера) используются:  $p_{|X_0|} = 4$ ,  $p_{w^0} = 12.0$ ,  $p_{v^0} = 15.0$ .

Оценка структуры опорного кластера равна:  $e_{p^0} = (1, 1, 2, 0)$ . В результате, получаются значения индексов баланса по методу 2:  $\hat{B}^c(\tilde{X}') = 1$ ,  $\hat{B}^w(\tilde{X}') = 4.7$ ,  $\hat{B}^v(\tilde{X}') = 7.3$ ,  $\hat{B}^s(\tilde{X}') = 2$ .

**Таблица 6.** Параметры элементов (пример на Рис. 3)

Элемент $a_j$	Номер кластера $\iota (X'_l)$	Вес элемента $w(a_j)$	Тип элемента $\theta(a_j)$
$a_1$	2	4.2	1
$a_2$	3	5.1	1
$a_3$	2	1.1	3
$a_4$	2	2.0	3
$a_5$	3	3.1	2
$a_6$	3	3.2	2
$a_7$	4	1.0	3
$a_8$	1	3.4	2
$a_9$	1	5.0	1
$a_{10}$	3	0.9	3
$a_{11}$	4	4.5	1
$a_{12}$	4	4.8	1
$a_{13}$	1	0.8	3
$a_{14}$	1	3.4	2
$a_{15}$	4	3.7	2

**Таблица 7.** Веса ребер связи для (пример на Рис. 3)

$a_{j_1}$	$a_{j_2}$ :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1				4.1	2.1											
2						4.4	4.5				3.0					
3		4.1			1.5											
4		2.1		1.5					3.0	2.9	0.7					
5			4.4				3.6			3.0	1.0					
6			4.5				3.6				0.8					
7											3.2	2.5	3.2			
8					3.0					4.0						3.2
9					2.9	3.0			4.0		1.6			3.1	6.0	
10			3.0		0.7	1.0	0.8	3.2		1.6		5.0				3.3
11									2.5		5.0		6.2			4.3
12									3.2			6.2				4.2
13										3.1						5.0
14									3.2	6.0				5.0		2.5
15											3.3	4.3	4.2			2.5

**Таблица 8.** Близость структур кластеров (пример на Рис. 3):  $\delta(e(X'_{l_1}), e(X'_{l_2}))$

Кластер $X'_{l_1}$	Кластер $X'_{l_2}$ :	$X'_1$	$X'_2$	$X'_3$	$X'_4$
$X'_1$		0	2	0	1
$X'_2$		2	0	3	4
$X'_3$		0	3	0	1
$X'_4$		1	4	1	0

**Таблица 9.** Параметры кластеров (пример на Рис. 3)

Кластер	$\eta_l(X'_l) =  X'_l $	$\sum_{a_j \in X'_l} w(a_j)$	$\sum_{a_{j_1}, a_{j_2} \in X'_l} v(a_{j_1}, a_{j_2})$	$e(X'_l)$
$X'_1$	4	12.6	21.3	(1, 2, 1, 0)
$X'_2$	3	7.3	7.7	(1, 0, 2, 1)
$X'_3$	4	12.3	18.4	(1, 2, 1, 0)
$X'_4$	4	14.0	21.6	(2, 1, 1, 0)

Рассматривается также второе решение кластеризации:  $\tilde{X}'' = \{X''_1, X''_2, X''_3, X''_4\}$  где  $X''_1 = \{4, 8, 9, 13, 14\}$ ,  $X''_2 = \{1, 3\}$ ,  $X''_3 = \{2, 5, 6\}$ ,  $X''_4 = \{7, 10, 11, 12, 15\}$ . Близости структур кластеров приведены в таблице 10. Результирующие параметры кластеров представлена в таблице 11.

**Таблица 10.** Близость структур кластеров (пример на Рис. 3):  $\delta(e(X''_{l_1}), e(X''_{l_2}))$

Кластер $X''_{l_1}$	Кластер $X''_{l_2}$ :	$X''_1$	$X''_2$	$X''_3$	$X''_4$
$X''_1$		0	5	2	1
$X''_2$		5	0	3	6
$X''_3$		2	3	0	3
$X''_4$		1	6	3	0

Окончательно, индексы баланса по методу 1 получаются следующими:  $B^c(\tilde{X}'') = 3$ ,  $B^w(\tilde{X}'') = 9.6$ ,  $B^v(\tilde{X}'') = 29.0$ ,  $B^s(\tilde{X}'') = 6$ .

Опорная структура кластера используется как и для первого решения (т.е.,  $\tilde{X}'$ ):  $p_{|X^0|} = 4$ ,  $p_w^0 = 12.0$ ,  $p_v^0 = 15.0$ . В результате, опорная оценка имеет вид:  $e_{p^0} = (1, 1, 3, 0)$ , Индексы

баланса по методу 2 получаются следующими:  $\widehat{B}^c(\widetilde{X}'') = 2$ ,  $\widehat{B}^w(\widetilde{X}'') = 6.7$ ,  $\widehat{B}^v(\widetilde{X}'') = 14.2$ ,  $\widehat{B}^s(\widetilde{X}'') = 4$ .

Таблица 12 содержит полученные значения индексов баланса для решений кластеризации  $\widetilde{X}'$  и  $\widetilde{X}''$ .

**Таблица 11.** Параметры кластеров (пример на Рис. 3)

Кластер	$\eta_l(X_l'') =  X_l'' $	$\sum_{a_j \in X_l''} w(a_j)$	$\sum_{a_{j_1}, a_{j_2} \in X_l''} v(a_{j_1}, a_{j_2})$	$e(X_l'')$
$X_1''$	5	14.6	27.2	(1, 2, 2, 0)
$X_2''$	2	5.3	4.1	(1, 0, 1, 3)
$X_3''$	3	11.4	12.5	(1, 2, 0, 2)
$X_4''$	5	14.9	33.1	(2, 1, 2, 0)

**Таблица 12.** Индексы баланса решений кластеризации  $\widetilde{X}'$  и  $\widetilde{X}''$

Ном.	Решение $\widetilde{X}'$	Решение $\widetilde{X}''$
1.1.	$B^c(\widetilde{X}') = 1$	$B^c(\widetilde{X}'') = 3$
1.2.	$B^w(\widetilde{X}') = 6.7$	$B^w(\widetilde{X}'') = 9.6$
1.3.	$B^v(\widetilde{X}') = 13.9$	$B^v(\widetilde{X}'') = 29.0$
1.4.	$B^s(\widetilde{X}') = 4$	$B^s(\widetilde{X}'') = 6$
2.1.	$\widehat{B}^c(\widetilde{X}') = 1$	$\widehat{B}^c(\widetilde{X}'') = 2$
2.2.	$\widehat{B}^w(\widetilde{X}') = 4.7$	$\widehat{B}^w(\widetilde{X}'') = 6.7$
2.3.	$\widehat{B}^v(\widetilde{X}') = 7.3$	$\widehat{B}^v(\widetilde{X}'') = 14.2$
2.4.	$\widehat{B}^s(\widetilde{X}') = 2$	$\widehat{B}^s(\widetilde{X}'') = 4$

5.3. Балансная кластеризация для формирования бригад студентов

Рассматриваемый пример является модификацией численного примера из [44]. Имеется 10 студентов  $A = \{a_1, \dots, a_j, \dots, a_{10}\}$  (таблица 13) в области радиотехники. Четыре типа мастерства (склонности, навыков, компетенции) используются в качестве параметров/критериев (таблица 13): (1) компетенции в математике  $C_1$ , (2) теоретическая подготовка в радиотехнике  $C_2$ , (3) навыки работы с радиотехническими устройствами  $C_3$ , (4) умение подготовки отчетов (статей)  $C_4$ . Используется шкала  $[0, 1, 2, 3]$ , где 0 - отсутствие подготовки (навыков), 1 - низкий уровень подготовки (навыков), 2 - средний уровень подготовки (навыков), 3 - высокий уровень подготовки (навыков). Оценка каждого студента  $j$  ( $j = \overline{1, 10}$ ) представляет собой 4-компонентный вектор  $\xi(a_j) = (\xi^1(a_j), \xi^2(a_j), \xi^3(a_j), \xi^4(a_j))$ . Предполагается, что каждый студент имеет позитивную оценку хотя бы по одному критерию.

В таблице 14 содержатся оценки совместимости студентов  $R^f = \{\varepsilon(a_{j_1}, a_{j_2})\}$  ( $j_1, j_2 = \overline{1, 10}$ ) (порядковая шкала  $[0, 1, 2, 3]$ , 0 - несовместимость).

Структура кластера имеет вид:  $X_l = \{b_1, \dots, b_\tau, \dots, b_\mu\}$ .

Тогда исследуются параметры кластера:

1. Общая векторная оценка кластера:  $\xi(X_l) = (\xi^1(X_l), \xi^2(X_l), \xi^3(X_l), \xi^4(X_l))$ ,

где  $\xi^\kappa(X_l) = \max_{\tau=\overline{1, \mu}} \xi^\kappa(b_\tau)$ ,  $\forall \kappa = \overline{1, 4}$  (индекс критерия/параметра).

2. Общая оценка качества совместимости в кластере (оценки по дугам):

$$\varepsilon(X_l) = \sum_{b_{\tau_1}, b_{\tau_2} \in X_l, b_{\tau_1} < b_{\tau_2}} \varepsilon(b_{\tau_1}, b_{\tau_2}).$$

Задача кластеризации может быть рассмотрена следующим образом:

Найти решение кластеризации  $\widetilde{X} = \{X_1, \dots, X_l, \dots, X_\lambda\}$  (т.е., набор групп/кластеров студентов без пересечения) для выполнения специальных лабораторных работ (например, в области радиотехники) с учетом следующих требований:



(а) ограничения на число элементов (студентов) в каждом кластере (на мощность кластера):  $\eta^- \leq |X_\iota| \leq \eta^+$ ;

(б) ограничение на общее мастерство (компетенцию) в кластере:  $\xi(X_\iota) \succeq \xi_0 = (2, 3, 3, 2)$ ;

(в) целевая функция 1 направлена на минимизацию индекса баланса по мощности кластера  $B^c(\tilde{X})$ ;

(г) целевая функция 2 направлена на максимизацию (многокритериальный случай) для оценки структуры наихудшего кластера;

(д) целевая функция 3 направлена на максимизацию оценки по весу ребер (оценок совместности) в “наихудшем” кластере.

**Таблица 13.** Студенты, оценки по критериям

Студент $a_j$	$C_1$ $(\xi^1(a_j))$	$C_2$ $(\xi^2(a_j))$	$C_3$ $(\xi^3(a_j))$	$C_4$ $(\xi^4(a_j))$	Векторная оценка $\xi(a_j)$
Студент 1 ( $a_1$ )	1	2	3	3	(1,2,3,3)
Студент 2 ( $a_2$ )	0	1	2	1	(0,1,2,1)
Студент 4 ( $a_3$ )	2	2	1	3	(3,2,1,3)
Студент 5 ( $a_4$ )	0	1	2	1	(0,1,2,1)
Студент 6 ( $a_5$ )	3	3	3	3	(3,3,3,3)
Студент 9 ( $a_6$ )	3	3	3	3	(3,3,3,3)
Студент 10 ( $a_7$ )	3	3	2	3	(3,3,2,3)
Студент 11 ( $a_8$ )	0	1	3	2	(0,1,3,2)
Студент 12 ( $a_9$ )	0	2	3	1	(0,2,3,1)
Студент 13 ( $a_{10}$ )	0	1	1	1	(0,1,1,1)

**Таблица 14.** Порядковые оценки совместности студентов  $R^f = \{\varepsilon(a_{j_1}, a_{j_2})\}$

$a_{j_1} / a_{j_2}$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$
$a_1$	2	3	3	2	2	2	2	3	2
$a_2$		1	2	2	0	0	2	2	1
$a_3$		1	3	2	3	0	3	3	2
$a_4$			3	1	2	2	2	3	1
$a_5$				1	2	0	3	3	3
$a_6$						3	2	3	1
$a_7$							2	3	2
$a_8$								3	3
$a_9$									3

Получаемая версия оптимизационной модели имеет вид:

$$\min B^c(\tilde{X}) = \max_{\iota=1, \bar{\lambda}} |X_\iota| - \min_{\iota=1, \bar{\lambda}} |X_\iota|, \quad \max_{\iota=1, \bar{\lambda}} \min_{\iota=1, \bar{\lambda}} \xi(X_\iota), \quad \max_{\iota=1, \bar{\lambda}} \min_{\iota=1, \bar{\lambda}} \varepsilon(X_\iota)$$

$$s.t. \quad 3 \leq |X_\iota| \leq 4 \quad \forall \iota = \overline{1, \bar{\lambda}}, \quad \xi(X_\iota) \succeq \xi_0 = (2, 2, 3, 2) \quad \forall \iota = \overline{1, \bar{\lambda}}.$$

Задачи комбинаторной оптимизации такого типа относятся к классу NP-трудных. В этом случае используются переборные методы или эвристики. Упрощенная эвристика может иметь вид:

*Стадия 1.* Вычисление приближенного числа кластеров (например, 3).

*Стадия 2.* Выбор наиболее важных критериев: первый выбор: критерий 3, второй выбор: критерий 2.

*Стадия 3.* Выбор лучших элементов (примерно 4) с точки зрения выбранных двух критериев (например, по правилу Парето) как ядер (лидеров) для групп/кластеров. В общем случае, клики или квази-клики могут рассматриваться как указанные ядра. В численном примере получаются элементы (т.е., лидеры):  $a_1, a_5, a_6$ .

*Стадия 4.* Расширение каждого ядра за счет присоединения других элементов с учетом оценок совместимости. В результате можно рассмотреть следующее решение кластеризации:  $\tilde{X} = \{X_1, X_2, X_3\}$  ( $B^c(\tilde{X}) = 1$ ), где кластеры имеют вид:

(i)  $X_1 = \{a_1, a_2, a_3\}$ ,  $\xi(X_1) = (2, 2, 3, 3)$ ,  $\varepsilon(X_1) = 8$ ;

(ii)  $X_2 = \{a_5, a_4, a_8\}$ ,  $\xi(X_2) = (3, 3, 3, 3)$ ,  $\varepsilon(X_2) = 8$ ;

(iii)  $X_3 = \{a_6, a_7, a_9, a_{10}\}$ ,  $\xi(X_3) = (3, 3, 3, 3)$ ,  $\varepsilon(X_3) = 15$ .

Индекс баланса по весу ребер в каждом кластере:  $B^v(\tilde{X}) = 7$ .

На Рис. 4 представлено полученное решение кластеризации  $\tilde{X} = \{X_1, X_2, X_3\}$  (оценки весов/совместимости также указаны).

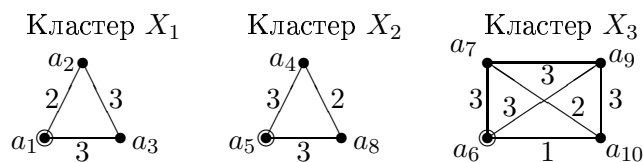


Рис. 4. Решение кластеризации

В заключение следует заметить, что рассмотренная задача похожа на задачу формирования экипажей для гражданских авиалайнеров.

## 6. ЗАКЛЮЧЕНИЕ

В данной статье предлагается подход к задачам балансной кластеризации на основе использования ряда индексов баланса для решений кластеризации. При этом оценка баланса (индексы баланса) основана на различии характеристик получаемых кластеров (число элементов, общий вес элементов, общий вес связей между элементами, структура кластера по типам элементов). Индексы баланса решений кластеризации используются как компоненты оптимизационных моделей кластеризации: целевых функций и ограничений.

Можно указать некоторые направлений дальнейших исследований: (1) исследование различных индексов баланса для решений кластеризации, (2) исследование моделей комбинаторной оптимизации с точки зрения получения балансного решения (например, задачи рюкзака, задачи упаковки в контейнеры, задача назначения), (3) исследование моделей балансной кластеризации с учетом неопределенности исходных данных, (4) исследование приложений задач балансной кластеризации (например, организация вычислений, системы связи).

## СПИСОК ЛИТЕРАТУРЫ

1. Abbasi A.A., Younis M., A survey on clustering algorithms for wireless sensor networks. *Comp. Commun.*, 2007, vol. 30, nos. 14–15, pp. 2826–2841.
2. Adelson-Velsky G.M., Landis E.M., An algorithm for the organization of information. *Soviet Math. Doklady*, 1962, vol. 3, pp. 1259–1263.
3. Aho A.V., Hopcroft J.E., Ullman J.D., *Data Structures and Algorithms*. Addison-Wesley, Reading, MA, 1983.
4. Akiyama J., Avis D., Chvatal C., Era H., Balancing sign graphs. *Discr. Appl. Math.*, 1981, vol. 3, no. 4, pp. 227–233.

5. Amen M., Cost-oriented assembly line balancing: Model formulations, solution difficulty, upper and lower bounds. *Eur. J. Oper. Res.*, 2006, vol. 168, no. 3, pp. 747–770.
6. Andreev K., Racke H., Balanced graph partitioning. *Theory of Comput. Syst.*, 2006, vol. 39, no. 6, pp. 929–939.
7. Argyres N.S., Silverman B.S., R&D, organization structure, and the development of corporate technological knowledge. *Strategic Management J.*, 2004, vol. 25, nos. 8–9, pp. 929–958.
8. Bader D.A., Meyerhenke H., Sanders P., Wagner D. (eds), *Graph Partitioning and Graph Clustering*. AMS, 2013.
9. Baer J.-L., Schwab B., A comparison of tree-balancing algorithms. *Commun. of the ACM*, 1977, vol. 29, no. 5, pp. 322–330.
10. Bar-Ilan J., Kortsarz G., Peleg D., How to allocate network centers. *J. of Algorithms*, 1993, vol. 15, no. 3, pp. 385–415.
11. Benlic U., Hao J.K., An effective multilevel tabu search approach for balanced graph partitioning. *Comp. and Oper. Res.*, 2011, vol. 38, no. 7, pp. 1066–1075.
12. Berge C., Balanced matrices. *Mathematical Programming*, 1972, vol. 2, no. 1, pp. 19–31.
13. van Bevern R., Bredereck R., Chopin M., Hartung S., Huffner F., Nichterlein A., Suchy O., Fixed-parameter algorithms for DAG partitioning. *Electr. prepr.*, 33 p., Nov. 27, 2016. <http://arxiv.org/abs/1611.08809> [cs.DS]
14. Bichot C.-E., Siarry P. (eds), *Graph Partitioning*, Wiley-ISTE, 2013.
15. Bonivento A., Fischione C., Necchi L., Pianegiani F., Sangiovanni-Vincentelli A., System level design for clustered wireless sensor networks. *IEEE Trans. Ind. Inf.*, 2007, vol. 3, no. 3, pp. 202–214.
16. Borisovsky P.A., Delorme X., Dolgui A., Balancing reconfigurable machining lines via a set partitioning model. *Int. J. of Prod. Res.*, 2014, vol. 52, no. 13, pp. 4052–4036.
17. Boysen N., Fliedner M., Scholl A., A classification of assembly line balancing problems. *Eur. J. Oper. Res.*, 2007, vol. 183, no. 2, pp. 674–693.
18. Cartwright D., Harary F., Structural balance: A generalization of Heider’s theory. *Psychol. Rev.*, 1956, vol. 63, pp. 277–293.
19. Chang R.-S., Chang J.-S., Lin P.-S., An ant algorithm for balanced job scheduling in grids. *Future Generation Computer Systems*, 2009, vol. 25, no. 1, pp. 20–27.
20. Conforti M., Cornuejols G., Vuskovic K., Balanced matrices. In: Aardal K., Nemhauser G.L., Weismantel R. (eds), *Handbooks in Operations Research and Management Science*, Elsevier, vol. 12, pp. 277–319, 2005.
21. Conway R.W., Maxwell W.L., Miller L.W., *Theory of Scheduling*. Addison-Wesley, Readings, Mass., 1967.
22. Cormen T.H., Leiserson C.E., Rivest R.L., *Introduction to Algorithms*. 3rd ed., MIT Press and McGraw-Hill, 2009.
23. Dai J., Wang S., Clustering-based interference management in densely deployment femtocell networks. *Digital Commun. and Netw.*, 2016, vol. 2, no. 4, pp. 175–183.
24. A.P. Davies, E.E. Bischoff, Weight distribution considerations in container loading. *Eur. J. Oper. Res.*, 1999, vol. 114, no. 3, pp. 509–527.
25. Dell’Olmo P., Speranza M.G., Approximation algorithms for partitioning small items in unequal bins to minimize the total size. *Discr. Appl. Math.*, 1999, vol. 94, nos. 1–3, pp. 181–191.
26. Dolgui A., Balancing assembly and transfer lines. *Eur. J. Oper. Res.*, 2006, vol. 168, no. 3, pp. 663–665.
27. Dolgui A., Guschinsky N., Levin G., A special case of transfer lines balancing by graph approach. *Eur. J. Oper. Res.*, 2006, vol. 168, no. 3, pp. 732–746.

28. Erel E., Sarin S.C., A survey of the assembly line balancing procedures. *Production Planning and Control*, 1998, vol. 9, no. 5, pp. 414–434.
29. Feldmann A.E., Foschini L., Balanced partitions of trees and applications. *Algorithmica*, 2015, vol. 71, no. 2, pp. 354–376.
30. Fu Y.-G., Zhou J., Deng L., Surveillance of a 2D plane area with 3D deployment cameras. *Sensors*, 2014, vol. 14, no. 2, pp. 1988–2011.
31. Garey M.R., Johnson D.S., *Computers and Intractability. The Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, San Francisco, 1979.
32. Harary F., On the notion of balance in signed graph. *Michigan Math.*, 1953, vol. 2, pp. 143–146.
33. Hoffmann T.R., Assembly line balancing: a set of challenging problems. *Int. J. of Prod. Res.*, 1990, vol. 28, pp. 1807–1815.
34. Horn G., Oommen B.J., Towards a learning automata solution to the multi-constraint partitioning problem. In: *2006 IEEE Conf. on Cybernetics and Intelligent Systems*, pp. 1–8, 2006.
35. Horowitz E., Sahni S., Computing partitions with applications to the knapsack problem. *J. of the ACM*, 1974, vol. 21, no. 2, pp. 277–292.
36. Karlton P.L., Fuller S.H., Scroggs R.E., Kaehler E.B., Performance of height-balanced trees. *Commun. of the ACM*, 1976, vol. 19, no. 1, pp. 23–28.
37. Khan A.W., Abdullah A.H., Anisi M.H., Bangash J.I., A comprehensive study of data collection schemes using mobile sinks in wireless sensor networks. *Sensors*, 2014, vol. 14, no. 2, pp. 2510–2548.
38. Knuth D., *The Art of computing Programming. Vol. 3: Sorting and Searching*, 3rd ed., Addison-Wesley, 1997.
39. Korf R.E., A complete anytime algorithm for number partitioning. *Artif. Intell.*, 1998, vol. 106, no. 2, pp. 181–203.
40. Krauthgamer R., Naor J., Schwattz R., Partitioning graphs into balanced components. In: *Proc. of the 20th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pp. 942–949, 2009.
41. Kuila P., Jana P.K., Approximation schemes for load balanced clustering in wireless sensor networks. *The J. of Supercomput.*, 2014, vol. 68, no. 1, pp. 87–105.
42. Левин М.Ш., Технология поддержки решений для модульных систем. Электр. книга, Москва, 2013. <http://www.mslevin.iitp.ru/Levin-bk-Nov2013-071.pdf>
43. Levin M.Sh., *Modular System Design and Evaluation*. Springer, 2015.
44. Levin M.Sh., Towards combinatorial clustering: preliminary research survey. *Electr. prepr.*, 102 p., May 28, 2015. <http://arxiv.org/abs/1505.07872> [cs.AI]
45. Левин М.Ш., О комбинаторной кластеризации: обзор литературы, методы, примеры. *Информационные процессы*, 2015, том 15, ном. 2, pp. 215–248.
46. Левин М.Ш., О динамической комбинаторной кластеризации. *Информационные процессы*, 2016, том 16, ном. 2, pp. 177–193.
47. Levin M.Sh., Towards balanced clustering - part 1 (preliminaries). *Electr. prepr.*, 21 p., June 9, 2017. <http://arxiv.org/abs/1706.03065> [cs.DS]
48. Liu X., A survey on clustering routing protocols in wireless sensor networks. *Sensors*, 2012, vol. 12, no. 8, pp. 11113–11153.
49. Low C.P., Fang C., Ng J.M., Ang Y.H., Efficient load-balanced clustering algorithms for wireless sensor networks. *Comp. Commun.*, 2008, vol. 31, no. 4, pp. 750–759.
50. Mertens S., A complete anytime algorithm for balanced number partitioning. *Electr. prepr.*, 12 p., Mar. 9, 1999. <http://arxiv.org/abs/9903011> [cs.DS]

51. Meyerhenke H., Sanders P., Schulz C., Partitioning (hierarchically clustered) complex networks via size-constrained graph clustering. *J. of Heuristics*, 2016, vol. 22, no. 5, pp. 759–782.
52. Monma C.L., Carpenter T., Variations on Matrix Balancing for Telecommunication Demand Forecasting. Technical Report, Telcordia, 1997.
53. Musunoori S.B., Horn G., Ant-based approach to the quality aware application service partitioning in a grid environment. In: 2006 IEEE Int. Conf. on Evolutionary Computation, pp. 589–596, 2006.
54. Musunoori S.B., Horn G., Intelligent ant-based solution to the application service partitioning problem in a grid environment. In: Sixth Int. Conf. on Intell. Syst. Des. and Appl., vol. 1, pp. 416–424, 2006.
55. Raisch S., Balanced structures: Designing organizations for profitable growth. *Long Range Planning*, 2008, vol. 41, no. 5, pp. 483–508.
56. Roberts F.R., *Discrete Mathematical Models with Applications Social, Biological and Environmental Problems*. Prentice Hall, Englewood Cliffs, NJ, 1976.
57. Roberts F.R., *Graph Theory and Its Applications to Problems of Society*. SIAM, Philadelphia, 1976.
58. Shang W., Fortes J.A.B., Independent partitioning of algorithms with uniform dependencies. *IEEE Trans. on Computers*, 1992, vol. 41, no. 2, pp. 190–206.
59. Shang F., Lei Y., An energy-balanced clustering routing algorithm for wireless sensor network. *Wir. Sens. Netw.*, 2010, vol. 2, no. 10, pp. 777–783.
60. Sleater D.D., Tarjan R.E., Self-adjusting binary search trees. *J. of the ACM*, 1985, vol. 32, pp. 652–686.
61. Tsai L.-H., Asymptotic analysis of an algorithm for balanced parallel processor scheduling. *SIAM J. Comput.*, 1992, vol 21, no. 1, pp. 59–64.
62. Wang L., Da Xu L., Bi Z., Xu Y., Data cleaning for RFID and WSN integration. *IEEE Trans. on Industrial Informatics*, 2014, vol. 10, no. 1, pp. 408–418.
63. Yildiz E., Akkaya K., Sisiloglu E., Sir M., Optimal camera placement for providing angular coverage in wireless video sensor networks. *IEEE Trans. on Computers*, 2014, vol. 63, no. 7, pp. 487–503.
64. Zhao M., Yang Y., Wang C., Mobile data gathering with load balanced clustering and dual data uploading in wireless sensor networks. *IEEE Trans. Mob. Comput.*, 2015, vol. 14, no. 4, pp. 770–785.

## On balanced clustering (indices, models, examples)

**Levin M.Sh.**

The article describes an approach to balanced clustering problems. The list of basic balanced structures involves: balanced set partition, balanced trees, balanced graph partition, balanced multi-layer structures. Balance indices (characteristics) for the balanced structures (clustering solutions) are based on difference between cluster parameters: cardinality of cluster, total cluster weight, total weight edges/arcs in cluster, structure of clusters by element types. The suggested balance indices are used as components of optimization balancing clustering models: objective function(s), constraints. Three numerical examples are presented: (1) calculation of balance indices for clustering solution based on structure of clusters by element types; (2) calculation of balance indices of clustering solution for network-like example; (3) balanced clustering for designing several student teams.

**KEYWORDS:** balanced clustering, balance indices, combinatorial optimization, heuristic