

## О задаче редактирования в кластеризации

Марк Ш. Левин

*Институт проблем передачи информации, Российская академия наук  
Большой Каретный пер. 19, Москва 127051, Россия  
email: mslevin@acm.org*

Поступила в редколлегию 1.09.2024 г. Принята 11.10.2024 г.

**Аннотация**—В статье рассмотрена задача редактирования в кластеризации (удаление и добавление ребер и/или вершин исходного графа с целью формирования кластерной структуры), включая ее различные версии (задача редактирования ребер, задача редактирования вершин и др.). Приведен обзор литературы по данным задачам (типы задач, методы их решения). Особое внимание уделено задаче редактирования ребер. Для этой задачи представлены математические оптимизационные формулировки на минимизацию: (1) базовая задача редактирования ребер с минимизацией общего числа удаления ребер и добавления вершин, (2) версии указанной базовой задачи с весами для всех пар вершин (включая двух-критериальный случай), (3) многокритериальная задача при векторных весах пар вершин. Кратко описаны несколько других задач редактирования: (а) задача редактирования на основе только удаления ребер, (б) задача редактирования с учетом нескольких типов вершин, (в) задача редактирования на основе только удаления вершин. Численные примеры иллюстрируют рассмотренные задачи. Приведены перспективные направления исследований.

**КЛЮЧЕВЫЕ СЛОВА:** кластеризация, редактирование кластеров, комбинаторная кластеризация, комбинаторная оптимизация.

**DOI:** 10.53921/18195822\_2024\_24\_3\_248

### 1. ВВЕДЕНИЕ

В последние годы особое внимание уделяется различным задачам и методам комбинаторной кластеризации [1, 2, 3, 4, 5, 6, 7, 8, 9]. Данная статья посвящена специальной задаче комбинаторной кластеризации: задаче редактирования в кластеризации (cluster editing problem - CEP) [10, 11, 12, 13, 14]. Следует отметить, что эта задача часто исследуется как задача корреляционной кластеризации (correlation clustering): трансформация исходного графа в множество несвязанных кластеров-клик на основе небольшого числа модификаций ребер [3, 15, 16]. С другой стороны, CEP близка к задаче разбиения на основе клик (clique clustering) [17, 18, 19].

Версии задачи редактирования CEP используются как части комбинаторных схем решения в теоретических и прикладных исследованиях, например: (i) исследование сетей (разбиение сетей, модификация сетей); (ii) различные исследования в вычислительной биологии; (iii) исследование и проектирование в производственных системах. С вычислительной точки зрения основные варианты задач редактирования в кластеризации относятся к классу NP-трудных задач [20, 21, 22]. Иллюстрация области задач редактирования в кластеризации представлена на Рис. 1. Некоторые основные типы задач редактирования в кластеризации имеют следующий вид:

1. Задача редактирования на основе модификации ребер исходного графа является, практически, основной [14, 23, 24, 25]: Преобразовать исходный граф за счет не более  $k$  модификаций ребер (добавление или удаление) в множество несвязанных клик.

2. Задача редактирования на основе ребер с целочисленными весами всех пар вершин исходного графа (ребер и не-ребер) определяются так [26]: Преобразовать исходный граф (за счет модификаций ребер - добавление или удаление) в множество несвязанных клик с общим весом модификации ребер не более  $k$ .

3. Специальная задача редактирования для двух-дольного графа (bicluster editing problem) имеет вид [27,28,29,30]: Добавить или удалить не более  $k$  ребер в исходном двух-дольном графе для получения множества несвязанных полных двух-дольных подграфов.

4.  $k$  (или  $p$ )-задача редактирования ( $k$ -СЕР или  $p$ -СЕР) имеет вид [31]: Добавить или удалить ребра в исходном графе так, чтобы множество из  $p$  несвязанных клик (кластеров) были получены на основе минимального числа операций модификации ребер.

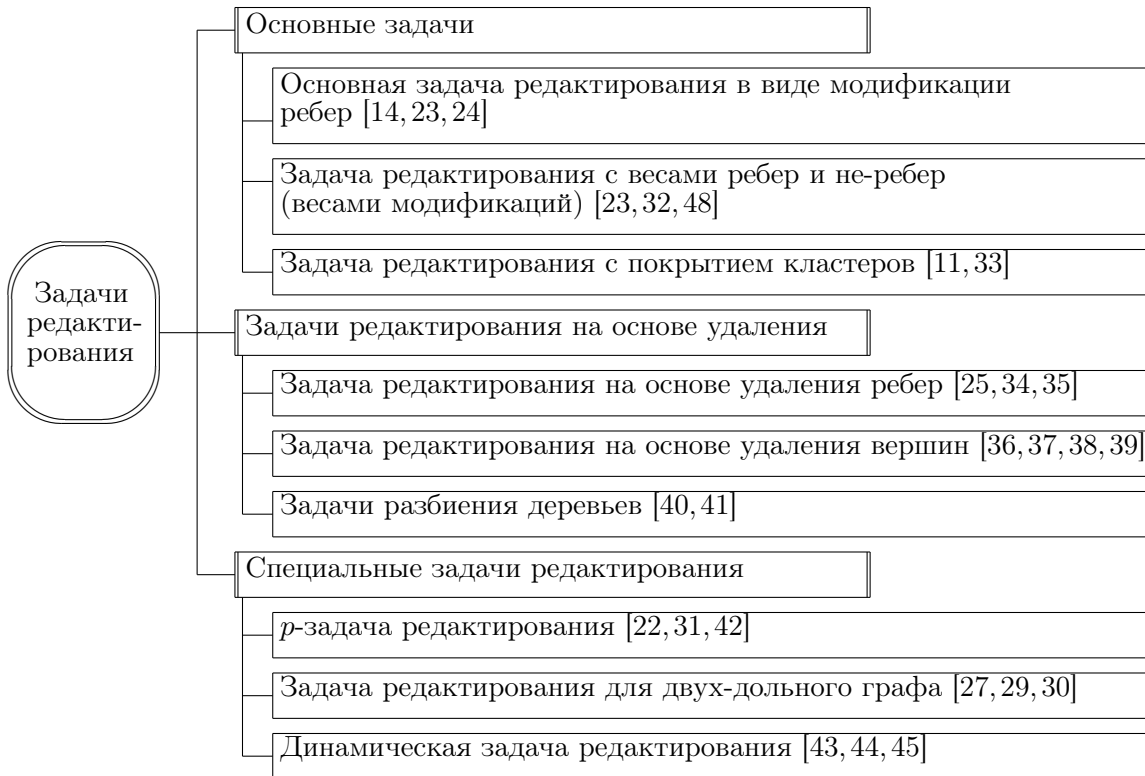


Рис. 1. Иллюстрация области задач редактирования в кластеризации

Иллюстративный пример для задачи редактирования в кластеризации представлен на Рис. 2. Здесь использованы следующие операции модификации ребер исходного графа: 5 удалений ребер и 5 добавлений ребер.

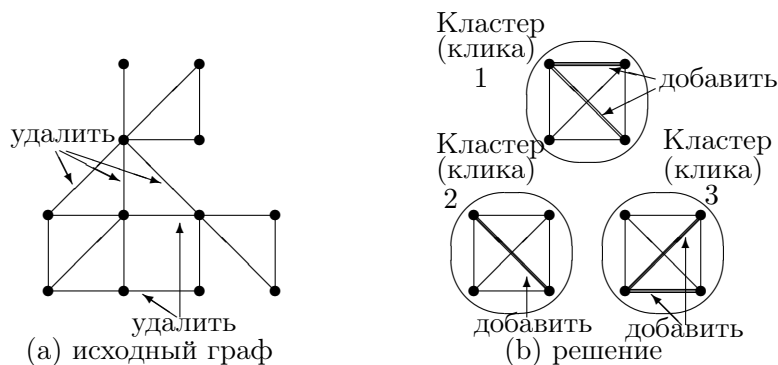


Рис. 2. Пример задачи редактирования

Данная статья содержит следующее: (а) обзор литературы в области задач редактирования в кластеризации (типы задач, методы решения); (б) описание задач редактирования ребер (cluster edge editing problem) включая комбинаторные оптимизационные модели и иллюстративные примеры; (в) краткое описание нескольких других задач редактирования; (г) краткое описание будущих направления исследований.

Материал данной статьи основан на предварительной публикации в виде препринта [46].

## 2. ОБЗОР ЛИТЕРАТУРЫ

Основные типы задач в области задач редактирования в кластеризации представлены в Таблице 1. Таблица 2 содержит список основных подходов к решению задач.

**Таблица 1.** Задачи редактирования в кластеризации

Ном.	задача/исследование	Источники
1.	Основные задачи:	
1.1.	Основные задачи редактирования в кластеризации	[14, 24, 25, 47]
1.2.	Взвешенные задачи редактирования	[23, 26, 32, 48]
1.3.	Задача редактирования с покрывающими (overlapping) кластерами	[11, 33]
1.4.	Задача редактирования на планарном графе	[49]
2.	Задачи редактирования на основе только удалений:	
2.1.	Задача редактирования на основе удалений	[25, 34]
2.2.	Задача редактирования на основе удалений для ко-графа (co-graph)	[50]
2.3.	Задача редактирования на основе удалений с фиксированным параметром	[51]
2.4.	Задача редактирования на основе удалений для интервального графа	[52]
2.5.	Задача редактирования на основе удалений вершин	[21, 36, 37, 38, 39]
2.6.	Задача редактирования на основе удалений с весами для хордового графа	[53]
2.7.	Задача разбиения дерева	[40, 41]
3.	Задачи редактирования с фиксированным параметром:	
3.1.	Задача редактирования с фиксированным параметром	[23, 51, 54, 55]
3.2.	Задача редактирования с фиксированным параметром с разбиением (splitting) вершин	[56, 57]
3.3.	Динамическая задача редактирования с фиксированным параметром	[45]
3.4.	Задача редактирования с многими фиксированными параметрами	[10]
4.	Специальные задачи редактирования:	
4.1.	Задача редактирования для двух-дольного графа	[27, 29, 30]
4.2.	$p$ -задача редактирования	[22, 31, 42]
4.3.	Задача редактирования с разбиением вершин	[56, 57]
4.4.	Задача редактирования с локально ограниченными модификациями	[58]
4.5.	Задача редактирования с небольшим числом кластеров	[59]
4.6.	Задача редактирования кластерами малого размера	[60]
4.7.	Задача редактирования на много-уровневом графе	[44]
5.	Динамические задачи редактирования:	
5.1.	Основная динамическая задача редактирования	[45]
5.2.	Задача редактирования для многовременного (temporal) графа	[43, 44]
6.	Связанные задачи:	
6.1.	Задача редактирования для ко-графа (co-graph)	[61]
6.2.	Оптимальная кластеризация для много-дольного графа	[62]
6.3.	Задача модификации вершин графа с балансом по кластерам	[63]
6.4.	Задача удаления вершин графа	[64]

Таблица 2. Основные подходы к решению

Ном.	Исследование	Источник
1.	Обзоры:	
1.1.	Теоретические исследования задач редактирования в кластеризации	[15]
1.2.	Эффективные алгоритмы для задач редактирования в кластеризации	[12]
1.3.	Точные алгоритмы для задач редактирования в кластеризации (оценивание и эксперименты)	[13]
2.	Переборные(enumerative) точные методы:	
2.1.	Метод ветвей и границ (branch-and-bound algorithm)	[65]
2.2.	Методы ветвей и разрезов (branch-and-cut approaches) для $p$ -задачи редактирования	[13, 31]
2.3.	Метод ветвей и оценок для $p$ -задачи редактирования (branch-and-price algorithm)	[42]
2.4.	Точные методы для задачи редактирования с весами	[48]
3.	Эвристики, мета-эвристики, приближенные алгоритмы:	
3.1.	Эвристики для задачи редактирования с весами	[48]
3.2.	Эвристика на основе размещения для задачи редактирования с весами	[32]
3.3.	2-аппроксимационный алгоритм для задачи редактирования на основе удаления вершин	[36]
3.4.	Гибридная эвристика для задачи редактирования с перекрытием кластеров	[33]
3.5.	Эвристики и мета-эвристики для задачи редактирования на двух-дольном графе	[27]
3.6.	Жадная (greedy) эвристика для задачи редактирования с разбиением вершин	[57]
4.	Полиномиальные алгоритмы (с полиномиальным временем выполнения):	
4.1.	Полиномиальный алгоритм для задачи редактирования на основе удаления вершин (случай блочного графа)	[66]
4.2.	Полиномиальный алгоритм для задачи редактирования на основе удаления вершин (случай расщепленного графа - split graphs)	[66]
4.3.	Полиномиальный алгоритм для задачи редактирования на основе удаления (случай ко-графа)	[50]
4.4.	Полиномиальный алгоритм для специальной задачи редактирования ((1, 1)-Cluster Editing problem)	[67]
4.5.	Полиномиальный алгоритм для специальной задачи редактирования (cluster editing problem for points on the real line)	[68]
4.6.	Полиномиальный алгоритм для задачи редактирования на интервальном графе	[69]
5.	Полиномиальные аппроксимационные схемы (PTAS):	
5.1.	PTAS для задачи редактирования на планарных графах	[49]
5.2.	PTAS для $p$ -задачи редактирования	[70]
5.3.	PTAS для задачи редактирования с кластерами небольшого размера	[60]
6.	Параметризованные алгоритмы (parameterized algorithms):	
6.1.	Параметризованный алгоритм для задачи редактирования	[23]
6.2.	Быстрый параметризованный алгоритм для задачи редактирования на основе удаления вершин	[39]
6.3.	Быстрый параметризованный алгоритм для задачи редактирования на двух-дольном графе	[30]
6.4.	Быстрый ветвистый алгоритм для задач редактирования на основе операций удаления с фиксированным параметром (fixed-parameter algorithm)	[37]
7.	Специальные подходы:	
7.1.	Точный алгоритм генерации клик	[25]
7.2.	Двух-стадийная стратегия решения для задачи редактирования на основе разбиения множества	[12]
7.3.	Комбинаторная аппроксимационная схема для задачи редактирования на основе операций удаления	[71]

3. БАЗОВАЯ ЗАДАЧА РЕДАКТИРОВАНИЯ НА ОСНОВЕ РЕБЕР

3.1. Обозначения

Рассматривается неориентированный граф  $G = (A, E)$  (или  $G = (A(G), E(G))$ ):

- (i)  $A(G) = \{a_1, \dots, a_i, \dots, a_n\}$  - множество вершин;
- (ii)  $E(G) = \{(a_i, a_j)\} \ (a_i, a_j \in A)$  - множество ребер  $E(G) \subseteq \{A \times A\}$ ;
- (iii)  $\bar{E}(G)$  - множество пар вершин, для которых нет ребер в графе (т.е., не-ребер) ( $\bar{E}(G) \subseteq \{A \times A\}$ ):  $E \cup \bar{E} = \{A \times A\}$  и  $|E \cap \bar{E}| = 0$ .

3.2. Базовая математическая формулировка

Базовая математическая формулировка для задачи редактирования ребер в виде целочисленного линейного программирования была предложена в [72]. Здесь используется факт, что  $G$  является кластерным графом если и только если  $G$  не содержит граф  $P_3$  (путь сформированный из 3-х вершин) как порожденный подграф.

Для каждой пары вершин  $a_i, a_j \ (\forall a_i, a_j \in A(G))$  при  $i < j$  рассматривается переменная  $x_{ij} \ (x_{ij} \in \{0, 1\})$ :  $x_{ij} = 0$  только в случае, если вершины  $a_i$  и  $a_j$  включены в один и тот же кластер (клику) в результирующем кластерном решении.

Базовая математическая формулировка в виде минимизационной модели имеет вид:

$$\min [ C^d = \sum_{i < j, (a_i, a_j) \in E(G)} x_{ij} + C^a = \sum_{i < j, (a_i, a_j) \in \bar{E}(G)} (1 - x_{ij}) ] \quad (1)$$

$$s.t. \ x_{ik} \leq x_{ij} + x_{jk}, \ x_{ij} \leq x_{ik} + x_{jk}, \ x_{jk} \leq x_{ij} + x_{ik}, \ i < j < k \quad (2)$$

$$x_{ij} \in \{0, 1\} \ i < j \quad (3)$$

Таким образом минимизируется целевая функция (1) в виде суммы двух компонентов:

- (а) число удаленных ребер исходного графа  $C^d$ ,
- (б) число добавленных ребер в исходном графе  $C^a$ .

Здесь имеется  $O(n^3)$  "треугольных" неравенств (2), которые исключают порождение подграфа, изоморфного структуре  $P_3$ .

Два численных примера задачи редактирования в кластеризации (СЕР) представлены ниже.

**Пример 1.** Иллюстративный пример представлен на Рис. 3:

- (а) исходный граф (Рис. 3а),
- (б) решение проведено на основе 9 операций редактирования/модификации (5 удалений ребер, 4 добавления ребер); получено множество из трех кластеров/клик (Рис. 3б).

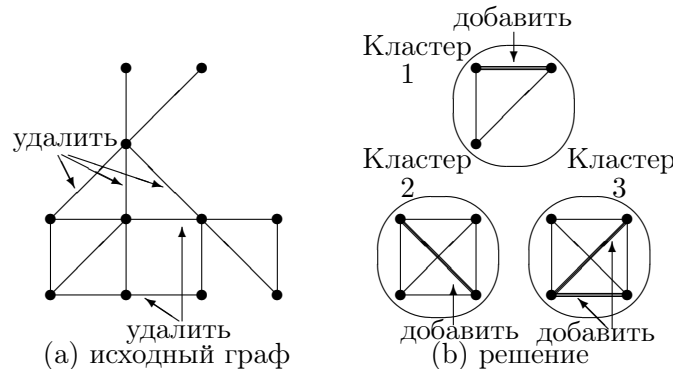


Рис. 3. Иллюстрация для примера 1

**Пример 2.** Другой иллюстративный пример представлен на Рис. 4 (используется исходный граф как в примере 1 на Рис. 3):

- (а) исходный граф (Рис. 4а),
- (б) решение проведено на основе 8 операций редактирования/модификации (7 удалений ребер, 1 добавление ребра); получено множество из четырех кластеров/клик (Рис. 4б).

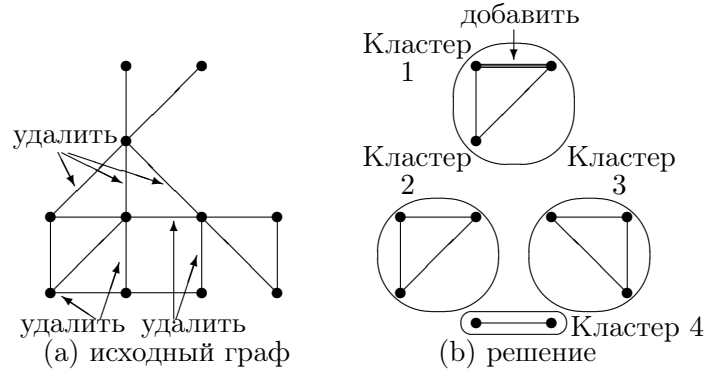


Рис. 4. Иллюстрация для примера 2

3.3. Модель с двумя целевыми функциями

Здесь можно рассматривать следующие две целевые функции (т.е., отдельно число удаленных ребер и число добавленных ребер):

$$\min C^d = \sum_{i < j, (a_i, a_j) \in E(G)} x_{ij} \tag{4a}$$

$$\min C^a = \sum_{i < j, (a_i, a_j) \notin E(G)} (1 - x_{ij}) \tag{4b}$$

$$s.t. \ x_{ik} \leq x_{ij} + x_{jk}, \ x_{ij} \leq x_{ik} + x_{jk}, \ x_{jk} \leq x_{ij} + x_{ik}, \ i < j < k \tag{5}$$

$$x_{ij} \in \{0, 1\} \quad i < j \tag{6}$$

В результате минимизируется целевая функция в виде двух-компонентного вектора:

$$\min \bar{C}^{da} = (C^d, C^a).$$

3.4. Модели с весами ребер

Базовая модель с весами ребер (Weighted Cluster Edge Editing problem) исследуется в [23, 48, 32]. Здесь используются неотрицательный вес для каждой пары вершин графа  $w_{i,j}$  ( $\forall a_i, a_j \in A(G)$ ). Таким образом веса рассмотрены и для пар вершин с ребрами и для пар вершин без ребер:

$$\min [ C^{dw} = \sum_{i < j, (a_i, a_j) \in E(G)} w_{i,j} x_{ij} + C^{aw} = \sum_{i < j, (a_i, a_j) \notin E(G)} w_{i,j} (1 - x_{ij}) ] \tag{7}$$

$$s.t. \ x_{ik} \leq x_{ij} + x_{jk}, \ x_{ij} \leq x_{ik} + x_{jk}, \ x_{jk} \leq x_{ij} + x_{ik}, \ i < j < k \tag{8}$$

$$x_{ij} \in \{0, 1\} \quad i < j \tag{9}$$

Здесь минимизируется сумма двух функций:

- (а) взвешенное число удаленных ребер как функция (1):  $C^{dw}$ ,  
 (б) взвешенное число добавленных ребер как функция (2):  $(C^{aw})$ .

Двух-критериальная модель с весами имеет вид:

$$\min C^{dw} = \sum_{i < j, (a_i, a_j) \in E(G)} w_{ij} x_{ij} \quad (10a)$$

$$\min C^{aw} = \sum_{i < j, (a_i, a_j) \notin E(G)} w_{ij} (1 - x_{ij}) \quad (10b)$$

$$s.t. \quad x_{ik} \leq x_{ij} + x_{jk}, \quad x_{ij} \leq x_{ik} + x_{jk}, \quad x_{jk} \leq x_{ij} + x_{ik}, \quad i < j < k \quad (11)$$

$$x_{ij} \in \{0, 1\} \quad i < j \quad (12)$$

В результате минимизируется взвешенная целевая функция в виде двух-компонентного вектора:

$$\min \bar{C}^{daw} = (C^{dw}, C^{aw}).$$

Многокритериальная модель основана на векторных весах пар вершин в виде:  $\bar{w}_{ij} = (w_{ij}^1, \dots, w_{ij}^\xi, \dots, w_{ij}^\lambda)$  (для каждой пары вершин  $\forall a_i, a_j \in A(G)$ , т.е., для ребер и не-ребер).

Многокритериальная модель имеет вид:

$$\min \bar{C}^{d\bar{w}} = \left( \sum_{i < j, (a_i, a_j) \in E(G)} w_{ij}^1 x_{ij}, \dots, \sum_{i < j, (a_i, a_j) \in E(G)} w_{ij}^\xi x_{ij}, \dots, \sum_{i < j, (a_i, a_j) \in E(G)} w_{ij}^\lambda x_{ij} \right) \quad (13a)$$

$$\min \bar{C}^{a\bar{w}} = \left( \sum_{i < j, (a_i, a_j) \notin E(G)} w_{ij}^1 (1 - x_{ij}), \dots, \sum_{i < j, (a_i, a_j) \notin E(G)} w_{ij}^\xi (1 - x_{ij}), \dots, \sum_{i < j, (a_i, a_j) \notin E(G)} w_{ij}^\lambda (1 - x_{ij}) \right) \quad (13b)$$

$$s.t. \quad x_{ik} \leq x_{ij} + x_{jk}, \quad x_{ij} \leq x_{ik} + x_{jk}, \quad x_{jk} \leq x_{ij} + x_{ik}, \quad i < j < k \quad (14)$$

$$x_{ij} \in \{0, 1\} \quad i < j \quad (15)$$

Здесь взвешенная векторная целевая функция используется (число компонентов вектора равно  $\lambda$ ):  $\bar{C}^{d\bar{w}}$  и  $\bar{C}^{a\bar{w}}$ . Очевидно, что две основные схемы решения могут быть использованы:

- (а) анализ интегрированной целевой функции (т.е., суммы компонентов целевой функции):

$$\min (\bar{C}^{d\bar{w}} + \bar{C}^{a\bar{w}});$$

- (б) исследование векторной целевой функции с двумя компонентами:

$$\{ \min \bar{C}^{d\bar{w}}, \min \bar{C}^{a\bar{w}} \}.$$

#### 4. КРАТКОЕ ОПИСАНИЕ ДРУГИХ ЗАДАЧ РЕДАКТИРОВАНИЯ

##### 4.1. Задача редактирования на основе удаления ребер

Данная задача редактирования основана только на операциях удаления ребер исходного графа (разбиение исходного графа) [50]. Иллюстративный пример имеет вид:

**Пример 3.** Пример задачи редактирования на основе удаления ребер представлен на Рис. 5:

- (а) исходный граф (Рис. 5а),
- (б) решение на основе 6 операций редактирования/модификации (6 удалений) для получения 4-х клик/класетов (Рис. 5б).

Следует отметить, что каждый кластер содержит 3 вершины (т.е., решение является сбалансированным по размеру кластера).

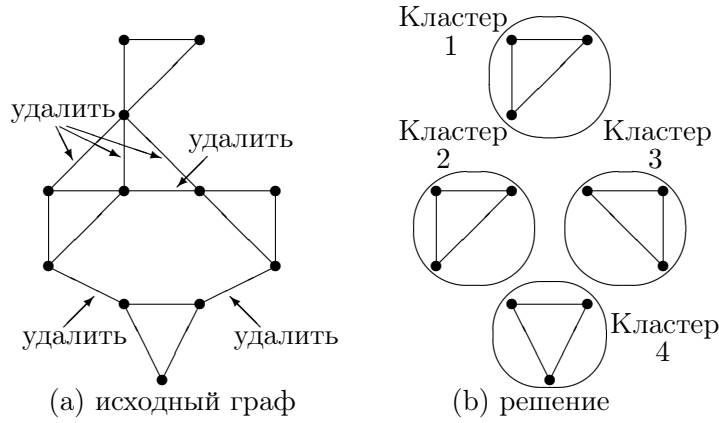


Рис. 5. Пример сбалансированного решения

В данной задаче предполагается минимизация числа удаленных ребер между кластерами/кликками в получаемом решении. Задача относится к классу NP-трудных задач и интенсивно исследуется с точки зрения применения специальных быстрых алгоритмов: разрешимость с фиксированным параметром (fixed-parameter tractability - FPT). Последние годы многие такие FPT-алгоритмы разработаны для задач кластеризации на специальных ограниченных графовых структурах и имеют полиномиальную вычислительную сложность.

Другой иллюстративный пример для данной задачи представлен на Рис. 6 (7 операций модификации/удаление ребер).

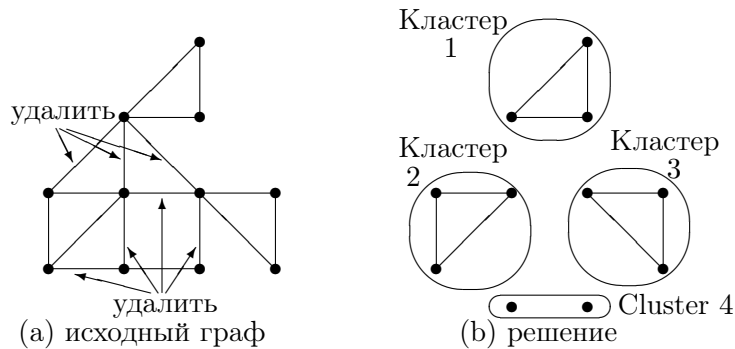


Рис. 6. Задача удаления ребер

Следует отметить, данная задача также исследуется при наличии весов ребер (weighted edge deletion problem) [53].

#### 4.2. Задача редактирования с учетом нескольких типов вершин

Задачу редактирования с учетом нескольких типов вершин можно отнести к специальному новому типу. При этом рассматривается баланс с учетом типов вершин [73]. Иллюстративный пример имеет следующий вид:



**Пример 4.** Пример представлен на Рис. 7:

(а) исходный граф (три типа вершин) (Рис. 7а),

(б) решение на основе 6-ти операций модификации дуг (4 удаления, 2 добавления) в виде 4-х клик/кластеров (Рис. 7б).

В полученном решении каждый кластер содержит 3 вершины различного типа (т.е., баланс по числу вершин и по составу кластеров).

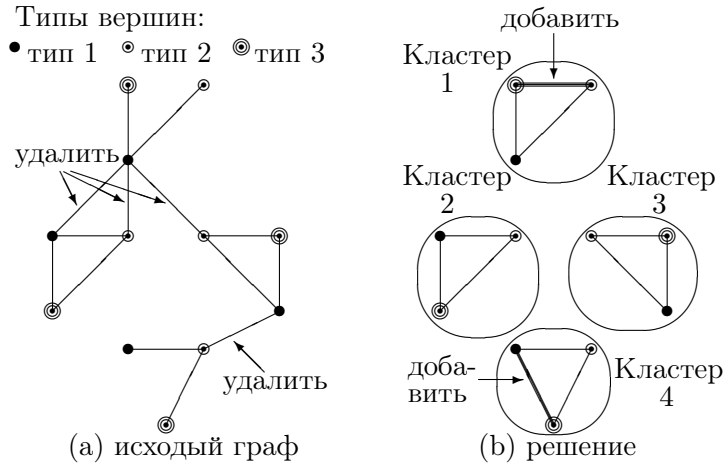


Рис. 7. Решение с балансом по типам вершин

#### 4.3. Оптимальная кластеризация на много-дольном графе

Задача оптимальная кластеризация на много-дольном графе (или разбиение на основе клик) представляет собой специальную постановку [62]:

Разбиение вершин много-дольного графа на множество несвязанных клик на основе добавления или удаления минимального множества ребер.

Эта задача относится к классу NP-трудных задач.

#### 4.4. О задаче редактирования на основе удаления вершин

Задача редактирования на основе удаления вершин (cluster vertex deletion problem) является важной для теории и приложений [36, 37, 38, 39]. Обычно исследуются следующие основные варианты данной задачи: (i) базовая версия задачи, (ii) версия задачи с весами вершин, (iii) многокритериальная версия задачи.

## 5. ЗАКЛЮЧЕНИЕ

Статья посвящена задаче редактирования в кластеризации (удаление и добавление ребер и/или вершин исходного графа с целью формирования кластерной структуры). Материал содержит следующие части: (а) обзор литературы в области задач редактирования в кластеризации (типы задач, методы решения); (б) описание нескольких задач редактирования на основе модификации ребер (cluster edge editing problem): комбинаторные оптимизационные модели, модели с весами для всех пар вершин, иллюстративные примеры; (в) краткое описание нескольких других задач задач редактирования в кластеризации (задача редактирования на

основе удаления ребер - cluster edge deletion problem, задача редактирования учетом нескольких типов вершин, задача редактирования на основе удаления вершин - cluster vertex deletion problem); (г) краткое описание будущих направления исследований.

Можно указать следующие перспективные направления исследования:

- (1) исследование задач редактирования в кластеризации в условиях неопределенности;
- (2) исследование задач с редактированием вершин (например, [36, 39]);
- (3) исследование задач с одновременным редактированием ребер и вершин (например, [28]);
- (4) исследование динамических задач редактирования в кластеризации (например, [43, 44, 45]);
- (5) исследование связей между задачами редактирования в кластеризации и другими близкими комбинаторными задачами;
- (6) использование задач редактирования в кластеризации как базовых компонентов в задачах модификации графов (например, [22]);
- (7) исследование специальных новых стратегий решения;
- (8) применение задач редактирования в кластеризации при проектировании и управлении в динамических коммуникационных системах;
- (9) использование данной задачи при обучении студентов.

Автор подтверждает, что нет конфликтов интересов.

Данное исследование выполнено в рамках проекта по гранту Министерства науки и высшего образования Российской Федерации, проект № FFNU-2022-0036 (Институт проблем передачи информации им. А.А. Харкевича Российской академии наук).

#### СПИСОК ЛИТЕРАТУРЫ

1. Broderick T., Mackey L., Paisley J., Jordan M.I., Combinatorial clustering and the beta negative binomial process. *IEEE Trans. PAMI*, 2015, vol. 37, no. 2, pp. 290–306.
2. Cao Y., Wang Z., Combinatorial optimization-based clustering algorithm for wireless sensor networks. *Mathematical Problems in Engineering*, 2020, art. ID 613704, pp. 1–13.
3. Cohen-Addad V., Lolck D.R., Pilipczuk M., Thorup M., Yan S., Zhang H., Combinatorial correlation clustering. In: *Proc. of the 56th Annual ACM Symp. on Theory of Computing STOC'24*, pp. 1617–1628, 2024.
4. Festa P., Combinatorial optimization approaches for data clustering. In: M. Celebi, K. Aydin (eds), *Unsupervised Learning Algorithms*, Springer, pp. 109–134, 2015.
5. Jarbouri B., Cheikh M., Siarry P., Rebai A., Combinatorial particle swarm optimization (CPSO) for partitional clustering problem. *Appl. Math. Comput.*, 2007, vol. 192, no. 2, pp. 337–345.
6. Kim K., Kumagai M., Yamamoti Y., Combinatorial clustering with a coherent XY machine. *Optics Express*, 2024, vol. 32, no. 19, pp. 33737–33757.
7. Kumar V., Bass G., Tomlin C., Dulny III J., Quantum annealing for combinatorial clustering. *Quantum Inform. Process.*, 2018, vol. 17, art. 39, pp. 1–14.
8. Levin M.S., On combinatorial clustering: literature review, methods, examples. *J. of Commun. Technol. Electr.*, 2015, vol. 60, no. 12, pp. 1403–1428.
9. Masoud H., Jalili S., Hasheminejad S.M.H., Dynamic clustering using combinatorial particle swarm optimization. *Appl. Intell.*, 2013, vol. 38, no. 3, pp. 289–314.

10. Abu-Khizam F.N., On the complexity of multi-parameterized cluster editing. *J. Discrete Algorithms*, 2017, vol. 45, pp. 26–34.
11. Arrighi E., Bentert M., Drange P.G., Sullivan B.D., Wolf P., Cluster editing with overlapping communities. In: 18th Int. Symp. on Parameterized and Exact Computation IPEC 2023, vol. 285, pp. 2:1–2:12, 2023.
12. Bastos L., Ochi L.S., Protti F., Subramanian A., Martins L.C., Pinheiro R.G.S., Efficient algorithms for cluster editing. *J. Combin. Optim.*, 2016, vol. 31, no. 1, pp. 347–371.
13. Bocker S., Briesemeister S., Klau G.W., Exact algorithms for cluster editing: evaluation and experiments. *Algorithmica*, 2011, vol. 60, pp. 316–334.
14. Bocker S., Baumbach J., Cluster editing. In: Bonizzoni P., Brattka V., Lowe B. (eds), *Proc. of the 9th Conf. on Computability in Europe CiE 2013*, LNCS 7921, Springer, pp. 33–44, 2013.
15. Bansal N., Blum A., Chawla S., Correlation clustering. *Machine Learning*, 2004, vol. 56, nos. 1–3, pp. 89–113.
16. Demaine E.D., Emanuel D., Fiat A., Immorlica N., Correlation clustering in general weighted graphs. *Theor. Comp. Sci.*, 2006, vol. 361, pp. 172–187.
17. Du Y., Kochenberger G., Glover F., Wang H., Lewis M., Xie W., Tsuyuguchi T., Solving clique partitioning problems: a comparison of models and commercial solvers. *Int. J. of Information Technology & Decision Making*, 2022, vol. 21, no. 01, pp. 59–81.
18. Jovanovic R., Sanfilippo A.P., Voss S., Fixed set search applied to the clique partitioning problem. *Eur. J. of Oper. Res.*, 2023, vol. 309, no. 1, pp. 65–81.
19. Levin M.S., On the clique partitioning of a graph. *J. of Commun. Technol. Electr.*, 2022, vol. 67, no. S2, pp. S267–S274.
20. Krivanek M., Moravek J., NP-hard problems in hierarchical-tree clustering. *Acta Inform.*, 1986, vol. 23, no. 3, pp. 311–323.
21. Le H.-O., Le V.B., Complexity of the cluster vertex deletion problem on  $H$ -free graphs. In: 47th Int. Symp. on Math. Foundations of CS (MFCS 2022), art. no. 68, pp. 68:1–68:10, 2022.
22. Shamir R., Sharan R., Tsur D., Cluster graph modification problems. *Discr. Appl. Math.*, 2004, vol. 144, no. 1, pp. 173–182.
23. Bocker S., Briesemeister S., Bui Q.B.A., Truss A., Going weighted: parameterized algorithms for cluster editing. *Theor. Computer Science*, 2009., vol. 410, no. 52, pp. 5467–5480.
24. Dehne F., Langston M.A., Luo X., Pitre S., Shaw P., Zhang Y., The cluster editing problem: implementation and experiments. In: 2nd Int. Workshop on Parameterized and Exact Computation IWPEC 2006, LNCS 4169, Springer, pp. 13–24, 2006.
25. Gramm J., Guo J., Huffner F., Niedermeier R., Graph-modeled data clustering: fixed-parameter algorithms for clique generation. *Theory of Computing Systems*, 2005, vol. 38, no. 4, pp. 373–392.
26. Bocker S., A golden ratio parameterized algorithm for Cluster Editing. *J. of Discrete Algorithms*, 2012, vol. 16, pp. 79–89.
27. de Sousa Filho G.F., Bulhoes Junior T.L., Cabral L.A.F., Ochi L.S., Protti F., New heuristics for the bicluster editing problem. *Ann. Oper. Res.*, 2017, vol. 258, no. 2, pp. 781–814.
28. Hochbaum D., Approximating clique and biclique problems. *J. of Algorithms*, 1998, vol. 29, no. 1, pp. 174–200.
29. Pinheiro R.G.S., Martins I.C., Protti F., Ochi L.S., Simonetti L.G., Subramanian A., On solving manufacturing cell formation via bicluster editing. *Eur. J. Oper. Res.*, 2016, vol. 254, no. 3, pp. 769–779.
30. Tsur D., Faster parameterized algorithm for Bicluster Editing. *Inform. Proc. Lett.*, 2021, vol. 168, art. 106095.

31. Bulhoes T., de Sousa Filho G.F., Subramanian A., dos Anjos F. Cabral L., Branch-and-cut approaches for  $p$ -Cluster Editing. *Discr. Appl. Math.*, 2017, vol. 219, pp. 51–64.
32. Wittkop T., Baumbach J., Lobo F., Rahmann S., Large scale clustering of protein sequences with FORCE – A layout based heuristic for weighted cluster editing. *BMC Bioinformatics*, 2007, vol. 8, no. 1, art. 396, pp. 1–12.
33. Chagas G.O., Lorena L.A.N., dos Santos R.D.C., A hybrid heuristic for the overlapping cluster editing problem. *Appl. Soft Computing*, 2019, vol. 81, art. 105482, pp. 1–12.
34. Damaschke P., Bounded-degree techniques accelerate some parameterized graph algorithms. In: *Proc. of Int. Workshop on Parameterized and Exact Computation (IWPEC 2009)*, LNCS 5917, Springer, pp. 98–109, 2009.
35. Hartuv E., Shamir R., A clustering algorithm based on graph connectivity. *Inf. Process Lett.*, 2000, vol. 76, pp. 175–181.
36. Aprile M., Drescher M., Florini S., Huynh T., A tight approximation algorithm for the cluster vertex deletion problem. *Electr. prepr.*, 26 p., Oct. 18, 2021. <http://arxiv.org/abs/2007.08057> [math.CO]
37. Boral A., Cygan M., Kociumaka T., Pilipczuk M., A fast branching algorithm for cluster vertex deletion. *Theory Comput. Syst.*, 2016, vol. 58, pp. 357–376.
38. Huffner F., Komusiewicz C., Moser H., Niedermaier R., Fixed-parameter algorithms for cluster vertex deletion. *Theory of Comput. Syst.*, 2010, vol. 47, no. 1, pp. 196–217.
39. Tsur D., Faster parameterized algorithm for cluster vertex deletion. *Theory Comput. Syst.*, 2021, vol. 65, pp. 323–343.
40. Cordone R., Franchi D., Scozzari A., Cardinality constrained connected balanced partitions of trees under different criteria. *Discr. Optim.*, 2022, vol. 46, art. 100742.
41. Feldmann A.E., Foschini L., Balanced partitions of trees and applications. *Algorithmica*, 2015, vol. 71, no. 2, pp. 354–376.
42. Bulhoes T., Subramanian A., de Sousa Filho G.F., dos Anjos F. Cabral L., Branch-and-price for  $p$ -cluster editing. *Computational Optimization and Applications*, 2017, vol. 67, pp. 293–316.
43. Bocci C., Capresi C., Meeks K., Sylvester J., A new temporal interpretation of cluster editing. *J. of Computer and System Sciences*, 2024, vol. 144, art. 103551, pp. 1–21.
44. Chen J., Molter H., Sorge M., Suchy O., Cluster editing for multi-layer and temporal graphs. *Theory Comput. Syst.*, 2024, in press.
45. Luo J., Molter H., Nichterlein A., Niedermeier R., Parameterized dynamic cluster editing. *Algorithmica*, 2021, vol. 83, no. 1, pp. 1–44.
46. Levin M.S., Towards cluster edge editing problems (literature survey, models; preliminary material). Preprint, 16 p., Oct. 7, 2024. DOI:10.13140/RG.2.2.32967.71849 (ResearchGate)
47. Damaschke P., Fixed-parameter enumerability of cluster editing and related problems. In: *Proc. 6th Int. Conf. on Algorithms and Complexity CIAC 2006*, LNCS 3998, Springer, pp. 344–355, 2006.
48. Rahmann S., Wittkop T., Baumbach J., Martin M., Truss A., Bocker S., Exact and heuristics for weighted cluster editing. In: *6th Annual Int. Conf. on Computational Systems Bioinformatics CSB 2007*, 2007, vol. 6, no. 1, pp. 391–401.
49. Berger A., Grigoriev A., Winokurow A., A PTAS for the Cluster Editing Problem on Planar Graphs. In: K. Jansen, M. Mastrolilli (eds), *WAOA 2016*, LNCS 10138, Springer, pp. 27–39, 2017.
50. Gao Y., Hare D.R., Nastos J., The cluster deletion problem for cographs. *Discr. Math.*, 2013, vol. 313, pp. 2763–2771.
51. Bocker S., Damaschke P., Even faster parameterized cluster deletion and cluster editing. *Inform. Process. Lett.*, 2011, vol. 111, no. 14, pp. 717–721.

52. Konstantinidis A.L., Papadopoulos C., Cluster deletion on interval graphs and split related graphs. *Algorithmica*, 2021, vol. 83, no. 7, pp. 2018–2026.
53. Bonomo F., Duran G., Napoli A., Velencia-Pabon M., Complexity of the cluster deletion problem on subclasses of chordal graphs. *Theor. Comput. Sci.*, 2015, vol. 600, pp. 59–69.
54. Fellow M.R., Langston M.A., Rosamond F.A., Shaw P., Efficient parameterized preprocessing for Cluster Editing. In: E. Csuhaj-Varju, Z. Esik (eds), *Proc. 16th FCT, Fundamentals of Computation Theory FCT*, LNCS 4639, Springer, pp. 312–321, 2009.
55. Protti F., da Silva M.D., Szwarcfiter, J.L., Applying modular decomposition to parameterized cluster editing problems. *Theory Comput. Syst.*, 2009, vol. 44, no. 1, pp. 91–104.
56. Abu-Khzam F.N., Egan J., Gaspers S., Shaw A., Shaw P., On the parameterized cluster editing with vertex splitting problem. *Electr. prepr.*, 23 p., Jan. 1, 2019. <http://arxiv.org/abs/1901.00156> [cs.CC]
57. Abu-Khzam F.N., Barr J.R., Fakhereldine A., Shaw P., A greedy heuristic for cluster editing with vertex splitting. In: *2021 4th Int. Conf. on Artif. Intell. for Industries (AI4I)*, pp. 38–41, 2021.
58. Komusiewicz C., Uhlmann J., Cluster editing with locally bounded modifications. *Discrete Appl. Math.*, 2012, vol. 160, no. 15, pp. 2259–2270.
59. Fomin F.V., Kratsch S., Pilipczuk M., Pilipczuk M., Villanger Y., Tight bounds for parameterized complexity of cluster editing with a small number of clusters. *J. Comput. Syst. Sci.*, 2014, vol. 80, no. 7, pp. 1430–1447.
60. Kononov A., Il'ev V., On cluster editing problem with clusters of small sizes. In: N. Olenev, Y. Evtushenko, M. Jacimovic, M. Khachay, V. Malkova (eds), *Optimization and Applications (OPTIMA 2023)*, LNCS 14395, Springer, pp. 316–328, 2023.
61. Liu Y., Wang J., Guo J., Chen J., Complexity and parameterized algorithms for Cograph Editing. *Theor. Comp. Sci.*, 2012, vol. 461, pp.v45–54.
62. Charon I., Hundry O., Optimal clustering of multipartite graphs. *Discrete Applied Math.*, 2008, vol. 156, no. 8, pp. 1330–1347.
63. Madathil J., Meeks K., Parameterized algorithms for balanced cluster edge modification problems. *Electr. prepr.*, 66 p., Sep. 4, 2024. <http://arxiv.org/abs/2403.03830> [cs.DS]
64. Kumar M., Mishra S., Devi N.S., Saurabh S., Approximation algorithms for node deletion problems on bipartite graphs with finite forbidden subgraph characterization. *Theor. Comp. Sci.*, 2014, vol. 526, pp. 90–86.
65. Blasius T., Fischbeck P., Gottesburen L., Hamann M., Heuer T., Spinner J., Weyand C., Wilhelm M., A branch-and-bound algorithm for cluster editing. In: C. Schulz, B. Ucar (eds), *20th Int. Symp. on Experimental Algorithms (SEA 2022)*, art. no. 13, pp. 13:1–13:19, 2022.
66. Cao Y., Ke Y., Otachi Y., You J., Vertex deletion problems on chordal graphs. *Theor. Comput. Sci.*, 2018, vol. 745, pp. 75–86.
67. Gutin G., Yeo A., (1,1)-Cluster Editing is polynomial-time solvable. *Discr. Appl. Math.*, 2023, vol. 340, pp. 259–271.
68. Bassel M., Cluster editing problem for points on the real line: a polynomial time algorithm. *Inf. Process. Lett.*, 2010, vol. 110, no. 21, pp. 961–965.
69. Lin M.C., Soullignac F.J., Szwarcfiter J.L., A faster algorithm for the cluster editing problem on proper interval graphs. *Inform. Process. Lett.*, 2015, vol. 115, no. 12, pp. 913–916.
70. Giotis T., Guruswami V., Correlation clustering with a fixed number of clusters. In: *Proc. of the Seventeenth Annual ACM-SIAM Symp. on Discrete Algorithm, SODA06*, pp. 1167–1176, 2006.
71. Balmaseda V., Xu Y., Cao Y., Veldt N., Combinatorial approximations for cluster deletion: simpler, faster, and better. *Electr. prepr.*, 29 p., Apr. 24, 2024. <http://arxiv.org/abs/2404.16131> [cs.DS]

72. Charikar M., Guruswami V., Wirth A., Clustering with qualitative information. *J. Comput. Syst. Sci.*, 2005, vol. 71, pp. 360–383.
73. Levin M.S., On balanced clustering (indices, models, examples). *J. of Commun. Technol. Electr.*, 2017, vol. 62, no. 12, pp. 1506–1515.

## On Editing Problem in Clustering

M.Sh. Levin

The paper addresses the editing problem in clustering (deletion and addition of edges and/or vertices in an initial graph to construct a cluster structure) including various versions of the problem (edge cluster editing, vertex cluster editing, etc.). A literature survey on the problems is presented (problem types, solving approaches). The cluster edge editing problem is under special study. Several mathematical optimization models are described for the problem: (1) the basic edge editing problem with minimization of the number of added edges and number of deleted edges, (2) the version of the above-mentioned problem with weights of all vertex pairs (including the bi-criteria problem case), (3) a multicriteria problem with vector weights of the vertex pairs. In addition a brief description of some other editing problems are presented: (a) edge deletion editing problem, (b) edge editing problem with several vertex types, (c) vertex deletion editing problem. Numerical examples illustrate the problems. Some future research directions are pointed out.

**KEYWORDS:** clustering, cluster editing, combinatorial clustering, combinatorial optimization