Discrete Applied Mathematics 365 (2025) 190–207

Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

# The complexity of cluster vertex splitting and company[☆]

Alexander Firbas[1],[*], Alexander Dobler[2], Fabian Holzer, Jakob Schafellner, Manuel Sorge[3], Anaïs Villedieu, Monika Wißmann

*TU Wien, Vienna, Austria*

ARTICLE INFO

*Article history:*
Received 12 April 2024
Received in revised form 9 December 2024
Accepted 8 January 2025
Available online 24 January 2025

*Keywords:*
Parameterized algorithms
Data reduction
Compact letter display
Computational complexity

ABSTRACT

Clustering a graph when the clusters can overlap can be seen from three different angles: We may look for cliques that cover the edges of the graph with bounded overlap, we may look to add or delete few edges to uncover the cluster structure, or we may split vertices to separate the clusters from each other. Splitting a vertex $v$ means to remove it and to add two new copies of $v$ and to make each previous neighbor of $v$ adjacent with at least one of the copies. In this work, we study underlying computational problems regarding the three angles to overlapping clusterings, in particular when the overlap is small. We show that the above-mentioned covering problem is NP-complete. We then make structural observations that show that the covering viewpoint and the vertex-splitting viewpoint are equivalent, yielding NP-hardness for the vertex-splitting problem. On the positive side, we show that splitting at most $k$ vertices to obtain a cluster graph has a problem kernel with $O(k)$ vertices. Finally, we observe that combining our hardness results with structural observations and a so-called critical-clique lemma yields a simple alternative NP-hardness proof for the Cluster Editing With Vertex Splitting problem, where we add or delete edges and split vertices to obtain a cluster graph.

## 1. Introduction

In classical graph-clustering, we want to partition the input graph into clusters that are densely connected, while there are few connections between different clusters. However, in clusterings of real-world graphs the clusters often overlap [44]. We are interested here in exact algorithms for and complexity of such overlapping clustering problems. Without overlap, these are well-studied (e.g. [8–13,17,29,30,34,35,37,38,40,43]), but less so if we allow overlap [2,3,5,23].

In some applications, clusters may overlap but not very strongly. We focus mainly on this case and build upon [24].

To understand the complexity, a basic formulation of a clustering with small overlaps can focus on perfect clusterings, i.e., clusters are cliques and all edges of the input graph occur in a cluster. This leads to the Sigma Clique Cover (SCC)

[☆] An extended abstract of this work appears in the *Proceedings of the 49th International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2024)* [26]. This research was initiated during a seminar at TU Wien in 2020/2021. The paper is based on the seminar discussion and on results of Alexander Firbas' master thesis [24]. During the seminar discussion it turned out that one of the results that we wanted to build on, a so-called critical clique lemma, was flawed [25]. We contacted the authors and they corrected the result [1]. NP-hardness of CEVS was proven independently in parallel to our work [5].

[*] Corresponding author.

*E-mail addresses:* afirbas@ac.tuwien.ac.at (A. Firbas), adobler@ac.tuwien.ac.at (A. Dobler), manuel.sorge@ac.tuwien.ac.at (M. Sorge), avilledieu@ac.tuwien.ac.at (A. Villedieu).

[1] Supported by the Vienna Science and Technology Fund (WWTF), Austria under grant 10.47379/ICT22029.

[2] Supported by the Vienna Science and Technology Fund (WWTF), Austria under grant 10.47379/ICT19035.

[3] Partly supported by the Alexander von Humboldt Foundation.

https://doi.org/10.1016/j.dam.2025.01.012

problem, where we seek a covering of the input graph by induced cliques and we want to minimize the total number of times the vertices are covered by the cliques or, in other words, minimize the sum of the sizes of the cliques in the cover (see Section 3 for a formal definition).[4] SCC was previously studied in the context of displaying information in bioinformatics [32] and in combinatorics [18]. To our knowledge, its complexity was not known. We prove that SCC is NP-complete (Theorem 3.5).

An alternative view on overlapping clustering with small overlaps is that of splitting vertices: A vertex split is a graph operation that takes a vertex $v$ and replaces it by two copies such that the union of the neighborhoods of the copies is equal to the neighborhood of the original vertex $v$. Given a graph and an integer $k$, we may then ask to perform at most $k$ vertex-splitting operations in order to obtain a cluster graph (a disjoint union of cliques). The cliques in the obtained cluster graph then correspond to the clusters in the original graph. This yields the Cluster Vertex Splitting (CVS) problem. To our knowledge, CVS has so far not been studied directly, but we feel it is a very natural simple formulation of clustering with small overlaps and, indeed, it is a simpler variant of the well-studied Cluster Editing With Vertex Splitting problem which we discuss below.

In Section 4 we show that SCC and CVS are indeed equivalent (Lemma 4.3), and thus both are NP-complete. On the positive side, we show that CVS is fixed-parameter tractable with respect to the number $k$ of allowed splits, that is, it can be solved in $f(k) \cdot n^{O(1)}$ time where $f$ is a computable function and $n$ the number of vertices. Indeed, in Section 5 we show a stronger result, namely, that CVS admits an $O(k)$-vertex problem kernel, that is, we may produce with polynomial processing time an equivalent instance that contains at most $3k + 3$ vertices (see Theorem 5.7). This result relies on an analysis of the structure of the so-called critical cliques of the input graph. Informally, a critical clique is an induced clique in the input graph with vertex set $C$ such that all vertices in $C$ have pairwise the same neighbors outside of $C$ and such that there is no critical clique that strictly contains $C$.[5]

The Cluster Editing With Vertex Splitting (CEVS) problem [3] is closely related to the above two problems. The difference is that the underlying clustering model allows the clusters to be imperfect, that is, the clusters may miss a small number of edges and there may be a small number of edges that are not contained in any cluster. More precisely, in CEVS we are given a graph $G$ and an integer $k$ and we want to obtain a cluster graph from $G$ by at most $k$ modifications. As modifications we are allowed to split vertices and to add or delete edges. Apart from earlier and applied work [2,3,6], CEVS has recently been shown to be NP-complete, to be solvable in $O(2^{9k \log k} + n + m)$ time, and to admit a problem kernel with $6k$ vertices [1].[6]

Here, building on the structural observations about CVS and SCC above, we contribute a formal proof for a characterization of solutions for CEVS as special covers of the input graph (Section 6). Using this characterization together with a so-called critical-clique lemma [1], we give a simple alternative NP-hardness proof for CEVS (Theorem 6.4). The critical-clique lemma essentially shows that there is always an optimal cover such that the cover sets respect the critical cliques in the input graph. Interestingly, not necessarily every optimal cover has this property: We give examples of instances and optimal covers that do not respect the critical cliques.

*Further Related Work.* The problems we study are related to two problems with similar context but that correspond to clusterings without overlap. First, there is the well-researched Cluster Editing (CE) problem, in which we want to add or delete a minimum number of edges in a given graph to obtain a cluster graph [8–13,17,23,29,30,34,35,37,38,40,43]. For instance, it is known that CE is NP-hard, fixed-parameter tractable, and admits a $2k$-vertex problem kernel. CE is one of a broad range of so-called edge-modification problems, see Crespelle et al. [14] for a recent survey.

Second, we have Edge Clique Cover (ECC), wherein we look for covering all edges of a graph with at most some given number $s$ of induced cliques. Here, it is known that covering all edges of a given graph with at most $s$ induced cliques can be done in $2^{O(4^s)} + n^{O(1)}$ time [31], but not substantially faster than that [16].

CE has been extended to a variant modeling overlapping clustering [23], where, instead of trying to get a cluster graph, we modify the edges to obtain a graph in which at most a bounded number of maximal cliques overlap in each vertex. If we can split a bounded number of vertices to obtain a cluster graph, then in the input graph indeed few maximal cliques overlap, but not necessarily vice versa.

Vertex splitting as a graph operation has also appeared in other contexts [20–22,42], such as splitting vertices towards obtaining a planar graph. Systematic investigation into the complexity of vertex-splitting towards obtaining a fixed graph property began only recently [7,27].

*Organization.* We will establish the following chain of polynomial-time reductions, based on the classical NP-hard Node Clique Cover (NCC) problem [36]:

$$\begin{aligned} \text{Node Clique Cover} &\leq_{\mathrm{P}} \text{Sigma Clique Cover} \\ &\leq_{\mathrm{P}} \text{Cluster Vertex Splitting} \\ &\leq_{\mathrm{P}} \text{Cluster Editing With Vertex Splitting.} \end{aligned}$$

[4] Note that this is a different optimization goal than the one of the well-studied Edge Clique Cover problem, where we seek a covering of all edges with a minimum number of induced cliques.

[5] Alternatively, a critical clique is a maximal set of pairwise true twins.

[6] These results were developed independently and in parallel to our work.

We give the first reduction in Section 3, the second in Section 4, and the last in Section 6. The informal definitions of these problems have been given above, the formal definitions will be given in the corresponding sections. The problem kernel is shown in Section 5. CEVS and the critical-clique lemma is treated in Section 6.

## 2. Preliminaries

For a positive integer $n \in \mathbb{N}$ we use $[n]$ to denote $\{1, 2, \ldots, n\}$. For a set $X$, we denote by $\mathcal{P}(X)$ its power set. Moreover, for a family of sets $\mathcal{X}$, we write $\bigcup \mathcal{X}$ for the union of all set members of $\mathcal{X}$, that is, $\bigcup_{X \in \mathcal{X}} X$. We denote disjoint unions by $\mathbin{\dot\cup}$. Unless explicitly mentioned otherwise, all graphs are undirected and without parallel edges or self-loops. In a graph $G$ with vertex set $V(G)$ and edge set $E(G)$, we denote the neighborhood of a vertex $v \in V(G)$ by $N_G(v)$ and its closed neighborhood by $N_G(v) \cup \{v\}$. If the graph $G$ is clear from the context, we omit the subscript $G$. For $V' \subset V(G)$, we write $G[V']$ for the graph induced by the vertices $V'$. For $u, v \in V(G)$ we write $uv$ as a shorthand for $\{u, v\}$, $G - v$ for $G[V \setminus \{v\}]$, and $d_G(v)$ for $|N_G(v)|$. The graph $K_n$ is the complete graph on $n$ vertices. For a graph $H$, we write $H \simeq G$ if $H$ is isomorphic to $G$, and $H \prec G$ if $H$ is an induced subgraph of $G$. A *cluster graph* is a graph in which every connected component is a clique. Equivalently, a cluster graph does not contain a path $P_3$ with three vertices as an induced subgraph. A *vertex split* operation applied to a graph $G = (V, E)$ and $u \in V$ results in a graph $G' = (V', E')$ such that $V' = V \setminus \{u\} \cup \{v, w\}$ with $v, w \notin V$, and $E'$ is obtained from $E$ by making each vertex adjacent to $u$ adjacent to at least one of $v$ and $w$; that is, $N_{G'}(v) \cup N_{G'}(w) = N_G(u)$.

Some of our results are in terms of parameterized complexity [15,19,28,41]. In a *parameterized problem*, each instance $x \in \Sigma^*$ is equipped with a parameter $k \in \mathbb{N}$. Such a problem is *fixed-parameter tractable* if it can be solved in $f(k) \cdot n^{O(1)}$ time, where $f$ is a computable function and $n$ the input size. A parameterized problem has a *problem kernel* if there is a polynomial-time self-reduction such that in the resulting instances the size is bounded by $g(k)$, where $g$ is a computable function and $k$ is the parameter. The function $g$ is also called the *size* of the problem kernel.

## 3. NP-completeness of SIGMA CLIQUE COVER

To start, we will fix some notation. Leading up to the formulation of the sigma clique cover problem, we first define the notion of a sigma clique cover:

**Definition 3.1.** Let $G$ be a graph. Then, $\mathcal{C} \subseteq \mathcal{P}(V)$ is called a *sigma clique cover* of $G$ if

1. $G[C]$ is a clique for all $C \in \mathcal{C}$ and
2. for each $e \in E(G)$, there is $C \in \mathcal{C}$ such that $e \in E(G[C])$, that is, all edges of $G$ are "covered" by some clique of $\mathcal{C}$.

The *weight* of a sigma clique cover $\mathcal{C}$ is denoted by $\mathrm{wgt}(\mathcal{C})$, where

$$\mathrm{wgt}(\mathcal{C}) := \sum_{C \in \mathcal{C}} |C|.$$

Now, we can formulate the associated decision problem:

SIGMA CLIQUE COVER (SCC)

**Input:** A tuple $(G, s)$, where $G$ is a graph and $s \in \mathbb{N}$.
**Question:** Is there a sigma clique cover $\mathcal{C}$ of $G$ with $\mathrm{wgt}(\mathcal{C}) \leq s$?

Note that SCC is not equivalent to the well-studied EDGE CLIQUE COVER problem, whose optimization goal is to minimize $|\mathcal{C}|$ rather than $\mathrm{wgt}(\mathcal{C})$. To show that SCC is NP-hard, we reduce from the NODE CLIQUE COVER problem. Analogous to the case of SCC, to define said problem formally, we first need to introduce the notion of a node clique cover:

**Definition 3.2.** Let $G$ be a graph. Then, $\mathcal{C} \subseteq \mathcal{P}(V)$ is called a *node clique cover* of $G$ if

1. $G[C]$ is a clique for all $C \in \mathcal{C}$ and
2. for each $v \in V(G)$, there is $C \in \mathcal{C}$ such that $v \in V(G[C])$, that is, all vertices of $G$ are "covered" by some clique $C \in \mathcal{C}$.

The *size* of a node clique cover $\mathcal{C}$ is denoted by $|\mathcal{C}|$.

With this, we formulate the NP-hard [36] NODE CLIQUE COVER problem:

NODE CLIQUE COVER (NCC)

**Input:** A tuple $(G, k)$, where $G$ is a graph and $k \in \mathbb{N}$.
**Question:** Is there a node clique cover $\mathcal{C}$ of $G$ with $|\mathcal{C}| \leq k$?

Note that the SCC and NCC problem are similar on a superficial level, but differ in two core aspects: Firstly, the notion of a sigma clique cover mandates that all edges be covered, in comparison to node clique covers, where all vertices need to be covered, and secondly, the "difficulty" of the SCC problem lies in minimizing a cumulative weight, in comparison to the NCC problem, where it is the number of cliques to be minimized. See Fig. 1 for a contrasting example.

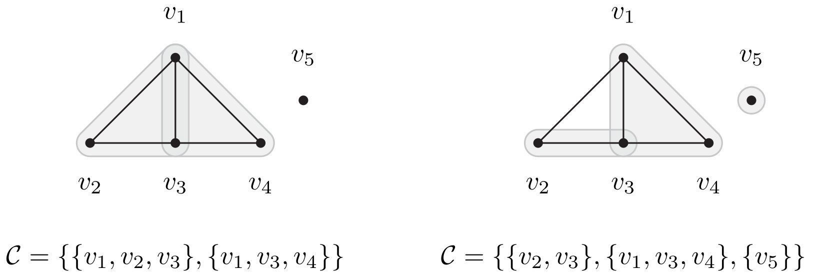
$v_1$
$v_5$
$v_2$
$v_3$
$v_4$
$\mathcal{C} = \{\{v_1, v_2, v_3\}, \{v_1, v_3, v_4\}\}$
$v_1$
$v_5$
$v_2$
$v_3$
$v_4$
$\mathcal{C} = \{\{v_2, v_3\}, \{v_1, v_3, v_4\}, \{v_5\}\}$

**Fig. 1.** A graph with its unique minimum-weight sigma clique cover (left) and one of its multiple minimum-cardinality node clique covers (right).

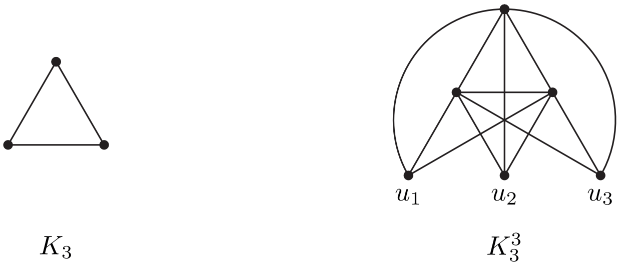

**Fig. 2.** $K_3$ and $K_3^3$, illustrating Definition 3.3.

To formulate our reduction from NCC to SCC, we introduce notation to extend a graph with independent universal vertices. See Fig. 2 for an example of Definition 3.3.

**Definition 3.3.** Let $G = (V, E)$ be a graph and $\ell \in \mathbb{N}$. Using a set $\{u_1, \ldots, u_\ell\}$ of $\ell$ new vertices called *universal vertices*, we construct a new graph $G^\ell$ with

$$G^\ell := (V \cup \{u_1, \ldots, u_\ell\}, E \cup \{u_i v \mid 1 \le i \le \ell, v \in V\}) .$$

Note that universal vertices themselves are not adjacent to each other. Informally, the main intuition behind our reduction from NCC is to add a sufficient number of universal vertices to the instances of NCC such that, concerning the derived instances of SCC, it will be "combinatorially favorable" to select cliques that contain a universal vertex. Refer to Fig. 3 for an example of the reduction.

**Lemma 3.4.** *Let $G = (V, E)$ be a graph and $\ell := 2|E| + 1$. Then, $(G, s)$ is a positive instance of* NCC *if and only if $\left(G^\ell, \ell\,(|V| + s + 1) - 1\right)$ is a positive instance of* SCC.

**Proof.** $(\Rightarrow)$: Let $\mathcal{C}$ be a node clique cover of $G$ with $|\mathcal{C}| \le s$. Without loss of generality, we assume that $\mathcal{C}$ is a partition of $V$ — for otherwise if there are distinct $C', C'' \in \mathcal{C}$ with $C' \cap C'' \ne \emptyset$, then $\mathcal{C}' := (\mathcal{C} \setminus \{C'\}) \cup \{C' \setminus C''\}$ is a node clique cover of $G$ with $|\mathcal{C}'| = |\mathcal{C}|$ and the number of nodes that are contained in more than one clique is strictly less. Thus, applying this observation a sufficient number of times always yields a partition of $V$.

Let

$$\mathcal{A} := \{C \cup \{u_i\} \mid C \in \mathcal{C}, 1 \le i \le \ell\} \text{ and}$$
$$\mathcal{B} := \{\{v_1, v_2\} \mid v_1 v_2 \in E\} .$$

We claim that $\mathcal{A} \uplus \mathcal{B}$ is a sigma clique cover of $G^\ell$ with

$$\text{wgt}(\mathcal{A} \uplus \mathcal{B}) \le \ell(|V| + s + 1) - 1.$$

First, we verify that $\mathcal{A} \uplus \mathcal{B}$ conforms to Definition 3.1, that is, it indeed is a sigma clique cover of $G^\ell$. To that end, we begin by verifying that $G[C]$ is a clique for all $C \in \mathcal{A} \uplus \mathcal{B}$. By construction, we need to differentiate two cases: Firstly, let $C \cup \{u_i\} \in \mathcal{A}$. Since $G[C]$ is a clique, $E \subseteq E(G^\ell)$ and $\forall v \in V : u_i v \in E(G^\ell)$, it follows that $G^\ell[C \cup \{u_i\}]$ is also a clique. Secondly, let $\{v_1, v_2\} \in \mathcal{B}$. Similarly, since $G[\{v_1, v_2\}] \simeq K_2$ and $E \subseteq E(G^\ell)$, we have $G^\ell[\{v_1, v_2\}] \simeq K_2$.

Now, we prove that all edges of $G^\ell$ are "covered" by $\mathcal{A} \uplus \mathcal{B}$. Two cases need to be verified: Consider any $v_1 v_2 \in E$, i.e., those edges that are "inherited" from $G$ to $G^\ell$. We see that $\{v_1, v_2\} \in \mathcal{B}$ by definition. Furthermore, consider any $u_i v \in E(G^\ell) \setminus E$, i.e., those edges added to $G$ in the construction of $G^\ell$. Observe that since $\exists C \in \mathcal{C}$ with $v \in C$, we have $\{u_i, v\} \subseteq C \cup \{u_i\} \in \mathcal{A}$.

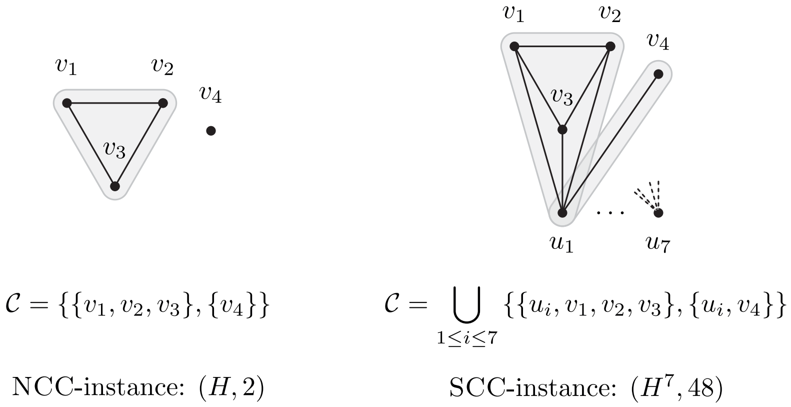
$v_1$
$v_2$
$v_3$
$v_4$
$v_1$
$v_2$
$v_3$
$v_4$
$u_1$
$u_7$
$\dots$
$\mathcal{C} = \{\{v_1, v_2, v_3\}, \{v_4\}\}$
$\mathcal{C} = \bigcup_{1 \le i \le 7} \{\{u_i, v_1, v_2, v_3\}, \{u_i, v_4\}\}$
NCC-instance: $(H, 2)$
SCC-instance: $(H^7, 48)$

**Fig. 3.** Example for our reduction from NCC to SCC. On the left, we see a NCC-instance, and on the right, we see the corresponding SCC-instance (only one universal node and its associated cliques are fully drawn). In both cases, a certificate is marked in the input graph, as well as stated explicitly.

We conclude that $\mathcal{A} \mathbin{\dot\cup} \mathcal{B}$ is a sigma clique cover of $G^\ell$ and proceed to verify that the claimed bound on the weight holds. By definition of $\mathcal{A}$, we obtain

$$\begin{aligned} \operatorname{wgt}(\mathcal{A}) &= \sum_{\substack{C \in \mathcal{C}, \\ 1 \le i \le \ell}} |C \cup \{u_i\}| && \\ &= \ell \sum_{C \in \mathcal{C}} |C| + 1 && \{u_1, \dots, u_\ell\} \cap C = \emptyset \textit{ for all } C \in \mathcal{C} \\ &= \ell \left( |\mathcal{C}| + \sum_{C \in \mathcal{C}} |C| \right) && \\ &= \ell \left( |\mathcal{C}| + |V| \right). && \mathcal{C} \textit{ is a partition of } V \end{aligned}$$

Clearly, $\operatorname{wgt}(\mathcal{B}) = 2|E| = \ell - 1$. Now, using $\mathcal{A} \cap \mathcal{B} = \emptyset$, we derive

$$\begin{aligned} \operatorname{wgt}(\mathcal{A} \mathbin{\dot\cup} \mathcal{B}) &= \operatorname{wgt}(\mathcal{A}) + \operatorname{wgt}(\mathcal{B}) \\ &= \ell(|\mathcal{C}| + |V| + 1) - 1 \\ &\le \ell(s + |V| + 1) - 1. && |\mathcal{C}| \le s. \end{aligned}$$

Thus, the forward direction of the proof is established.

($\Leftarrow$): Let $\mathcal{S}$ be a sigma clique cover of $G^\ell$ with

$$\operatorname{wgt}(\mathcal{S}) \le \ell(|V| + s + 1) - 1,$$

and let

$$u^* \in \operatorname*{argmin}_{u \in \{u_1, \dots, u_\ell\}} \operatorname{wgt}\left(\{C \in \mathcal{S} \mid u \in C\}\right),$$

$$\mathcal{X} := \left\{C \in \mathcal{S} \mid u^* \in C\right\}, \text{ and}$$

$$\mathcal{N} := \left\{C \setminus \{u^*\} \mid C \in \mathcal{X}\right\}.$$

We claim that $\mathcal{N}$ is a node clique cover of $G$ with $|\mathcal{N}| \le s$.

First, we verify that $\mathcal{N}$ conforms to Definition 3.2, i.e., it indeed is a node clique cover of $G$. Clearly, $G[C]$ is a clique for all $C \in \mathcal{N}$. It remains to verify that all vertices of $G$ are "covered" by $\mathcal{N}$: Let $v \in V$. Since $\mathcal{S}$ is a sigma clique cover of $G^\ell$ and $vu^* \in E(G^\ell)$, there is some $C \in \mathcal{S}$ s.t. $\{v, u^*\} \subseteq C$. It immediately follows that $v \in C \setminus \{u^*\} \in \mathcal{N}$.

Second, we establish that $|\mathcal{N}| \le s$. To that end, first, we derive $\operatorname{wgt}(\mathcal{X}) \le |V| + s$. Towards a contradiction, suppose that $\operatorname{wgt}(\mathcal{X}) \ge |V| + s + 1$. Observe that since no $C \in \mathcal{S}$ can contain two different universal nodes of $G^\ell$ we get

$$
\begin{aligned}
\operatorname{wgt}(\mathcal{S}) &\geq \sum_{u \in \{u_1,\ldots,u_\ell\}} \operatorname{wgt}\left(\{C \in \mathcal{S} \mid u \in C\}\right) \\
&\geq \ell \cdot \operatorname{wgt}(\mathcal{X}) && \text{choice of } u^*, \mathcal{X} \\
&\geq \ell(|V| + s + 1) && \operatorname{wgt}(\mathcal{X}) \geq |V| + s + 1 \\
&= \ell(|V| + s + 1) - 1 + 1 \\
&\geq \operatorname{wgt}(\mathcal{S}) + 1. && \ell(|V| + s + 1) - 1 \geq \operatorname{wgt}(\mathcal{S})
\end{aligned}
$$

In total, this yields $\operatorname{wgt}(\mathcal{S}) \geq \operatorname{wgt}(\mathcal{S}) + 1$, hence $\operatorname{wgt}(\mathcal{X}) \leq |V| + s$.

Now, towards the final contradiction, suppose $|\mathcal{N}| \geq s + 1$. We obtain

$$
\begin{aligned}
\operatorname{wgt}(\mathcal{X}) &= \sum_{\substack{C \in \mathcal{S}, \\ u^* \in C}} |C| \\
&= \sum_{C \in \mathcal{N}} \left|C \cup \{u^*\}\right| && \text{definition of } \mathcal{N} \\
&= |\mathcal{N}| + \sum_{C \in \mathcal{N}} |C| && u^* \notin C \text{ for all } C \in \mathcal{N} \\
&\geq s + 1 + \sum_{C \in \mathcal{N}} |C| && |\mathcal{N}| \geq s + 1 \\
&= s + 1 + \sum_{v \in V} |\{C \in \mathcal{N} \mid v \in C\}| && \text{double counting principle} \\
&\geq s + 1 + |V|. && \mathcal{N} \text{ covers } V
\end{aligned}
$$

Thus, we have derived both $\operatorname{wgt}(\mathcal{X}) \geq |V| + s + 1$ and $\operatorname{wgt}(\mathcal{X}) \leq |V| + s$, a contradiction. Hence, we conclude that $|\mathcal{N}| \leq s$. □

Using this preliminary work, the $\mathtt{NP}$-completeness proof is straightforward:

**Theorem 3.5.** Sigma Clique Cover *is* $\mathtt{NP}$*-complete.*

**Proof.** Lemma 3.4 directly yields a polynomial-time many-one reduction from NCC to SCC, i.e., deciding an instance $(G, s)$ of NCC is equivalent to deciding the instance $\left(G^\ell, \ell\,(|V| + s + 1) - 1\right)$ of SCC where $\ell := 2|E| + 1$. Because NCC is $\mathtt{NP}$-hard [36], so is SCC. Observe that SCC $\in \mathtt{NP}$, since a certificate for SCC can clearly be guessed and checked in polynomial-time. Consequently, we conclude that SCC is $\mathtt{NP}$-complete. □

## 4. NP-completeness of cluster vertex splitting

We will now build upon the $\mathtt{NP}$-completeness of SCC and attend to the $\mathtt{NP}$-completeness proof of CVS. The formal problem definition of the corresponding decision problem is given below.

Cluster Vertex Splitting (CVS)

**Input:** A tuple $(G, k)$, where $G$ is a graph and $k \in \mathbb{N}$.
**Question:** Is there a sequence of at most $k$ vertex splits that transforms $G$ into a cluster graph?

The reduction will be accomplished in a multi-step manner: We begin with introducing two lemmata, Lemmas 4.1 and 4.2, used to prove the forward and backward direction of Lemma 4.3, respectively. Then, in Lemma 4.3, we establish a close correspondence between instances of SCC and instances of CVS. Finally, in Theorem 4.4, we use said correspondence to show that CVS is $\mathtt{NP}$-complete.

Lemma 4.1 essentially states the following: Consider a graph $G'$ that has a sigma clique cover $\mathcal{C}'$. If we merge two non-adjacent vertices $v$ and $w$ in $G'$ into a vertex we call $u$, that is, we perform a reverse vertex split, we obtain a new graph, $G$. Then, we can replace each occurrence of $v$ or $w$ in $\mathcal{C}'$ with $u$ and obtain a sigma clique cover $\mathcal{C}$ of the same weight for $G$. Note that the "overlap" of $\mathcal{C}$, $\operatorname{wgt}(\mathcal{C}) - |V(G)|$, is one more than the "overlap" of $\mathcal{C}'$, $\operatorname{wgt}(\mathcal{C}') - |V(G')|$.

**Lemma 4.1.** *Let $G = (V, E)$ be a graph and let $G' = (V', E')$ be obtained from $G$ by splitting $u \in V$ into $v, w \in V'$. If $\mathcal{C}'$ is a sigma clique cover of $G'$, then there exists a sigma clique cover $\mathcal{C}$ of $G$ with $\operatorname{wgt}(\mathcal{C}) = \operatorname{wgt}(\mathcal{C}')$.*

**Proof.** Using

$$
f(C') := \begin{cases} (C' \setminus \{v, w\}) \cup \{u\} & \text{if } C' \cap \{v, w\} \neq \emptyset \\ C' & \text{otherwise} \end{cases}
$$

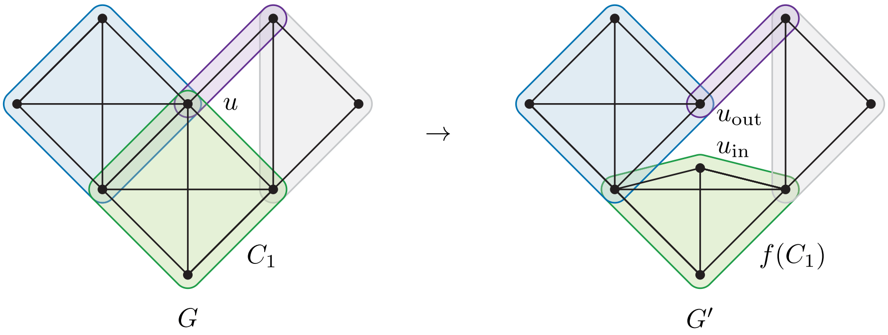

**Fig. 4.** On the left, a graph $G$ with a sigma clique cover $\mathcal{C}$ is depicted. The clique $C_1 \in \mathcal{C}$ is marked in green. On the right, a graph $G'$, obtained by splitting $u$ into $u_{\text{in}}$ and $u_{\text{out}}$, is drawn. Additionally, a sigma clique cover $\mathcal{C}'$ of $G'$ is shown. The clique $C_1$ of $\mathcal{C}$ was "pulled away" to form $f(C_1)$ in the derived $\mathcal{C}'$, creating a sigma clique cover of "decreased overlap". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

we define

$$\mathcal{C} := \left\{ f(C') \mid C' \in \mathcal{C}' \right\}.$$

Note that $f$ gives a bijection from $\mathcal{C}'$ to $\mathcal{C}$. We claim that $\mathcal{C}'$ satisfies the conditions of this lemma. First, we establish that $\mathcal{C}$ is a sigma clique cover of $G$ by verifying the two conditions of Definition 3.1. We begin by proving that all $C \in \mathcal{C}$ induce cliques in $G$.

Let $C \in \mathcal{C}$. Assume $f^{-1}(C) = C$. Observe that $C \cap \{v, w\} = \emptyset$. This implies that $G[C] = G'[C]$. Hence, since $G'[C]$ is a clique, so is $G[C]$.

Conversely, assume $f^{-1}(C) \neq C$. Without loss of generality, we assume that $v \in f^{-1}(C)$ and $w \notin f^{-1}(C)$, since $f^{-1}(C)$ cannot both contain $v$ and $w$ by the semantics of vertex splitting.

Towards the goal of showing $v_1v_2 \in E$, let $v_1, v_2 \in C$ with $v_1 \neq v_2$.

***Case*** $\{v_1, v_2\} \cap \{u\} = \emptyset$**:** We get that $v_1v_2 \in E$ if and only if $v_1v_2 \in E'$ by the way our vertex split was defined. From $\{v_1, v_2\} \subseteq f^{-1}(C)$, our assumption that $f^{-1}(C) \in \mathcal{C}'$ is a sigma clique cover of $G'$ and the correspondence just established, it follows that $v_1v_2 \in E$.

***Case*** $\{v_1, v_2\} \cap \{u\} \neq \emptyset$**:** Without loss of generality, assume $v_1 = u$. Since $\{v, v_2\}$ is a subset of $f^{-1}(C)$, again invoking that $\mathcal{C}'$ is a sigma clique cover of $G'$ to derive $vv_2 \in E'$ and $N_{G'}(v) \subseteq N_G(u)$, it follows that $uv_2 \in E$.

Now, we prove the second property, that is, all edges of $G$ are covered by $\mathcal{C}$. Again, let $v_1, v_2 \in C$ with $v_1 \neq v_2$.

***Case*** $\{v_1, v_2\} \cap \{u\} = \emptyset$**:** This edge is not affected by the split, therefore $v_1v_2 \in E'$, enabling us to choose $C' \in \mathcal{C}'$ such that $\{v_1, v_2\} \subseteq C'$. Thus $\{v_1, v_2\} \subseteq C' \setminus \{v, w\} \subseteq f(C') \in \mathcal{C}$.

***Case*** $\{v_1, v_2\} \cap \{u\} \neq \emptyset$**:** Without loss of generality, assume $v_1 = u$. By the semantics of our split, $vv_2 \in E'$ or $wv_2 \in E'$ must hold. Without loss of generality, assume the former. By the assumption of $\mathcal{C}'$ being a sigma clique cover of $G'$, we can choose $C'$ such that $\{v, v_2\} \subseteq C' \in \mathcal{C}'$. Thus, we find that $\{u, v_2\} \subseteq f(C') \in \mathcal{C}$.

Therefore, $\mathcal{C}$ is a sigma clique cover of $G$. Finally, observe that $f$ ranging over $\mathcal{C}'$ does not change the cardinality of any image it maps, implying that $\text{wgt}(\mathcal{C}) = \text{wgt}(\mathcal{C}')$. □

Now, we tend to the other direction. In essence, Lemma 4.2 states the following: Consider a graph $G$ that has a sigma clique cover $\mathcal{C}$ of "overlap" at most $\alpha \in \mathbb{N}$, that is, $\text{wgt}(\mathcal{C}) - |V(G)|$. If $\alpha$ is zero, then $G$ evidently is a cluster graph. Otherwise, there is a vertex $u$ covered by at least two cliques, $C_1$ and $C_2$. Then, we can define a vertex split acting on $u$ that "pulls the clique $C_1$ away from the other cliques of $\mathcal{C}$" while leaving the cliques of the sigma clique cover intact. One of $u$'s descendants is then only covered by a single clique. Refer to Fig. 4 for an illustration. Consequently, we obtain a graph $G'$ that has a sigma clique cover of the same weight, but with an "overlap" decremented by one.

**Lemma 4.2.** *Let $G = (V, E)$ be a graph without isolated vertices and let $\mathcal{C}$ be a sigma clique cover of $G$ with $\text{wgt}(\mathcal{C}) \leq |V| + \alpha \in \mathbb{N}$ as well as $|C| > 1$ for all $C \in \mathcal{C}$. Then, either $G$ is already a cluster graph or there is $u \in V$ such that $u$ can be split in $G$ to obtain $G' = (V', E')$ admitting a sigma clique cover $\mathcal{C}'$ satisfying*

1. $\text{wgt}(\mathcal{C}') \leq |V'| + \alpha - 1$,
2. $|C'| > 1$ *for all* $C' \in \mathcal{C}'$, *and*
3. $G'$ *does not contain isolated vertices.*

**Proof.** If $G$ is not already a cluster graph, there must exist $C_1 \neq C_2 \in \mathcal{C}$ such that $C_1 \cap C_2 \neq \emptyset$. In this case, let $u \in C_1 \cap C_2$.

We define $G' = (V', E')$ as the graph that is obtained when $u$ is split into the two vertices $u_{\text{in}}$ and $u_{\text{out}}$ obeying:

$$\begin{aligned} &N_{G'}(u_{\text{in}}) := N_G(u) \cap C_1, \\ &N_{G'}(u_{\text{out}}) := (N_G(u) \setminus C_1) \cup \{v \in N_G(u) \cap C_1 \mid \exists C \in \mathcal{C} \setminus \{C_1\} : u, v \in C\}. \end{aligned}$$

Furthermore, using the map

$$f(C) := \begin{cases} (C \setminus \{u\}) \cup \{u_{\text{in}}\} & \text{if } C = C_1 \\ (C \setminus \{u\}) \cup \{u_{\text{out}}\} & \text{if } u \in C \wedge C \neq C_1 \\ C & \text{otherwise} \end{cases}$$

we can define

$$\mathcal{C}' := \{f(C) \mid C \in \mathcal{C}\}.$$

Note that $f$ gives a bijection between $\mathcal{C}$ and $\mathcal{C}'$; thus $f^{-1}(\cdot)$ will be used to denote a single well-defined element in what follows.

Intuitively, this split corresponds to "pulling out" the vertex $u$ creating $u_{\text{in}}$, only keeping the part of $u$'s neighborhood contained in $C_1$, so that $u_{\text{in}}$ will only be contained in a single clique $f(C_1)$ in the derived sigma clique cover $\mathcal{C}'$ and letting $u_{\text{out}}$ inherit the rest of the neighborhood, plus a select set of vertices already neighbors of $u_{\text{in}}$, as to not destroy any cliques of $\mathcal{C} \setminus \{C_1\}$. See Fig. 4 for an example. We claim that $\mathcal{C}'$ is a sigma clique cover of $G'$ and that Conditions 1–3 of this lemma are met. To show the former, we need to verify the two conditions of Definition 3.1. We start with the first condition, that is, we verify that all $C' \in \mathcal{C}'$ induce cliques in $G'$:

Let $C' \in \mathcal{C}'$. Since $|C'| < 2$ is impossible, we select arbitrary $v_1, v_2 \in C'$ such that $v_1 \neq v_2$. We denote the intersection of $\{u_{\text{in}}, u_{\text{out}}\}$ and $\{v_1, v_2\}$ by $I$ and enumerate all arising cases:

**$I = \emptyset$:** We have $v_1v_2 \in E$ since $\{v_1, v_2\} \subseteq f^{-1}(C')$, further implying $v_1v_2 \in E'$, because this edge was not affected by the splitting operation.

**$I = \{u_{\text{in}}\}$:** Without loss of generality, assume $v_1 = u_{\text{in}}$. Let $C := f^{-1}(C') \in \mathcal{C}$. Since $\{v_2, u\} \subseteq C$, we obtain $v_2 \in N_G(u)$ and $v_2 \in C$. Thus, $v_2 \in N_{G'}(u_{\text{in}})$, implying $v_1v_2 \in E'$ by construction.

**$I = \{u_{\text{out}}\}$:** Without loss of generality, assume $v_1 = u_{\text{out}}$. Observe that this yields $v_2 \in N_G(u)$. In the case that $v_2 \notin C_1$ it holds that $v_2 \in N_G(u) \setminus C_1 \subseteq N_{G'}(u_{\text{out}})$, thus $u_{\text{out}}v_2 = v_1v_2 \in E'$.

Otherwise, if $v_2 \in C_1$, observe that $\{u, v_2\} \in f^{-1}(C')$. This implies $v_2 \in N_G(u)$, and using $v_2 \in C_1$, we get $v_2 \in N_G(u) \cap C_1$. Note also that $f^{-1}(C') \neq C_1$ by definition of $f$. Thus, $v_2 \in \left\{v \in N_G(u) \cap C_1 \mid \exists C'' \in \mathcal{C} \setminus C_1 : u, v \in C''\right\} \subseteq N_{G'}(u_{\text{out}})$ is witnessed by $f^{-1}(C')$. Hence, we have $u_{\text{out}}v_2 = v_1v_2 \in E'$.

**$I = \{u_{\text{in}}, u_{\text{out}}\}$:** Contradiction to the definition of the split yielding $G'$.

Now, we proceed with the second condition, demanding that all edges of $G'$ be covered by $\mathcal{C}'$: Let $v_1v_2 \in E'$. Again, we denote the intersection of $\{u_{\text{in}}, u_{\text{out}}\}$ and $\{v_1, v_2\}$ by $I$ and enumerate all arising cases:

**$I = \emptyset$:** Since this case mandates that $v_1v_2 \in E$, by assumption of $\mathcal{C}$ being a sigma clique cover of $G$, there exists $C \in \mathcal{C}$ such that $\{v_1, v_2\} \subseteq C$. By definition of $f$, it must also hold that $\{v_1, v_2\} \subseteq f(C)$.

**$I = \{u_{\text{in}}\}$:** Without loss of generality, assume $v_1 = u_{\text{in}}$. Because $v_2 \in N_{G'}(u_{\text{in}}) \subseteq C_1$, and $v_2 \neq u$, we know that $v_2 \in f(C_1)$. Furthermore, because $u_{\text{in}} \in f(C_1)$ by definition of $f$, we have $u_{\text{in}}v_2 \in E(G'[f(C_1)])$.

**$I = \{u_{\text{out}}\}$:** Without loss of generality, assume $v_1 = u_{\text{out}}$. It holds that $v_2 \in N_{G'}(u_{\text{out}})$. As $N_{G'}(u_{\text{in}})$ is defined as the union of two sets, we distinguish two cases: Firstly, assume $v_2 \in N_G(u) \setminus C_1$. By definition of the vertex split at hand, we have $uv_2 \in E$. Using the assumption that $\mathcal{C}$ is a sigma clique cover of $G$, there is $C^* \in \mathcal{C} \setminus \{C_1\}$ with $\{u, v_2\} \subseteq C^*$. By the second case of the definition of $f$, it thus follows that $\{u_{\text{out}}, v_2\} \subseteq f(C^*)$.

Secondly, assume $v_2 \in \{v \in N_G(u) \cap C_1 \mid \exists C \in \mathcal{C} \setminus \{C_1\} : u, v \in C\}$. This yields that there is $C^* \in \mathcal{C} \setminus \{C_1\}$ with $\{u, v_2\} \subseteq C^*$ and therefore, by the argument employed in the previous case, we have $\{u_{\text{out}}, v_2\} \subseteq f(C^*)$.

**$I = \{u_{\text{in}}, u_{\text{out}}\}$:** Contradiction to the definition of the split yielding $G'$.

Thus, $\mathcal{C}'$ indeed is a sigma clique cover of $G'$. Next, we derive Conditions 1–3.

*Condition 1.* Since $|C| = |f(C)|$ for all $C \in \mathcal{C}$, we have

$$\begin{aligned} \operatorname{wgt}(\mathcal{C}') &= \operatorname{wgt}(\mathcal{C}) \\ &\le |V| + \alpha && \Big\rangle \textit{by assumption} \\ &\le |V'| + \alpha - 1. && \Big\rangle |V| = |V'| - 1. \end{aligned}$$

*Condition 2.* Observe that $f$ preserves the cardinality of mapped sets, and that $|C| > 1$ for all $C \in \mathcal{C}$. Thus, Condition 2 follows immediately.

*Condition 3.* Towards a contradiction, suppose $G'$ contains an isolated vertex $v \in V'$. As the vertex degree of all vertices, except those of $u_{\text{in}}$ and $u_{\text{out}}$, are necessarily inherited from $G$ by the vertex split, we must have either $N_{G'}(u_{\text{in}}) = \emptyset$ or $N_{G'}(u_{\text{out}}) = \emptyset$.

Suppose $N_{G'}(u_{\text{in}}) = \emptyset$. Since $u \in C_1$ and $|C_1| > 1$, there exists $v_2 \neq u$ with $v_2 \in C_1$. Since $G[C_1]$ is a clique, we get $v_2 \in N_G(u)$. Therefore, $v_2 \in N_G(u) \cap C_1 = N_{G'}(u_{\text{in}})$, contradicting $N_{G'}(u_{\text{in}}) = \emptyset$.

Now, suppose $N_{G'}(u_{\text{out}}) = \emptyset$. Invoking the same argument as in the last case substituting $C_2$ for $C_1$, we derive $v_2 \in N_G(u)$. First, suppose $v_2 \in C_1$. Using $C_2$ as witness, we obtain

$$v_2 \in \left\{v' \in N_G(u) \cap C_1 \mid \exists C \in \mathcal{C} \setminus \{C_1\} : u, v' \in C\right\} \subseteq N_{G'}(u_{\text{out}}),$$

which contradicts $N_{G'}(u_{\text{out}}) = \emptyset$.

Now, suppose the contrary, that is, $v_2 \notin C_1$. We derive

$$v_2 \in N_G(u) \setminus C_1 \subseteq N_{G'}(u_{\text{out}}),$$

which again is a contradiction to $N_{G'}(u_{\text{out}}) = \emptyset$.

Thus, our initial assumption that $G'$ contains an isolated vertex $v \in V'$ is invalid. □

With this groundwork, we can formulate and prove Lemma 4.3. In essence, the lemma states that it is equivalent to search for sigma clique covers of bounded "overlap", and splitting sequences of bounded length that end in cluster graphs. Note that some special care needs to be taken to deal with the possibility of isolated vertices.

To prove the correspondence, we proceed as follows: Suppose we are given a graph with a sigma clique cover of "overlap" at most $k$. Then, we can apply Lemma 4.2 at most $k$ times to obtain a graph admitting a sigma clique cover of zero "overlap", which is a cluster graph.

Conversely, consider a splitting sequence of length at most $k$ that ends in a cluster graph. The last graph trivially has a sigma clique cover of zero "overlap". Then, we can work through the sequence in reverse order, and by repeatedly applying Lemma 4.1, obtain a sigma clique cover of the first graph that has an "overlap" of at most $k$.

**Lemma 4.3.** *Let $G = (V, E)$ be a graph, and let $I := \{v \in V \mid d_G(v) = 0\}$. Then, $(G, k)$ is a positive instance of* CVS *if and only if $(G, |V| - |I| + k)$ is a positive instance of* SCC.

**Proof.** ($\Rightarrow$): Let $G_0, \dots, G_\ell$ be a sequence of graphs with $G_0 = G$ and $\ell \leq k$ such that each graph, except $G_0$, is obtained from its predecessor via a vertex split, and $G_\ell$ is a cluster graph. Observe that a vertex split never results in a graph with fewer isolated vertices than the original graph, hence at least $|I|$ vertices of $G_\ell$ are isolated. By identifying all connected components of $G_\ell$ with their vertex sets, but omitting some $|I|$ trivial components, we can construct a sigma clique cover $\mathcal{C}_\ell$ of $G_\ell$ with $\text{wgt}(\mathcal{C}_\ell) = |V(G_\ell)| - |I|$. Each split used in the construction of $G_0, \dots, G_\ell$ introduces exactly one new vertex, therefore $|V(G_\ell)| = |V| + \ell$. Combining this with the fact that $\ell \leq k$, we derive $\text{wgt}(\mathcal{C}_\ell) \leq |V| - |I| + k$. Using the sequence $G_0, \dots, G_\ell$ in reverse order, we iteratively apply Lemma 4.1 $\ell$ times using $\mathcal{C}_\ell$ and $G_\ell$ as base case and obtain $\mathcal{C}_0, \dots, \mathcal{C}_\ell$. In particular, it follows that $\mathcal{C}_0$ is a sigma clique cover of $G$ satisfying $\text{wgt}(\mathcal{C}_0) \leq |V| - |I| + k$. Thus, $(G, |V| - |I| + k)$ is a positive instance of SCC.

($\Leftarrow$): Let $\mathcal{C}$ be a sigma clique cover of $G$ with $\text{wgt}(\mathcal{C}) \leq |V| - |I| + k$. Without loss of generality, we can assume that $\mathcal{C}$ contains no $C \in \mathcal{C}$ with $|C| \leq 1$, for $\mathcal{C} \setminus \{C\}$ still is a sigma clique cover of $G$ of weight not exceeding that of $\mathcal{C}$ for any such $C \in \mathcal{C}$. Observe that $\mathcal{C}$ is a sigma clique cover of $H_0 := G[V \setminus I]$ too, since $E(H_0) = E(G)$. Furthermore, set $\mathcal{C}_0 := \mathcal{C}$. By iteratively applying Lemma 4.2 for a number of times, call it $\ell$, either until a cluster graph is obtained as a direct result of the lemma, or alternatively, stopping after $l = k$ iterations, we can obtain the sequences $H_0, \dots, H_\ell$ and $\mathcal{C}_0, \dots, \mathcal{C}_\ell$.

We shall now verify that also in the latter case where $\ell = k$, $H_\ell$ must be a cluster graph. As a consequence of the $k$ applications of Lemma 4.2, we get $\text{wgt}(\mathcal{C}_\ell) \leq |V(H_\ell)|$. By considering the fact that for each vertex $v \in V(H_\ell)$ there exists $C \in \mathcal{C}_\ell$ with $v \in C$ (since $\mathcal{C}_\ell$ is a sigma clique cover of $H_\ell$ and $H_\ell$ contains no isolated vertices), we derive $\text{wgt}(\mathcal{C}_\ell) \geq |V(H_\ell)|$. Thus, we have that $\text{wgt}(\mathcal{C}_\ell) = |V(H_\ell)|$ and it follows that $\mathcal{C}_\ell$ forms a partition of $V(H_\ell)$. Using this partition property and the fact that $\mathcal{C}_\ell$ is a sigma clique cover of $H_\ell$ allows us to directly conclude that $H_\ell$ is a cluster graph. Thus, $H_\ell$ is a cluster graph in both cases.

We reintroduce the isolated vertices $I$ by constructing

$$H'_0, \dots, H'_\ell := \Big( (V(H_i) \cupdot I, E(H_i)) \Big)_{i \in \{0, \dots, \ell\}}.$$

$H'_0, \dots, H'_\ell$ forms a sequence of graphs where each constituent except the first is generated by performing a split in its predecessor for a total of no more than $k$ splits; this property is inherited from $H_0, \dots, H_\ell$. Note that in particular $H'_0 = G$ by definition, and furthermore, $H'_\ell$ is a cluster graph, since adding isolated vertices to a cluster graph yields another cluster graph. In total, we thus have obtained a certificate $H'_0, \dots, H'_\ell$ proving that $(G, k)$ is a positive instance of CVS. □

With the correspondence just established, the NP-hardness proof of CVS becomes immediate.

**Theorem 4.4.** Cluster Vertex Splitting *is* NP*-complete.*

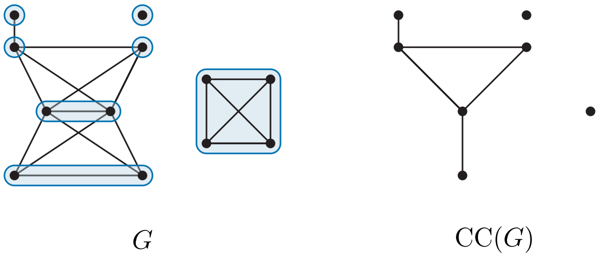

**Fig. 5.** A graph $G$ whose critical cliques are marked in blue (left) and CC($G$) (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Proof.** Let $(G, s)$ be an instance of SCC and $I := \{v \in V(G) \mid d_G(v) = 0\}$. We can leverage Lemma 4.3 to conclude that deciding this instance of SCC is equivalent to deciding the instance $(G, s - |V(G)| + |I|)$ of CVS. We have thus constructed a polynomial-time many-one reduction from SCC to CVS. Because SCC is NP-hard by Theorem 3.5, so is CVS. Observe that CVS $\in$ NP, since a certificate for CVS can clearly be guessed and checked in polynomial-time. Consequently, we conclude that CVS is NP-complete. □

## 5. A linear kernel for CLUSTER VERTEX SPLITTING

To start, we introduce the concept of *valency*, a straightforward tool that assists us in counting arguments. We also review the concept of *critical cliques* [39], where vertices that share identical closed neighborhoods are grouped together.

In Section 5.2, we establish the groundwork for the first reduction rule of the kernel, which allows us to reduce certain critical cliques in a SIGMA CLIQUE COVER instance. The second rule of the kernel is based on Section 5.3, where we determine that SIGMA CLIQUE COVER instances that have been exhaustively reduced using the previously explored mechanism and still contain more than $3k$ vertices are negative instances. We then give the kernel in Section 5.4.

### *5.1. The notions of valency and critical cliques*

We will frequently have to prove lower bounds for the weight that a sigma clique cover needs to have at minimum. This we will do by observing that particular vertices must be covered by at least a certain number of cliques each. To aid in such arguments, we introduce a new measure. The *valency* of a vertex $v$ with respect to a sigma clique cover $\mathcal{C}$ counts the number of cliques that contain $v$:

**Definition 5.1.** Let $\mathcal{C}$ be a sigma clique cover of a graph $G$. Then, for each vertex $v \in V(G)$, we define the *valency* of $v$ with respect to $\mathcal{C}$ as the number of cliques in $\mathcal{C}$ that cover $v$. Symbolically, we express this quantity as

$$\mathrm{val}_{\mathcal{C}}(v) := |\{C \in \mathcal{C} \mid v \in C\}|.$$

With this notation, we can express the weight of a sigma clique cover in an alternative manner: Via the definition of wgt(·) (Definition 3.1) and the principle of double counting, we obtain

$$\mathrm{wgt}(\mathcal{C}) = \sum_{C \in \mathcal{C}} |C| = \sum_{v \in V(G)} \mathrm{val}_{\mathcal{C}}(v).$$

Another key tool that we will use in this section is the concept of critical cliques, coined by Lin et al. [39]. This allows us to consider an equivalence relation, where vertices of a graph are in the same class if and only if their closed neighborhoods coincide. The equivalence classes under this relation are called the *critical cliques* of $G$. Consider a critical clique $C$ of $G$. Observe that it is fully connected "internally", that is, $G[C]$ is a clique, and that $N_G(v) \setminus C = N_G(w) \setminus C$ for any $v, w \in C$, which means that the vertices of $C$ share a common "external neighborhood".

If we delete all but one vertex from each critical clique, we obtain a graph isomorphic to what we will call the critical clique graph of $G$; we will use the shorthand CC($G$) to refer to it. See Fig. 5 for an example. Formally, we define this graph as follows:

**Definition 5.2.** Let $G$ be a graph. Consider the equivalence relation $R_G \subseteq V(G) \times V(G)$ where $(v, w) \in R_G$ if and only if $N(v) \cup \{v\} = N(w) \cup \{w\}$. We use $[v]_G$ to denote the equivalence class generated by $v \in V(G)$ and $R_G$. The *critical clique graph* of $G$, referred to using CC($G$), is given by

$$V(\mathrm{CC}(G)) := \{[v]_G \mid v \in V(G)\} \text{ and}$$
$$E(\mathrm{CC}(G)) := \{[v]_G[w]_G \mid vw \in E(G) \wedge [v]_G \neq [w]_G\}.$$

The main intuition we make use of here is that members of the same critical clique are essentially "clones" of one another. Thus, it seems reasonable that, provided certain conditions are met, we are allowed to "shrink" certain critical cliques without removing a significant amount of "computational complexity" when solving the combinatorial problems we are interested in.

### 5.2. Towards a rule to shrink critical cliques

Consider the critical clique graph $\mathrm{CC}(G)$ of a graph $G$. We distinguish between two kinds of critical cliques:

(1) Critical cliques $[v]_G$ such that their neighborhood, that is, $N_{\mathrm{CC}(G)}([v]_G)$, forms a clique in $\mathrm{CC}(G)$, and
(2) critical cliques $[v]_G$, where said neighborhood does not form a clique.

In this section, we show that, with respect to the sigma clique cover problem, critical cliques of the first kind consisting of at least two vertices, can either safely be reduced in size, or deleted altogether (Lemma 5.5). Correspondingly, we will refer to them as *reducible critical cliques*. The second kind of critical cliques we will call *irreducible critical cliques*.

To help prove Lemma 5.5, we first observe that in any minimum-weight sigma clique cover of a graph, a vertex member of a critical clique of the first kind is always covered by precisely one clique. Furthermore, this clique can be determined explicitly (Lemma 5.4). We start with a useful observation that we prove for completeness' sake:

**Lemma 5.3.** *Let $G$ be a graph, $\mathcal{C}$ a sigma clique cover of $G$, and $v \in C \in \mathcal{C}$. Then, $C \subseteq N_G(v) \cup \{v\}$.*

**Proof.** Suppose there is $w \in C \setminus \{N_G(v) \cup \{v\}\}$. Observe that $w$ differs from $v$. But then $w$ cannot be a neighbor of $v$ in $G$. Hence, $C$ cannot cover $v$ and $w$ simultaneously, contradicting our choice of $w$. □

Now, we are ready to prove our auxiliary lemma that offers insight into the structure of minimum-weight sigma clique covers:

**Lemma 5.4.** *Let $G$ be a graph without isolated vertices and let $[v]_G$ be a critical clique in $G$ such that $\mathrm{CC}(G)[N_{\mathrm{CC}(G)}([v]_G)]$ is a clique. Furthermore, let $\mathcal{C}$ be a minimum-weight sigma clique cover of $G$ and let $C^* := N_G(v) \cup \{v\}$. Then, $C^*$ is contained in $\mathcal{C}$. Moreover, $C^*$ is the only clique of $\mathcal{C}$ that covers $v$.*

**Proof.** We will first show that $G[C^*]$ is a clique; this will become useful later on. Since $v$ is not isolated, we can select two distinct vertices $a, b \in C^*$. We need to show that $ab \in E(G)$.

**Case** $[a]_G = [b]_G = [v]_G$: The vertices $a$ and $b$ are part of a shared critical clique. Hence, $N_G(a) \cup \{a\} = N_G(b) \cup \{b\}$, which implies $a \in N_G(b)$.

**Case** $[a]_G \neq [v]_G \wedge [b]_G \neq [v]_G$: Since $a \neq v$ and $b \neq v$, we have $\{a, b\} \subseteq N_G(v)$, implying $\{va, vb\} \subseteq E(G)$. Using Definition 5.2, we obtain that all of $\{[v]_G[a]_G, [v]_G[b]_G\}$ are edges of $\mathrm{CC}(G)$. If $[a]_G = [b]_G$, it is immediate that $ab \in E(G)$. Otherwise, we invoke the precondition that $\mathrm{CC}(G)[N_{\mathrm{CC}(G)}([v]_G)]$ is a clique, yielding $[a]_G[b]_G \in E(\mathrm{CC}(G))$, which implies $ab \in E(G)$.

**Case** $[a]_G = [v]_G \wedge [b]_G \neq [v]_G$: Similarly to the last case, $b \neq v$ gives $b \in N_G(v)$, implying $[v]_G[b]_G = [a]_G[b]_G \in E(\mathrm{CC}(G))$. Hence, $ab \in E(G)$.

**Case** $[a]_G \neq [v]_G \wedge [b]_G = [v]_G$: Symmetrical to the previous case.

Next, we show that $v$ is covered by at most one clique. Towards a contradiction, suppose that $\mathrm{val}_{\mathcal{C}}(v) \geq 2$. Let $C_1$ and $C_2$ be two distinct cliques of $\mathcal{C}$ such that $v \in C_1 \cap C_2$. By Lemma 5.3, we know that $C_1 \subseteq N_G(v) \cup \{v\}$ and $C_2 \subseteq N_G(v) \cup \{v\}$. Thus, $C_1 \cup C_2 \subseteq N_G(v) \cup \{v\} = C^*$. We have already shown that $G[C^*]$ is a clique. Since the family of clique graphs is closed under vertex deletion, we thus find that $G[C_1 \cup C_2]$ is a clique too. Now, let

$$\mathcal{C}' := (\mathcal{C} \setminus \{C_1, C_2\}) \cup (C_1 \cup C_2)\,.$$

Clearly, $\mathcal{C}'$ covers $G$ as $\mathcal{C}$ does. Also, we have just observed that $G[C_1 \cup C_2]$ is a clique, while all other $C \in \mathcal{C}'$ induce cliques in $G$ because $\mathcal{C}$ is a sigma clique cover of $G$. Therefore, $\mathcal{C}'$ is a sigma clique cover of $G$. But notice

$$\begin{aligned} \mathrm{wgt}(\mathcal{C}') &= \mathrm{wgt}(\mathcal{C}) - |C_1| - |C_2| + |C_1 \cup C_2| \\ &< \mathrm{wgt}(\mathcal{C}). \end{aligned} \qquad |C_1 \cup C_2| < |C_1| + |C_2|$$

This contradicts that $\mathcal{C}$ has minimum weight for $G$. Therefore, $\mathrm{val}_{\mathcal{C}}(v) < 2$. Since $v$ is not isolated, we additionally have that $\mathrm{val}_{\mathcal{C}}(v) \geq 1$. Thus, $\mathrm{val}_{\mathcal{C}}(v) = 1$.

We have shown that $v$ is covered by precisely one clique of $\mathcal{C}$; call it $C$. The last remaining step is to prove that $C = C^*$. Consider any edge $e \in E(G)$ incident with $v$. We observe that $e$ is covered by $C$, for were $e$ covered by any different $C' \in \mathcal{C}$, we would obtain $\mathrm{val}_{\mathcal{C}}(v) \geq 2$. Thus, considering all such edges lets us conclude that $N_G(v) \cup \{v\} = C^* \subseteq C$. At the same time, by Lemma 5.3, we get $C \subseteq N_G(v) \cup \{v\} = C^*$. Therefore, $C$ equals $C^*$ and the proof is complete. □

It remains to turn our previous observation into a lemma suitable to show the correctness of a reduction rule used in the kernel. More specifically, when we prove the correctness of Rule I formulated in Theorem 5.7, we will make direct use of the following lemma:

**Lemma 5.5.** *Let $G$ be a graph without isolated vertices and let $[v]_G$ be one of its critical cliques such that $|[v]_G| \geq 2$ and $\mathrm{CC}(G)[N_{\mathrm{CC}(G)}([v]_G)]$ is a clique. Then, $(G, |V(G)| + k)$ is a positive instance of* SCC *iff $(G - v, |V(G - v)| + k)$ is.*

**Proof.** $(\Rightarrow)$: Let $\mathcal{C}$ be a sigma clique cover of $G$ with $\mathrm{wgt}(\mathcal{C}) \leq |V(G)|+k$. We set $\mathcal{C}' := \{C \setminus \{v\} \mid C \in \mathcal{C}\}$. Clearly, $\mathcal{C}'$ is a sigma clique cover of $G-v$. Furthermore, since $v$ is not isolated, $\mathcal{C}$ covers $v$ and $\mathrm{wgt}(\mathcal{C}') < \mathrm{wgt}(\mathcal{C})$. Hence, $\mathrm{wgt}(\mathcal{C}') \leq |V(G - v)|+k$.

$(\Leftarrow)$: Let $\mathcal{C}'$ be a minimum-weight sigma clique cover of $G - v$ such that $\mathrm{wgt}(\mathcal{C}') \leq |V(G - v)| + k$. Furthermore, let $w \in [v]_G \setminus \{v\}$. We apply Lemma 5.4 to $G-v$, $[w]_{G-v}$, and $\mathcal{C}'$ to deduce that there is a single clique $C^* = N_{G-v}(w) \cup \{w\} \in \mathcal{C}'$ where $w \in C^*$. Next, let

$$\mathcal{C} := \left(\mathcal{C}' \setminus C\right)^* \cup \left(C^* \cup \{v\}\right).$$

We know that $v$ and $w$ are part of the same critical clique in $G$. Thus, $N_G(v) \cup \{v\} = N_G(w) \cup \{w\}$. Subtracting $v$ on both sides, we obtain

$$N_G(v) = N_{G-v}(w) \cup \{w\} = C^* \subseteq C^* \cup \{v\}.$$

Thus, all $e \in E(G) \setminus E(G - v)$ are covered by $C^* \cup \{v\}$. All remaining edges of $G$ are not incident with $v$; let $e$ be such an edge. Since there is $C' \in \mathcal{C}'$ that covers $e$ and $C' \subseteq C$ for some $C \in \mathcal{C}$, we have that $\mathcal{C}$ covers $e$.

It remains to show that all $C \in \mathcal{C}$ induce cliques in $G$. Let $C \in \mathcal{C}$. If $v \notin C$, then $G[C] = (G - v)[C]$. Otherwise, $C$ is equal to $C^* \cup \{v\}$. We know that $(G - v)[C^*]$ is a clique and that $G - v \prec G$. Thus, we only need to show that all edges between $C^*$ and $\{v\}$ exist in $G$. Let

$$\begin{aligned} a \in C^* &= N_{G-v}(w) \cup \{w\} \\ &\subseteq N_G(w) \cup \{w\} \\ &= N_G(v) \cup \{v\}. \end{aligned}$$

Since $a \neq v$, we have $a \in N_G(v)$, or phrased differently: $av \in E(G)$.

Observe that $\mathcal{C} \subseteq \mathcal{P}(V(G))$ and $\mathrm{wgt}(\mathcal{C}) = \mathrm{wgt}(\mathcal{C}') + 1 \leq |V(G - v)| + 1 + k = |V(G)| + k$. Therefore, we can finish our proof and conclude that $\mathcal{C}$ is a sigma clique cover of $G$ of the required weight. □

### 5.3. Towards a rule to recognize negative instances

In the previous section, we laid the foundation for a rule that minimizes the sizes of reducible critical cliques. Consider an instance $(G, |V(G)| + k)$ of Sigma Clique Cover that has been exhaustively reduced using the aforementioned rule. We now observe that, if this instance has more than $3k$ vertices, then it is a negative instance. This will serve as the basis for Rule II defined in Theorem 5.7.

We proceed as follows: We assume that $G$ has more than $3k$ vertices and consider an arbitrary sigma clique cover $\mathcal{C}$ of $G$. Then, we provide two separate lower bounds on $\mathrm{wgt}(\mathcal{C})$. One bound is based on reducible critical cliques, while the other bound is based on irreducible critical cliques. Each lower bound individually is too weak, but the maximum of both will be greater than $|V(G)| + k$ in all cases, yielding that $(G, |V(G)| + k)$ is a negative instance.

**Lemma 5.6.** *Let $G$ be a graph such that none of its connected components are cliques and $k \in \mathbb{N}$. We divide $V(\mathrm{CC}(G))$ into the partition $A \uplus B$ where $v \in A$ if and only if $\mathrm{CC}(G)[N_{\mathrm{CC}(G)}(v)]$ is a clique. Furthermore, we set*

$$\overline{A} := \bigcup A \text{ and } \overline{B} := \bigcup B,$$

*that is, the partition of $V(G)$ induced by $A \uplus B$. If $|A| = |\overline{A}|$ and $|V(G)| > 3k$, then $(G, |V(G)| + k)$ is a negative instance of* SCC.

**Proof.** We assume that $|A| = |\overline{A}|$ and $|V(G)| > 3k$. Let $\mathcal{C}$ be a sigma clique cover of $G$. We claim that

$$\mathrm{wgt}(\mathcal{C}) \geq \max\left\{2|\overline{A}|, |V(G)| + |\overline{B}|\right\} > |V(G)| + k.$$

First, we will derive $\mathrm{wgt}(\mathcal{C}) \geq 2|\overline{A}|$: Consider the set $B' \subseteq B$ with

$$B' := \left\{b \in B \mid N_{\mathrm{CC}(G)}(b) \cap A \neq \emptyset\right\}.$$

Phrased differently, $B'$ is the subset of $B$ where each element has at least one neighbor in $A$ in $\mathrm{CC}(G)$. Furthermore, let $f_B : B \to \overline{B}$ such that $f_B(b) \in b$ for all $b \in B$, that is, a function selecting an arbitrary vertex out of each critical clique contained in $B$. Additionally, we define a second function $f_A : A \to \overline{A}$ in a completely symmetric manner.

Now, consider some $b \in B'$ and $a \in A$ such that $ab \in E(\mathrm{CC}(G))$. By Lemma 5.4, there is precisely one $C \in \mathcal{C}$ such that $\{f_A(a), f_B(b)\} \subseteq C$. Thus, accounting for all such $a$, we obtain

$$\operatorname{val}_{\mathcal{C}}(f_B(b)) \geq |N_{\mathrm{CC}(G)}(b) \cap A|.$$

On the other hand, let $a \in A$. Suppose $N_{\mathrm{CC}(G)}(a) \subseteq A$. Then,

$$G\left[\bigcup (N_{\mathrm{CC}(G)}(a) \cup \{a\})\right]$$

is a connected component of $G$ that is a clique, which we required to never be the case. Thus $|N_{\mathrm{CC}(G)}(a) \cap B'| \geq 1$. Using these two facts, we obtain

$$\begin{aligned}
\sum_{b\in B'} \operatorname{val}_{\mathcal{C}}(f_B(b)) &\geq \sum_{b \in B'} |N_{\mathrm{CC}(G)}(b) \cap A| \\
&= \sum_{a \in A} |N_{\mathrm{CC}(G)}(a) \cap B'| && \textit{double counting principle} \\
&\geq |A| \\
&= |\overline{A}|.
\end{aligned}$$

In total, we calculate

$$\begin{aligned}
\operatorname{wgt}(\mathcal{C}) &= \sum_{\overline{v} \in V(G)} \operatorname{val}_{\mathcal{C}}(\overline{v}) \\
&= \sum_{c \in A \cup (B \setminus B')} \sum_{\overline{c} \in c} \operatorname{val}_{\mathcal{C}}(\overline{c}) + \sum_{b \in B'} \sum_{\overline{b} \in b} \operatorname{val}_{\mathcal{C}}(\overline{b}) \\
&\geq |\overline{A}| + \sum_{b \in B'} \sum_{\overline{b} \in b} \operatorname{val}_{\mathcal{C}}(\overline{b}) && A \subseteq A \cup (B \setminus B'),\ \forall \overline{v} \in V(G)\colon \operatorname{val}_{\mathcal{C}}(\overline{v}) \geq 1 \\
&\geq |\overline{A}| + \sum_{b \in B'} \operatorname{val}_{\mathcal{C}}(f_B(b)) && f_B(b) \in b \\
&\geq 2|\overline{A}|.
\end{aligned}$$

Next, we will derive $\operatorname{wgt}(\mathcal{C}) \geq |V(G)| + |\overline{B}|$: Let $[v]_G \in B$. By definition of $B$, there are distinct $[u]_G, [w]_G \in V(\mathrm{CC}(G))$ such that $\{vu, vw\} \subseteq E(G)$, but $uw \notin E(G)$. Let $C_1 \in \mathcal{C}$ such that $\{v, u\} \subseteq C_1$ and $C_2 \in \mathcal{C}$ such that $\{v, w\} \subseteq C_2$. Since $uw \notin E(G)$, we know that $C_1$ differs from $C_2$. Thus, $\operatorname{val}_{\mathcal{C}}(v) \geq 2$. In total, we obtain

$$\begin{aligned}
\operatorname{wgt}(\mathcal{C}) &= \sum_{\overline{v} \in V(G)} \operatorname{val}_{\mathcal{C}}(\overline{v}) \\
&= \sum_{b \in B} \sum_{\overline{b} \in b} \operatorname{val}_{\mathcal{C}}(\overline{b}) + \sum_{a \in A} \sum_{\overline{a} \in a} \operatorname{val}_{\mathcal{C}}(\overline{a}) \\
&\geq 2|\overline{B}| + |\overline{A}| \\
&= |V(G)| + |\overline{B}|. && |V(G)| = |\overline{A}| + |\overline{B}|
\end{aligned}$$

To finish our proof, we will combine these two bounds to obtain that $\operatorname{wgt}(\mathcal{C}) > |V(G)| + k$. First, suppose that $|\overline{A}| \geq \frac{2}{3}|V(G)|$. Then,

$$\begin{aligned}
\operatorname{wgt}(\mathcal{C}) &\geq 2|\overline{A}| \\
&\geq \frac{4}{3}|V(G)| \\
&> |V(G)| + k. && |V(G)| > 3k
\end{aligned}$$

If otherwise $|\overline{A}| < \frac{2}{3}|V(G)|$, then

$$\begin{aligned}
\operatorname{wgt}(\mathcal{C}) &\geq |V(G)| + |\overline{B}| \\
&\geq |V(G)| + \frac{1}{3}|V(G)| && |\overline{B}| \geq \tfrac{1}{3}|V(G)| \\
&> |V(G)| + k. && |V(G)| > 3k
\end{aligned}$$

Therefore, we conclude that $\mathrm{wgt}(\mathcal{C}) > |V(G)| + k$ in all cases. Since $\mathcal{C}$ was chosen generically, this implies $(G, |V(G)| + k)$ is a negative instance of SCC. □

### 5.4. Deriving the kernel

In the two preceding sections, we have essentially derived two reduction rules for the sigma clique cover problem. It remains to compile our results into a polynomial kernelization procedure for Cluster Vertex Splitting. Essentially, we convert a given instance $(G, k)$ of Cluster Vertex Splitting into an equivalent instance of Sigma Clique Cover, apply the two reduction rules exhaustively, until finally converting the reduced instance back to an instance of Cluster Vertex Splitting. Refer to Fig. 6 for an example.

**Theorem 5.7.** Cluster Vertex Splitting *admits a problem kernelization with running time $\mathcal{O}(|V(G)| + |E(G)|)$ mapping an instance $(G, k)$ to an equivalent instance $(G', k')$ satisfying $|V(G')| \leq 3k + 3$ and $k' \leq k$.*

**Proof.** Let an instance of Cluster Vertex Splitting be given through $(G, k)$ and let $G_0$ be obtained from $G$ by removing all isolated vertices. Observe that $(G, k)$ is equivalent to $(G_0, k =: k_0)$ with respect to CVS. We apply Lemma 4.3 and derive that $(G_0, k_0)$ is a positive instance of CVS if and only if $(G_0, |V(G_0)| + k_0)$ is a positive instance of Sigma Clique Cover. Next, we construct the sequences $G_0, \ldots$ and $k_0, \ldots$ by exhaustively applying the following set of rules:

**Rule I:** If there is a critical clique $[v]_{G_i} \in V(\mathrm{CC}(G_i))$ such that $[v]_{G_i}$ contains at least two vertices and $\mathrm{CC}(G_i)[N_{\mathrm{CC}(G_i)}([v]_{G_i})]$ is a clique, then $G_{i+1} := (G_i - v) - I$ and $k_{i+1} := k_i$, where $I$ is the set of isolated vertices in $G_i - v$.

**Rule II:** If Rule I is not applicable to $G_i$, Rule II has not been used so far, and $|V(G_i)| > 3k_i$, then $G_{i+1} := P_3$ and $k_{i+1} := 0$.

*Termination in linear time.* Observe that Rule I reduces the number of vertices of the current graph, and that Rule II is applicable at most once. Thus, both sequences are finite and of length $\ell = \mathcal{O}(|V(G)|)$. The time complexity of constructing the critical clique graph of a graph $H$ is in $\mathcal{O}(|V(H)| + |E(H)|)$ [33]. Note that it suffices to calculate the critical clique graph once and then update it in constant time per step as (up to graph isomorphism) applying Rule I may only do nothing or delete isolated vertices. Hence, we observe that our sequences can be constructed using a budget of $\mathcal{O}(|V(G)| + |E(G)|)$ steps.

*Correctness.* We claim that Rule I and Rule II are *correct*, that is, the instances $(G_i, |V(G_i)| + k_i)$ and $(G_{i+1}, |V(G_{i+1})| + k_{i+1})$ are equivalent with respect to the SCC problem for all $i \in \{0, \ldots, \ell - 1\}$. Let $G_i$ such that $G_{i+1}$ was obtained by applying Rule I, and let $v$ as well as $I$ as used in the definition of Rule I. First, consider the case when $I \neq \emptyset$. Let $w \in I$. We have that $d_{G_i}(w) \geq 1$, because $w$ is not isolated in $G_i$. At the same time, we know that $d_{G_i}(w) < 2$, for otherwise $w$ would not be isolated in $G_i - v$. Thus, $d_{G_i}(w) = 1$, which forces $|[v]_{G_i}| = 2$. Since $w \notin [v]_{G_i}$ would imply $d_{G_i}(w) \geq 2$, we conclude that $[v]_{G_i} = \{v, w\}$, that is, $G_i[\{v, w\}] \simeq K_2$ is a connected component of $G_i$. Now, it is easy to see that $(G_i, |V(G_i)| + k_i)$ is equivalent to $(G_{i+1}, |V(G_{i+1})| + k_{i+1})$ with respect to the SCC problem. Otherwise, $I = \emptyset$. By construction, $G_i$ is free of isolated vertices. Thus, applying Lemma 5.5 yields that $(G_i, |V(G_i)| + k_i)$ is equivalent to $(G_i - v, |V(G_i - v)| + k_i) = (G_{i+1}, |V(G_{i+1})| + k_{i+1})$ with respect to the SCC problem. Hence, Rule I is correct.

Next, let $G_i$ such that $G_{i+1}$ was obtained by applying Rule II, and let $A, \overline{A}, B, \overline{B}$ as defined in the header of Lemma 5.6 when substituting $G$ for $G_i$. Then, $|V(G_i)| > 3k_i$ and Rule I is not applicable to $G_i$. Hence, for all $[v]_{G_i} \in V(\mathrm{CC}(G_i))$ such that $\mathrm{CC}(G_i)[N_{\mathrm{CC}(G_i)}([v]_{G_i})]$ is a clique, we have $|[v]_{G_i}| = 1$. Note that this implies $|A| = |\overline{A}|$. Again, notice also that $G_i$ cannot contain isolated vertices. Now, suppose that $C \subseteq V(G_i)$ induces a connected component of $G_i$ that is a clique with $|C| > 1$ and let $v \in C$. Then, $C$ "spans" the whole of $G_i[C]$, that is, $[v]_{G_i} = C$ and $\mathrm{CC}(G_i)[N_{\mathrm{CC}(G_i)}([v]_{G_i})] = \emptyset$. Thus, applying the above, we have $|C| = 1$, which cannot be since $G_i$ is free of isolated vertices. Therefore, none of $G_i$'s connected components are cliques. Hence, all conditions are met to apply Lemma 5.6 to $G_i, k_i, A, \overline{A}, B$ and $\overline{B}$, showing that $(G_i, k_i)$ is a negative instance of SCC. Since $(P_3, |P_3| + 0)$ is a negative instance of SCC too, Rule II is correct.

In total, we have that $(G_\ell, |V(G_\ell)| + k_\ell)$ is a positive instance of SCC if and only if $(G_0, |V(G_0)| + k_0)$ is. Another application of Lemma 4.3 (using that $G_\ell$ is free of isolated vertices) yields that $(G_\ell, |V(G_\ell)| + k_\ell)$ is a positive instance of SCC if and only if $(G_\ell, k_\ell)$ is a positive instance of CVS. Finally, we conclude that $(G, k)$ is equivalent to $(G_\ell, k_\ell)$ with respect to the CVS problem.

*Problem kernel size.* First, we observe that $k_\ell \leq k$, as no rule may increase the current value for $k$. If Rule II was used in the construction of the sequence at any step, then $|V(G_\ell)| = |P_3| \leq 3 + k$. Otherwise, Rule II was not used. As $G_\ell$ is the last element of $G_0, \ldots, G_\ell$, no rule is applicable to it. Suppose $|V(G_\ell)| > 3k_\ell$. But then, Rule II is applicable, which is a contradiction. Hence, $|V(G_\ell)| \leq 3k_\ell \leq 3k + 3$. □

## 6. Characterization and hardness of cluster editing with vertex splitting

We now give a characterization of solutions to Cluster Editing With Vertex Splitting (CEVS) and give an alternative hardness reduction. CEVS is defined as follows, where by a *graph modification* we mean a vertex split, an edge addition, or an edge deletion.

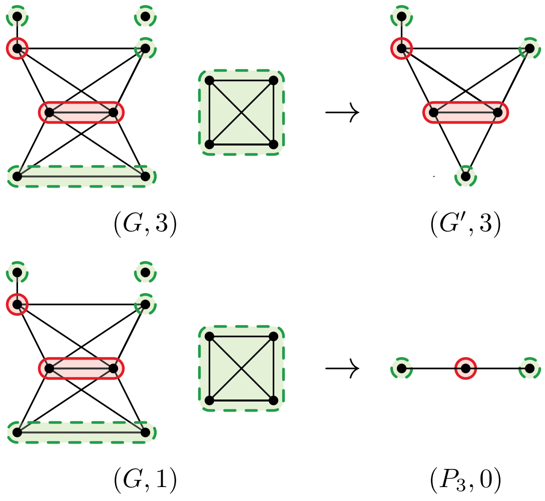

**Fig. 6.** Two instances of CVS and their corresponding kernel as given by Theorem 5.7. Reducible critical cliques are marked in green with dashed outlines, while irreducible critical cliques are marked in red with solid outlines. Note that the graph $G$ is taken from Fig. 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Cluster Editing With Vertex Splitting (CEVS)

**Input:** A tuple $(G, k)$, where $G$ is a graph and $k \in \mathbb{N}$.
**Question:** Is there a sequence of at most $k$ graph modifications that transforms $G$ into a cluster graph?

The alternative hardness result uses the critical-clique lemma for CEVS. To state it conveniently, we first need an equivalence between the sequence of modifications in CEVS and a cover of the input graph by clusters, similar to the correspondence between sigma clique covers and cluster vertex splittings in Lemma 4.3.

A *cover* of a graph $G$ is a collection $\mathcal{C}$ of subsets of $V(G)$ such that $\bigcup_{C\in\mathcal{C}} C = V(G)$. The *cost* $\mathrm{cst}_G(\mathcal{C})$ of a cover $\mathcal{C}$ is the number of non-edges contained in a set of $\mathcal{C}$ plus the number of edges not contained in any set of $\mathcal{C}$ plus the number of times each vertex is covered by a set beyond the first time. In formulas,

$$\mathrm{cst}_G(\mathcal{C}) = \left|\left\{uv \in \binom{V}{2} \setminus E(G) \mid \exists C \in \mathcal{C} \colon \{u,v\} \subseteq C\right\}\right| +$$
$$\left|\{uv \in E(G) \mid \forall C \in \mathcal{C} \colon \{u,v\} \nsubseteq C\}\right| + \left(\sum_{C\in\mathcal{C}} |C|\right) - |V(G)|.$$

Herein, $\binom{V}{2}$ denotes the set of all two-element subsets of $V$. If $G$ is clear from the context, we omit the subscript $G$ in $\mathrm{cst}_G$.

The following lemma has been used implicitly by Abu-Khzam et al. [3] but we are not aware of a formal proof.

**Lemma 6.1.** *Let $G$ be a graph and $k$ a positive integer. There is a sequence of at most $k$ graph modifications to obtain from $G$ a cluster graph if and only if $G$ admits a cover of cost at most $k$.*

**Proof.** Let $S$ be a sequence of at most $k$ graph modifications such that applying them to $G$ results in a cluster graph. By a reordering argument of Abu-Khzam et al. [3] (see [4, Theorem 1]) we may assume that $S$ consists of a possibly empty sequence of edge additions, then a possibly empty sequence of edge deletions, and then a possibly empty sequence of vertex splits. Consider the graph $\tilde{G}$ obtained after performing all edge additions and edge deletions but none of the vertex splits. Let $\ell$ be the number of vertex splits in $S$ and $n_0$ the number of degree-0 vertices in $\tilde{G}$. By Lemma 4.3 there is a sigma clique cover of $\tilde{G}$ of weight at most $n - n_0 + \ell$ were $n$ is the number of vertices of $\tilde{G}$. By adding to this sigma clique cover the degree-0 vertices of $\tilde{G}$ as singleton sets, we obtain a cover $\mathcal{C}$ of $\tilde{G}$. Observe that the cost of $\mathcal{C}$ (with respect to $\tilde{G}$) is at most $\ell$. Notice that the number of edges of $G$ that are not contained in any set in $\mathcal{C}$ is at most the number of edge deletions in $S$ and that the number of non-edges of $G$ that are contained in at least one set in $\mathcal{C}$ is at most the number of edge additions in $S$. Hence, $\mathcal{C}$ is a cover of $G$ of cost at most $k$.

Now let $\mathcal{C}$ be a cover of $G$ of cost at most $k$. Delete each edge from $G$ that is not in any set in $\mathcal{C}$ and for each non-edge of $G$ that is contained in some set of $\mathcal{C}$, add the corresponding edge to $G$. Denote by $\tilde{G}$ the so-obtained graph. Let $k'$ be obtained from $k$ by subtracting the number of performed graph modifications so far. Note that $\mathcal{C}$ is a sigma clique cover of $\tilde{G}$. Remove the isolated vertices from $\mathcal{C}$, obtaining $\mathcal{C}'$, which is still a sigma clique cover of $\tilde{G}$. Moreover, the weight of $\mathcal{C}'$

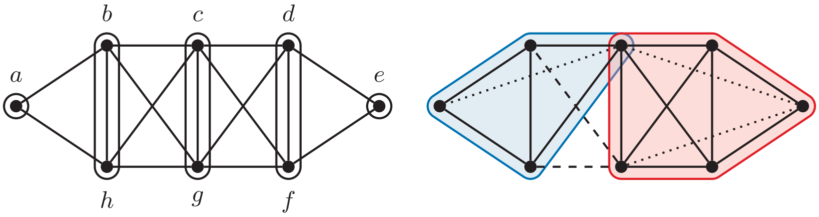

**Fig. 7.** A graph with an optimal cover that cuts a critical clique.

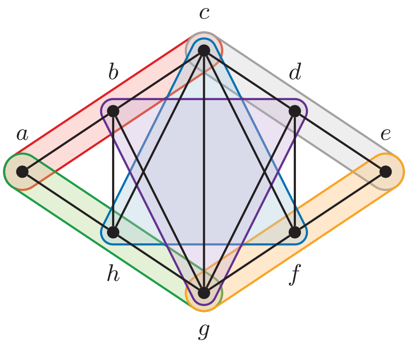

**Fig. 8.** The $P_3$ packing in Proposition 6.3.

with respect to $\tilde{G}$ is at most $n - n_0 + k'$, where $n_0$ is the number of isolated vertices in $\tilde{G}$, by the definition of the cost of $\mathcal{C}$. Thus, by Lemma 4.3 we may split at most $k'$ vertices in $\tilde{G}$ to obtain a cluster graph. □

Recall the definition of critical cliques from Definition 5.2. The critical-clique lemma is stated as follows.

**Lemma 6.2** (*Abu-Khzam et al. [1]*)**.** *Let $G$ be a graph and $k$ a positive integer. If $(G, k)$ admits a solution for* CEVS*, then there is a cover $\mathcal{C}$ of cost at most $k$ such that for each critical clique $K$ of $G$ and each set $C \in \mathcal{C}$ we have either $K \subseteq C$ or $K \cap C = \emptyset$.*

We mention in passing that we cannot assume that *every* cover corresponding to an optimal solution respects critical cliques. (Indeed, a previous version of the critical clique lemma claimed this stronger property, which we observed to be incorrect [25,26]. The authors since provided a sound proof of the above, weaker version [1].) This is shown in Fig. 7: The left shows the input graph with marked critical cliques. The right shows a minimum-cost cover in which the left cover set contains the central critical clique only partially. The cover has cost 6. That this is optimal is shown in the following proposition:

**Proposition 6.3.** *The graph shown on the left in Fig. 7 needs at least 6 modifications to turn it into a cluster graph.*

**Proof.** We show that there is a modification-disjoint packing of six induced $P_3$s. In the following, we denote a $P_3$ by $xyz$, where $x$, $y$, and $z$ are its three vertices and $y$ is the center vertex. Two $P_3$s $xyz$ and $abc$ are *modification disjoint* if they do not contain the same vertex pair (that is, the same edge or non-edge) and they do not contain the same center vertex. In formulas, $|\{a, b, c\} \cap \{x, y, z\}| \leq 1$ and $y \neq b$.

A modification-disjoint packing of $P_3$s is a collection of induced $P_3$s that are pairwise modification disjoint. Note that, if a graph admits a modification-disjoint packing of $\ell$ $P_3$s then we need at least $\ell$ modifications to turn the graph into a cluster graph.

Consider the following $P_3$s in the graph in Fig. 7: $abc$, $cde$, $ahg$, $gfe$, $hcf$, $bgd$. See also Fig. 8. Note that they form a modification-disjoint packing. Thus we need at least 6 modifications to turn the graph into a cluster graph. □

Based mainly on our NP-hardness proof of Cluster Vertex Splitting in conjunction with the critical-clique lemma we obtain NP-hardness of Cluster Editing With Vertex Splitting:

**Theorem 6.4.** *There is a polynomial-time many-one reduction from* CVS *to* CEVS*, showing that* CEVS *is* NP*-hard.*

**Proof.** We give a reduction from Cluster Vertex Splitting (CVS) to Cluster Editing With Vertex Splitting (CEVS). Let $(G, k)$ be an instance of CVS. Without loss of generality, we assume that $G$ does not contain isolated vertices. We construct an instance $(H, s)$ of CEVS. To obtain $H$ from $G$, replace each vertex in $G$ by a clique with $k + 1$ vertices. That is, $V(H) = \{v_i \mid v \in V(G), i \in [k + 1]\}$ and $E(H) = \{u_i v_j \mid uv \in E(G), i, j \in [k + 1]\}$. We say that $v_i \in V(H)$ is a *copy* of $v \in V(G)$

and for each $v \in V(G)$ we let $K_v := \{v_i \in V(H) \mid i \in [k+1]\}$ denote the *clique of* $v$. Put $s = k(k+1)$. Clearly, the reduction can be carried out in polynomial time. It remains to prove that $(G, k)$ has a solution (for CVS) if and only if $(H, s)$ has a solution (for CEVS).

Let $S$ be a solution to $(G, k)$. By Lemma 4.3 there is a sigma clique cover $\mathcal{C}$ for $G$ of weight at most $n+k$. From $\mathcal{C}$, construct a cover $\mathcal{C}'$ for $H$ by replacing in each set of $\mathcal{C}$ each vertex by all of its copies. That is $\mathcal{C}' = \{\{v_i \mid v \in C, i \in [k+1]\} \mid C \in \mathcal{C}\}$. Observe that, since each set in $\mathcal{C}$ is a clique with $k+1$ vertices, we have $\mathrm{cst}(\mathcal{C}') \leq k(k+1)$. Thus, $(H, s)$ has a solution by Lemma 6.1.

Let $S$ be a solution to $(H, s)$. By Lemma 6.1 there is a cover $\mathcal{C}'$ of cost at most $k(k+1)$. By Lemma 6.2 we may assume that $\mathcal{C}'$ is such that for each critical clique in $H$ with vertex set $K$ and each set $C' \in \mathcal{C}'$ we have either $K \subseteq C'$ or $K \cap C' = \emptyset$. We claim that $\mathcal{C}'$ is a sigma clique cover for $H$. Observe that for each $v \in V(G)$ we have that $K_v$ is contained in some critical clique of $H$. Hence, for all $C' \in \mathcal{C}'$ we have either $K_v \subseteq C'$ or $K_v \cap C' = \emptyset$. We claim that each edge of $H$ is contained in a set of $\mathcal{C}'$. For a contradiction, assume the contrary, that is, there are $i, j \in [k+1]$ and $uv \in E(G)$ such that $u_i v_j \in E(H)$ is not contained in any set of $\mathcal{C}'$. It follows that indeed for all $i, j \in [k+1]$ we have that $u_i v_j \in E(H)$ is not contained in any set of $\mathcal{C}'$. That is, the cost of $\mathcal{C}'$ is at least $(k+1)^2$, a contradiction to the fact that $\mathcal{C}'$ has cost at most $k(k+1)$. Analogously we can show that no non-edge of $H$ is contained in a set of $\mathcal{C}'$. Hence, indeed $\mathcal{C}'$ is a sigma clique cover of $H$. Construct a sigma clique cover $\mathcal{C}$ for $G$ by replacing each clique $K_v$ by $v$, that is, put $\mathcal{C} = \{\{v \in V(G) \mid K_v \subseteq C'\} \mid C' \in \mathcal{C}'\}$. Observe that $\mathcal{C}$ has weight at most $n+k$. Thus, by Lemma 4.3, $(G, k)$ has a solution, as required. □

## 7. Conclusion

We conclude with directions for future research. The constants in our kernelization for CVS (at most $3k+3$ vertices, see Theorem 5.7) are already quite small, but it would be interesting to see whether they can be further improved. A problem kernel with a linear number of edges would also be interesting. Our technique for obtaining the kernel for CVS follows the same basic idea as the independently developed technique by Abu-Khzam et al. [1] for CEVS but differs in crucial aspects in the reduction rules and analysis. Since we obtain a better bound on the number of remaining vertices, it would be interesting to see whether our technique can be used to improve Abu-Khzam et al.'s bound as well.

In terms of solution algorithms for CVS, a straightforward brute-force search on the kernel yields an algorithm solving CVS in $2^{O(k^2)} \cdot n^{O(1)}$ time, which can be improved to $2^{O(k \log k)} \cdot n^{O(1)}$ with further straightforward observations. Is it possible to obtain $2^{O(k)} \cdot n^{O(1)}$ time as well?

Finally, we focused here on the case where the overlap between clusters is small. There are applications where the overlap is relatively large [44]. Thus, to get efficient algorithms in this case, it would be interesting to study parameterizations dual to $k$ that measure the non-overlapping parts of the clustering.

## Data availability

No data was used for the research described in the article.

## References

[1] F.N. Abu-Khzam, E. Arrighi, M. Bentert, P.G. Drange, J. Egan, S. Gaspers, A. Shaw, P. Shaw, B.D. Sullivan, P. Wolf, Cluster editing with vertex splitting, 2023, CoRR, abs/1901.00156v2, arXiv:1901.00156v2.

[2] F.N. Abu-Khzam, J.R. Barr, A. Fakhereldine, P. Shaw, A greedy heuristic for cluster editing with vertex splitting, in: Proceedings of the 4th International Conference on Artificial Intelligence for Industries, AI4I 2021, IEEE, 2021, pp. 38–41, http://dx.doi.org/10.1109/AI4I51902.2021.00017.

[3] F.N. Abu-Khzam, J. Egan, S. Gaspers, A. Shaw, P. Shaw, Cluster editing with vertex splitting, in: J. Lee, G. Rinaldi, A.R. Mahjoub (Eds.), Proceedings of the 5th International Symposium of Combinatorial Optimization, ISCO 2018, in: Lecture Notes in Computer Science, vol. 10856, Springer, 2018, pp. 1–13, http://dx.doi.org/10.1007/978-3-319-96151-4_1.

[4] F.N. Abu-Khzam, J. Egan, S. Gaspers, A. Shaw, P. Shaw, On the parameterized cluster editing with vertex splitting problem, 2019, CoRR, abs/1901.00156v1, arXiv:1901.00156v1.

[5] E. Arrighi, M. Bentert, P.G. Drange, B.D. Sullivan, P. Wolf, Cluster editing with overlapping communities, in: N. Misra, M. Wahlström (Eds.), 18th International Symposium on Parameterized and Exact Computation, IPEC 2023, September 6–8, 2023, Amsterdam, The Netherlands, in: LIPIcs, vol. 285, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023, pp. 2:1–2:12, http://dx.doi.org/10.4230/LIPICS.IPEC.2023.2.

[6] G. Askeland, Overlapping Community Detection Using Cluster Editing with Vertex Splitting (MA thesis), University of Bergen, 2022, URL https://hdl.handle.net/11250/3045483.

[7] J. Baumann, M. Pfretzschner, I. Rutter, Parameterized complexity of vertex splitting to pathwidth at most 1, 2023, CoRR abs/2302.14725, arXiv:2302.14725.

[8] S. Böcker, A golden ratio parameterized algorithm for cluster editing, J. Discrete Algorithms 16 (2012) 79–89, http://dx.doi.org/10.1016/j.jda.2012.04.005.

[9] S. Böcker, S. Briesemeister, Q. Bui, A. Truss, Going weighted: Parameterized algorithms for cluster editing, Theoret. Comput. Sci. 410 (52) (2009) 5467–5480, http://dx.doi.org/10.1016/j.tcs.2009.05.006.

[10] S. Böcker, S. Briesemeister, G.W. Klau, Exact algorithms for Cluster Editing: Evaluation and experiments, Algorithmica 60 (2) (2011) 316–334, http://dx.doi.org/10.1007/s00453-009-9339-7.

[11] S. Böcker, P. Damaschke, Even faster parameterized cluster deletion and cluster editing, Inform. Process. Lett. 111 (14) (2011) 717–721, http://dx.doi.org/10.1016/j.ipl.2011.05.003.

[12] H.L. Bodlaender, M.R. Fellows, P. Heggernes, F. Mancini, C. Papadopoulos, F.A. Rosamond, Clustering with partial information, Theoret. Comput. Sci. 411 (7–9) (2010) 1202–1211, http://dx.doi.org/10.1016/j.tcs.2009.12.016.

[13] N. Bousquet, J. Daligault, S. Thomassé, Multicut is FPT, SIAM J. Comput. 47 (1) (2018) 166–207, http://dx.doi.org/10.1137/140961808.

[14] C. Crespelle, P.G. nås Drange, F.V. Fomin, P.A. Golovach, A survey of parameterized algorithms and the complexity of edge modification, Comput. Sci. Rev. 48 (2023) 100556, http://dx.doi.org/10.1016/j.cosrev.2023.100556.

[15] M. Cygan, F.V. Fomin, Ł. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, S. Saurabh, Parameterized Algorithms, vol. 5, Springer, 2015, http://dx.doi.org/10.1007/978-3-319-21275-3.

[16] M. Cygan, M. Pilipczuk, M. Pilipczuk, Known algorithms for edge clique cover are probably optimal, SIAM J. Comput. 45 (1) (2016) 67–83.

[17] P. Damaschke, Fixed-parameter enumerability of cluster editing and related problems, Theory Comput. Syst. 46 (2) (2010) 261–283, http://dx.doi.org/10.1007/s00224-008-9130-1.

[18] A. Davoodi, R. Javadi, B. Omoomi, Edge clique covering sum of graphs, Acta Math. Hungar. 149 (1) (2016) 82–91, http://dx.doi.org/10.1007/s10474-016-0586-1.

[19] R.G. Downey, M.R. Fellows, Parameterized Complexity, Springer Science & Business Media, 1999, http://dx.doi.org/10.1007/978-1-4471-5559-1.

[20] P. Eades, C.F.X. de Mendonça Neto, Vertex splitting and tension-free layout, in: Proceedings of the International Symposium on Graph Drawing, GD 1995, in: Lecture Notes in Computer Science, vol. 1027, Springer, 1995, pp. 202–211, http://dx.doi.org/10.1007/BFb0021804.

[21] D. Eppstein, P. Kindermann, S. Kobourov, G. Liotta, A. Lubiw, A. Maignan, D. Mondal, H. Vosoughpour, S. Whitesides, S. Wismath, On the planar split thickness of graphs, Algorithmica 80 (2018) 977–994, http://dx.doi.org/10.1007/s00453-017-0328-y.

[22] L. Faria, C.M.H. de Figueiredo, C.F.X. de Mendonça Neto, Splitting number is NP-complete, Discrete Appl. Math. 108 (1–2) (2001) 65–83, http://dx.doi.org/10.1016/S0166-218X(00)00220-1.

[23] M.R. Fellows, J. Guo, C. Komusiewicz, R. Niedermeier, J. Uhlmann, Graph-based data clustering with overlaps, Discrete Optim. 8 (1) (2011) 2–17, http://dx.doi.org/10.1016/j.disopt.2010.09.006.

[24] A. Firbas, Establishing Hereditary Graph Properties via Vertex Splitting (MA Thesis), TU Wien, 2023, http://dx.doi.org/10.34726/hss.2023.103864.

[25] A. Firbas, A. Dobler, F. Holzer, J. Schafellner, M. Sorge, A. Villedieu, M. Wißmann, The complexity of cluster vertex splitting and company, 2023, CoRR abs/2309.00504v3, arXiv:2309.00504v3.

[26] A. Firbas, A. Dobler, F. Holzer, J. Schafellner, M. Sorge, A. Villedieu, M. Wißmann, The complexity of cluster vertex splitting and company, in: H. Fernau, S. Gaspers, R. Klasing (Eds.), Proceedings of the 49th International Conference on Current Trends in Theory and Practice of Computer Science, SOFSEM 2024, in: Lecture Notes in Computer Science, vol. 14519, Springer, 2024, pp. 226–239, http://dx.doi.org/10.1007/978-3-031-52113-3_16.

[27] A. Firbas, M. Sorge, On the complexity of establishing hereditary graph properties via vertex splitting, in: J. Mestre, A. Wirth (Eds.), Proceedings of the 35th International Symposium on Algorithms and Computation, ISAAC 2024, in: Leibniz International Proceedings in Informatics (LIPIcs), vol. 322, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2024, pp. 30:1–30:15, http://dx.doi.org/10.4230/LIPIcs.ISAAC.2024.30.

[28] J. Flum, M. Grohe, Parameterized Complexity Theory, in: Texts in Theoretical Computer Science. An EATCS Series, Springer, 2006, http://dx.doi.org/10.1007/3-540-29953-X.

[29] F.V. Fomin, S. Kratsch, M. Pilipczuk, M. Pilipczuk, Y. Villanger, Tight bounds for parameterized complexity of cluster editing with a small number of clusters, J. Comput. System Sci. 80 (7) (2014) 1430–1447, http://dx.doi.org/10.1016/j.jcss.2014.04.015.

[30] J. Gramm, J. Guo, F. Hüffner, R. Niedermeier, Graph-modeled data clustering: Exact algorithms for clique generation, Theory Comput. Syst. 38 (4) (2005) 373–392, http://dx.doi.org/10.1007/s00224-004-1178-y.

[31] J. Gramm, J. Guo, F. Hüffner, R. Niedermeier, Data reduction and exact algorithms for clique cover, ACM J. Exp. Algorithm. 13 (2009) 2:2.2–2:2.15, http://dx.doi.org/10.1145/1412228.1412236.

[32] J. Gramm, J. Guo, F. Hüffner, R. Niedermeier, H.-P. Piepho, R. Schmid, Algorithms for compact letter displays: Comparison and evaluation, Comput. Statist. Data Anal. 52 (2) (2007) 725–736, http://dx.doi.org/10.1016/j.csda.2006.09.035.

[33] J. Guo, A more effective linear kernelization for cluster editing, Theoret. Comput. Sci. 410 (8) (2009) 718–726, http://dx.doi.org/10.1016/j.tcs.2008.10.021.

[34] J. Guo, I.A. Kanj, C. Komusiewicz, J. Uhlmann, Editing graphs into disjoint unions of dense clusters, Algorithmica 61 (4) (2011) 949–970, http://dx.doi.org/10.1007/s00453-011-9487-4.

[35] J. Guo, C. Komusiewicz, R. Niedermeier, J. Uhlmann, A more relaxed model for graph-based data clustering: s-Plex cluster editing, SIAM J. Discrete Math. 24 (4) (2010) 1662–1683, http://dx.doi.org/10.1137/090767285.

[36] R.M. Karp, Reducibility among combinatorial problems, in: R.E. Miller, J.W. Thatcher (Eds.), Proceedings of a Symposium on the Complexity of Computer Computations, in: The IBM Research Symposia Series, Plenum Press, New York, 1972, pp. 85–103, http://dx.doi.org/10.1007/978-1-4684-2001-2_9.

[37] C. Komusiewicz, J. Uhlmann, Cluster editing with locally bounded modifications, Discrete Appl. Math. 160 (15) (2012) 2259–2270, http://dx.doi.org/10.1016/j.dam.2012.05.019.

[38] S. Li, M. Pilipczuk, M. Sorge, Cluster editing parameterized above modification-disjoint $P_3$-packings, in: Proceedings of the 38th International Symposium on Theoretical Aspects of Computer Science, STACS 2021, in: LIPIcs, vol. 187, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, pp. 49:1–49:16, http://dx.doi.org/10.4230/LIPIcs.STACS.2021.49.

[39] G.-H. Lin, P.E. Kearney, T. Jiang, Phylogenetic $k$-root and steiner $k$-root, in: G. Goos, J. Hartmanis, J. van Leeuwen, D.T. Lee, S.-H. Teng (Eds.), Proceedings of the 11th International Symposium on Algorithms and Computation, ISAAC 2000, Springer Berlin Heidelberg, 2000, pp. 539–551, http://dx.doi.org/10.1007/3-540-40996-3_46.

[40] D. Marx, I. Razgon, Fixed-parameter tractability of multicut parameterized by the size of the cutset, SIAM J. Comput. 43 (2) (2014) 355–388, http://dx.doi.org/10.1137/110855247.

[41] R. Niedermeier, Invitation to Fixed-Parameter Algorithms, Oxford University Press, 2006, http://dx.doi.org/10.1093/ACPROF:OSO/9780198566076.001.0001.

[42] M. Nöllenburg, M. Sorge, S. Terziadis, A. Villedieu, H.-Y. Wu, J. Wulms, Planarizing graphs and their drawings by vertex splitting, in: Proceedings of the 30th International Symposium on Graph Drawing and Network Visualization, GD 2022, Springer International Publishing, 2023, pp. 232–246, http://dx.doi.org/10.1007/978-3-031-22203-0_17.

[43] F. Protti, M.D. da Silva, J.L. Szwarcfiter, Applying modular decomposition to parameterized cluster editing problems, Theory Comput. Syst. 44 (1) (2009) 91–104, http://dx.doi.org/10.1007/s00224-007-9032-7.

[44] J. Yang, J. Leskovec, Structure and overlaps of ground-truth communities in networks, ACM Trans. Intell. Syst. Technol. 5 (2) (2014) 26:1–26:35, http://dx.doi.org/10.1145/2594454.