Editing Graphs into Disjoint Unions of Dense Clusters

Jiong Guo[1*], Iyad A. Kanj[2**], Christian Komusiewicz[3***], and Johannes Uhlmann[3†]

[1] Universität des Saarlandes,
Campus E 1.4, D-66123 Saarbrücken, Germany
jguo@mmci.uni-saarland.de
[2] School of Computing, DePaul University,
243. S. Wabash Avenue, Chicago, IL 60604, USA
[3] Institut für Informatik, Friedrich-Schiller-Universität Jena
Ernst-Abbe-Platz 2, D-07743 Jena, Germany
{c.komus,johannes.uhlmann}@uni-jena.de ikanj@cs.depaul.edu

Abstract. In the Π-CLUSTER EDITING problem, one is given an undirected graph G, a density measure Π, and an integer $k \geq 0$, and needs to decide whether it is possible to transform G by editing (deleting and inserting) at most k edges into a dense cluster graph. Herein, a dense cluster graph is a graph in which every connected component $K = (V_K, E_K)$ satisfies Π. The well-studied CLUSTER EDITING problem is a special case of this problem with $\Pi :=$"being a clique". In this work, we consider three other density measures that generalize cliques: 1) having at most s missing edges (s-defective cliques), 2) having average degree at least $|V_K| - s$ (average-s-plexes), and 3) having average degree at least $\mu \cdot (|V_K| - 1)$ (μ-cliques), where s and μ are a fixed integer and a fixed rational number, respectively. We first show that the Π-CLUSTER EDITING problem is NP-complete for all three density measures. Then, we study the fixed-parameter tractability of the three clustering problems, showing that the first two problems are fixed-parameter tractable with respect to the parameter (s, k) and that the third problem is W[1]-hard with respect to the parameter k for $0 < \mu < 1$.

1 Introduction

Graph-based data clustering is an important tool for analyzing real-world data, ranging from biological to social network data. In such applications, data items

* Supported by the Excellence Cluster on Multimodal Computing and Interaction (MMCI). Main work was done while the author was with Friedrich-Schiller-Universität Jena.
** Part of this work was done while the author was visiting Friedrich-Schiller-Universität Jena.
*** Supported by a PhD fellowship of the Carl-Zeiss-Stiftung.
† Supported by the DFG, research project PABI, NI 369/7.

are represented as vertices, and there is an edge between two vertices iff the interrelation between the two corresponding data items exceeds some threshold value. A clustering with respect to such a graph means a partition of the vertices into dense subgraphs, also called clusters, such that there are few edges between the clusters. When formulated as a graph modification problem, one thus asks for a minimum-cardinality set of edge modifications, such that the resulting graph is a graph in which every connected component is a cluster. More precisely, the algorithmic task can be formalized as follows:

> Π-CLUSTER EDITING:
> **Input**: An undirected graph $G = (V, E)$, a density measure Π, and an integer $k \geq 0$.
> **Task**: Find a set of at most k edge modifications to transform G into a Π-cluster graph, that is, a graph in which every connected component satisfies Π.

Herein, an edge modification is to insert or delete an edge. Analogously, one defines Π-CLUSTER DELETION by allowing only edge deletions and Π-CLUSTER ADDITION by allowing only edge insertions.

One of the most prominent problems in this context is the NP-hard CLUSTER EDITING problem (also known as CORRELATION CLUSTERING) [2, 16], where the required density measure is $\Pi :=$"being a clique". CLUSTER EDITING finds applications in various fields, such as computational biology [3] and machine learning [2], and has been intensively studied from the viewpoints of polynomial-time approximability as well as parameterized algorithmics. In terms of approximability, the currently best known approximation factor is 2.5 [17]. CLUSTER EDITING can be solved in $O(1.83^k + |E|)$ time [3] and several studies concerning provably efficient and effective preprocessing by data reduction have been performed [8, 11]. Successful experimental studies of the parameterized algorithms for CLUSTER EDITING have been conducted mainly in the context of computational biology [3, 6]. The related CLUSTER DELETION problem is also NP-hard [16].

The density requirement of being cliques has been often criticized for its overly restrictive nature and modeling disadvantages [5, 15]. In this work, we attempt to chart the tractability borderlines of Π-CLUSTER EDITING when the density requirement is relaxed. Therefore, we consider three relaxed density measures, namely, *s-defective cliques*, *average-s-plexes*, and *μ-cliques*. The corresponding modification problems are s-DEFECTIVE CLIQUE EDITING, AVERAGE-s-PLEX EDITING, and μ-CLIQUE EDITING. We study the classical and the parameterized complexity of the aforementioned problems. The proposed density measures may provide more realistic models for practical applications and fixed-parameter tractability (FPT) results can serve as a first step in a series of algorithmic improvements, eventually leading to applicability in practice, as it was the case for CLUSTER EDITING [3, 6, 8, 11]. An overview of our results is given in Table 1. Note that for all three density measures the polynomial-time solvability of the addition problem can be easily seen and is included only for the sake of completeness. In the following, we give the exact definitions of the density

Table 1. The complexity of the problems considered in this work. For s-defective cliques and average-s-plexes, the considered parameter is (s, k), for μ-cliques, the considered parameter is k.

	Deletion	Editing	Addition
s-Defective Clique	NP-complete (Thm. 1) FPT (Thm. 3)	NP-complete (Thm. 1) FPT (Thm. 3)	$\in P$
Average-s-Plex	NP-complete (Thm. 4) FPT (Thm. 5)	NP-complete (Thm. 4) FPT (Thm. 5)	$\in P$
μ-Clique	NP-complete (Thm. 7)	W[1]-hard (Thm. 6)	$\in P$

measures studied in this work, point to related work, and describe our results.

Defective Cliques. The concept of defective cliques has been used in biological networks to represent a clique with exactly one edge missing [18]. Here, we generalize this notion[1] by allowing up to s missing edges: A graph $G = (V, E)$ is called an s-*defective clique*, if G is connected and $|E| \geq |V| \cdot (|V| - 1)/2 - s$. On the negative side, we prove that s-Defective Clique Deletion and Editing are NP-complete. On the positive side however, we show that s-defective cliques can be characterized by forbidden subgraphs of size at most $2(s+1)$, thus showing the fixed-parameter tractability of s-Defective Clique Deletion and Editing with respect to the parameter (s, k).

Average-s-Plexes. With average-s-plexes, we propose a density measure that concerns the *average* degree of a graph $G = (V, E)$, which is defined as $\bar{d} = 2|E|/|V|$. We call a connected graph $G = (V, E)$ an *average-s-plex* if the average degree $\bar{d}$ of G is at least $|V| - s$ for an integer $1 \leq s \leq |V|$. This density measure is a relaxation of the s-plex notion, which demands that the *minimum* degree of a graph $G = (V, E)$ is $|V| - s$. s-Plexes find applications for example in social network analysis [15]. For s-plexes, the clustering problem s-Plex Editing has been previously shown to be NP-hard but fixed-parameter tractable with respect to the parameter (s, k) [12]. Here, we complement this result by showing that Average-s-Plex Deletion and Editing are also NP-hard as well as fixed-parameter tractable with respect to the parameter (s, k). The fixed-parameter tractability result is achieved by a reduction to a more general problem and a subsequent polynomial-time data reduction for the general problem that produces a graph with at most $4k^2 + 8sk$ vertices.

μ-Cliques. With this density measure, we capture the ratio of edges in a graph versus the number of edges in a complete graph of the same size. More precisely, the *density* of a graph $G = (V, E)$ is defined as $2|E|/(|V|(|V| - 1))$. A connected graph $G = (V, E)$ is then called a μ-*clique* for a rational constant $0 \leq \mu \leq 1$ if the density of G is at least μ. We assume that μ is represented by two constant integers a, b such that $\mu = a/b$ (note that a and b are not part of the input). Observe that for $\mu = 0$ every graph is a μ-clique, and that a graph is a 1-clique iff it is a clique. The μ-clique concept was studied for example

[1] Note that Yu et al. [18] introduced a different generalization of defective cliques that is more restrictive than the one considered here.

by Abello et al. [1] and is sometimes also referred to as μ-dense graph [13]. We show that—in contrast to s-Defective Clique Editing and Average-s-Plex Editing—μ-Clique Editing is W[1]-hard and thus presumably fixed-parameter intractable with respect to the parameter k for *any* fixed $0 < \mu < 1$. Note that for $\mu = 1$, the problem is equivalent to Cluster Editing and is thus fixed-parameter tractable. For μ-Clique Deletion we show the NP-hardness, the parameterized complexity remains open.

Preliminaries. We only consider *undirected* graphs $G = (V, E)$, where $n := |V|$ and $m := |E|$. The *(open) neighborhood* $N(v)$ of a vertex $v \in V$ is the set of vertices that are adjacent to v in G. The *degree* of a vertex v, denoted by $\deg(v)$, is the cardinality of $N(v)$. For a set U of vertices, $N(U) := \bigcup_{v \in U} N(v) \setminus U$. We use $N[v]$ to denote the *closed* neighborhood of v, that is, $N[v] := N(v) \cup \{v\}$. For a set of vertices $V' \subseteq V$, the *induced subgraph* $G[V']$ is the graph over the vertex set V' with edge set $\{\{v, w\} \in E \mid v, w \in V'\}$. For $V' \subseteq V$ we use $G - V'$ as an abbreviation for $G[V \setminus V']$ and for a vertex $v \in V$ let $G - v$ denote $G - \{v\}$. A vertex $v \in V(G)$ is called a *cut-vertex* if $G - v$ has more connected components than G. For a graph $G = (V, E)$ let $\overline{G} := (V, \overline{E})$ with $\overline{E} := \{\{u, v\} \mid u, v \in V \wedge u \neq v \wedge \{u, v\} \notin E\}$ denote the *complement graph* of G.

A parameterized problem is *fixed-parameter tractable (FPT)* with respect to a parameter k, if there exists an algorithm solving the problem in time $f(k) \cdot n^{O(1)}$, where n denotes the overall input size and f is a computable function. Downey and Fellows [7] developed a formal framework to show *fixed-parameter intractability*; the basic complexity class for fixed-parameter intractability is called $W[1]$ and there is good reason to believe that $W[1]$-hard problems are not FPT [7, 14].

Due to lack of space, some proofs are deferred to the full version.

2 Defective Cliques

First, we focus on the s-Defective Clique Editing problem. A graph is called an *s-defective clique graph* if every connected component forms an s-defective clique. An edge $\{u, v\} \in \overline{E}$ is called a *missing* edge of G. The following theorem can be obtained by reductions from Cluster Deletion and Editing.

Theorem 1. *s-Defective Clique Deletion and Editing are NP-complete.*

If we delete an arbitrary vertex of an s-defective clique graph, then clearly the resulting graph is still an s-defective clique graph. A graph property that is closed under the operation of deleting vertices (and hence, taking induced subgraphs) is called *hereditary*. It is well known that hereditary graph properties can be described by forbidden induced subgraphs [10]. This means that there exists a set $\mathcal{F}$ of graphs such that a given graph G is an s-defective clique graph iff G is $\mathcal{F}$-free, that is, G does not contain any graph from $\mathcal{F}$ as induced subgraph. A forbidden induced subgraph is *minimal* if each of its proper induced subgraphs is an s-defective clique graph. Clearly, a graph is an s-defective clique graph iff it does not contain any minimal forbidden induced subgraph. Next, we show that, for $s \geq 1$, every minimal forbidden subgraph of s-defective clique graphs contains at most $2(s+1)$ vertices. Note that for $s = 0$ the only forbidden induced subgraph is a path on 3 vertices.

Theorem 2. *For $s \geq 1$, every minimal forbidden induced subgraph of s-defective clique graphs contains at most $2(s+1)$ vertices. Given a graph that is not an s-defective clique graph, a minimal forbidden induced subgraph can be found in $O(nm)$ time.*

Proof. Assume towards a contradiction that there exists a minimal forbidden subgraph $G = (V, E)$ with $|V| > 2(s+1)$. Clearly, we can assume that G is connected, since otherwise we can keep one connected component that is not an s-defective clique and delete all other connected components.

First, we consider the case when G contains a cut-vertex v. Let U denote a set of $s+2$ vertices which together with v induce a connected graph $G' := G[U \cup \{v\}]$ and v remains a cut-vertex in G'. We show that G' is not an s-defective clique graph, a contradiction to the fact that G is minimal (note that $s+3 \leq 2s+2$ for $s \geq 1$). Let $U_1, \ldots, U_\ell$ denote the connected components of $G' - v$. It is not hard to see that there are at least $\frac{1}{2}\sum_{i=1}^{\ell} |U_i| \cdot (|U \setminus U_i|) > s$ edges missing in G', and, hence, G' is not an s-defective clique graph.

In the following, we assume that G does not contain any cut-vertex. Moreover, we can assume that no vertex of G is adjacent to all other vertices of G, since otherwise we can delete it to get a connected graph that has the same number of missing edges as G, thus contradicting the minimality of G. Hence, there are more than $s+1$ missing edges in G, since every vertex is incident to at least one missing edge. Let v be an arbitrary vertex of G and let $A := V \setminus N[v]$. Since the deletion of v results in an s-defective clique graph, it follows that in $G - v$ there are at most s missing edges. Hence, there exists a vertex u that is adjacent to all vertices of $G - v$. Clearly, $u \in A$, since, otherwise, u would be adjacent to all vertices in G. Then, the deletion of u reduces the number of missing edges by one. Thus, $G - u$ is connected and has at least $s+1$ missing edges, and, hence, is not an s-defective clique graph, contradicting the fact that G is minimal.

To find a minimal forbidden induced subgraph proceed as follows. Given a graph $G = (V, E)$ that is not an s-defective clique graph we check for every $v \in V$ whether $G - v$ is an s-defective clique graph in $O(n+m)$ time and delete v if not. It is not hard to observe that we have to consider every vertex at most once. Hence, the overall running time is $O(nm)$. □

The forbidden subgraph characterization given in Theorem 2 directly leads to a search tree algorithm for s-DEFECTIVE CLIQUE EDITING [4].

Theorem 3. s-DEFECTIVE CLIQUE EDITING *and* DELETION *are fixed-parameter tractable with respect to the parameter (s, k).*

3 Average-s-Plexes

Here, we consider the AVERAGE-s-PLEX EDITING problem, showing its NP-completeness and fixed-parameter tractability with respect to (s, k). To this end, we need the following problem, called Equal-Size Clique Editing: Given an undirected graph $G = (V, E)$ and two integers $k, d \geq 0$, decide whether it is possible to transform G, by adding and deleting at most k edges, into a vertex-disjoint union of d cliques which have the same size. The edge deletion version of this problem allows only edge deletions. The NP-completeness of both problems can be shown by a reduction from the well-known Clique problem. Reducing from Equal-Size Clique Editing and Deletion we can show the following.

Theorem 4. Average-s-Plex Editing *and* Deletion *are NP-complete.*

In the following, we describe a fixed-parameter algorithm for Average-s-Plex Editing parameterized by (s, k). Our algorithm consists of two main steps. First, we reduce the original problem to a weighted version. Then, we show the fixed-parameter tractability of the weighted version by describing two polynomial-time data reduction rules that yield instances which contain at most $4k^2 + 8sk$ vertices. Note that being an average-s-plex graph is not a hereditary graph property. Hence, the fixed-parameter tractability of Average-s-Plex Editing and Average-s-Plex Deletion cannot be shown by a forbidden subgraph characterization as in the case of s-Defective Clique Deletion and s-Defective Clique Editing.

We begin with describing a weighted version of Average-s-Plex Editing. We introduce three types of weights: two vertex weights and one edge weight. The idea behind these weight types is the following: whenever there are two vertices in G that cannot be separated by at most k edge modifications, we can merge them into a new "super-vertex", since it is clear that they end up in the same connected component of the solution. We say that a super-vertex v "comprises" a vertex u of the input graph, if u is merged into v. When doing so, we must remember for each such super-vertex v:

- How many vertices of the input graph v comprises,
- How many edges there are between the vertices that v comprises, and
- For each vertex w outside v, how many vertices that v comprises are adjacent to w.

The first two aspects can be remembered by introducing two weights for v, $\sigma(v)$ which keeps track of the number of vertices comprised by v, and $\delta(v)$ which keeps track of the number of edges between these vertices. The third aspect can be stored as the edge weight $\omega(e)$ for the edge $e = \{w, v\}$. Herein, we call a vertex pair having no edge between them a non-edge. Then, edges have edge weights at least one and non-edges have edge weight zero.

The "size" of a vertex set S is then simply defined as $\sigma(S) := \sum_{v \in S} \sigma(v)$. The average degree $\bar{d}(V_i)$ of a connected component V_i can be computed as follows:

$$\bar{d}(V_i) = \frac{2 \sum_{v \in V_i} \delta(v) + \sum_{v \in V_i} \sum_{u \in N(v)} \omega(\{u, v\})}{\sigma(V_i)}.$$

Similar to the definition of average-s-plex graphs, we say that a graph is a weighted average-s-plex graph, if for each connected component V_i, the average degree $\bar{d}(V_i)$ is at least $\sigma(V_i) - s$. For modifying the weighted graph, we allow the following modifications: increasing $\delta(u)$ by one for some $u \in V$, increasing $\omega(\{u,v\})$ by one for some $\{u,v\} \in E$, decreasing $\omega(\{u,v\})$ by one for some $\{u,v\} \in E$, deleting some $\{u,v\} \in E$ with $\omega(\{u,v\}) = 1$, and adding some edge $\{u,v\}$ to E and setting $\omega(\{u,v\}) := 1$. Each of these operations has cost one, and the overall cost of a modification set S is thus exactly $|S|$. The weighted problem version is then defined as

> Weighted Average-s-Plex Editing
> **Input:** A graph $G = (V, E)$, with two vertex-weight functions $\sigma : V \to [1, n]$ and $\delta : V \to [0, n^2]$, an edge weight function $\omega : E \to [1, n^2]$, and a nonnegative integer k.
> **Question:** Is there a set of edge modifications S such that applying S to G yields a weighted average-s-plex graph, and such that $|S| \leq k$?

Observe that we can easily reduce an instance $((V, E), k)$ of Average-s-Plex Editing to an instance of Weighted Average-s-Plex Editing, by setting $\sigma(v) := 1$ and $\delta(v) := 0$ for each $v \in V$, and $\omega(\{u,v\}) := 1$, if $\{u,v\} \in E$; otherwise, $\omega(\{u,v\}) := 0$. Note that this reduction is parameter-preserving, that is, s and k are not changed.

In the following, we present two data reduction rules for Weighted Average-s-Plex Editing which (as we show in Theorem 5) yield instances that contain at most $4k^2 + 8sk$ vertices.

Rule 1 *Remove connected components that are weighted average-s-plexes from G.*

The rule is obviously correct, since no optimal solution modifies any edges incident to vertices of such a connected component.

The second reduction rule identifies two vertices that have a large common neighborhood, or a heavy edge between them and "merges" these vertices into a new "super-vertex".

Rule 2 *If G contains two vertices u and v such that $\omega(\{u,v\}) > k$ or u and v have more than k common neighbors, then remove u from G and set*
– $\sigma(v) := \sigma(u) + \sigma(v)$,
– $\delta(v) := \delta(u) + \delta(v) + \omega(\{u,v\})$, *and*
– $\omega(\{v,w\}) := \omega(\{v,w\}) + \omega(\{u,w\})$ *for each $w \in V \setminus \{u,v\}$.*

To see the correctness of the rule, consider the following: we cannot separate u and v using at most k edge modifications, they thus end up in the same connected component. Hence, we can remove one of them, and store the information about its adjacency in the vertex weights and edge weights of the other vertex.

With these two reduction rules we can show our main result of this section.

Theorem 5. (Weighted) Average-s-Plex Editing *and* Deletion *are fixed-parameter tractable with respect to the parameter (s, k).*

Proof. We first show that a yes-instance I of Weighted Average-s-Plex Editing that is reduced with Rules 1 and 2 contains at most $4k^2 + 8sk$ vertices. Let I be such a reduced instance, and let G be the input graph of I. Since I is a yes-instance, there is a weighted average-s-plex graph G' that can be obtained from G by applying at most k edge modifications. We now bound the size of G'. Herein, we call a vertex v "affected" if v is an endpoint of a modified edge.

First, since G is reduced with respect to Rule 1, there is at least one affected vertex in each connected component of G'. Hence, there can be at most $2k$ connected components in G'.

Next, we show that each connected component of G' contains at most $2k+4s$ vertices. Suppose towards a contradiction that there is a connected component V_i of G' such that $|V_i| > 2k + 4s$. Let $u \in V_i$ be a vertex of maximum degree in V_i. Since G' is a weighted average-s-plex graph, the average vertex degree in $G'[V_i]$ is at least $\sigma(V_i) - s$. Since $|V_i| \leq \sigma(V_i)$, u must be adjacent to at least $|V_i| - s \geq 2k + 3s$ vertices in $G'[V_i]$. We consider two cases for $\sigma(v)$.

Case 1: $\sigma(u) \geq \sigma(V_i)/2$**.** We show that the average degree of $G'[V_i]$ is less than $\sigma(V_i) - s$, contradicting the assumption that G' is a weighted average-s-plex graph. Since G is reduced with respect to Rule 2, there is no edge that has weight at least $k+1$ in G. In G', the edge weights of edges incident to u have increased by at most k overall. Without loss of generality we can assume that the solution distributes these edge weight increases evenly, since for the average degree of a connected component only the *overall* edge and vertex weights are relevant. The maximum edge weight of edges incident to u in G' is thus at most $k + 2$, since u has $2k + 3s$ neighbors in G' and spreading the edge additions equally leads to a weight increase of at most one for each edge. However, with $\sigma(u) \geq \sigma(V_i \setminus \{u\}) \geq 2k + 3s$, this means that for each edge incident to u, the edge weight is at most $\sigma(u)/2$. This leads to a low average degree. More precisely, we can bound the average degree of V_i as

$$\bar{d}(V_i) \overset{(*)}{<} \frac{2\binom{\sigma(V_i)}{2} - (\sigma(u)/2) \cdot \sigma(V_i \setminus \{u\})}{\sigma(V_i)} \overset{(**)}{<} \frac{2\binom{\sigma(V_i)}{2} - (\sigma(u)/2) \cdot 4s}{\sigma(V_i)}$$
$$\overset{(***)}{\leq} \frac{2\binom{\sigma(V_i)}{2} - (\sigma(V_i)/4) \cdot 4s}{\sigma(V_i)} < \sigma(V_i) - s.$$

Inequality (*) follows from assuming the (maximum-density) case that with exception of the edges incident to u all edges are present and have maximum possible weight, and the fact that the maximum edge weight of edges incident to u is $k+1$, which means that for each such edge, a weight of at least $\sigma(u) - (k+1) > \sigma(u)/2$ is "missing". Inequality (**) follows from $\sigma(V_i \setminus \{u\}) \geq |V_i| - 1 > 4s$ and inequality (***) follows from $\sigma(u) \geq \sigma(V_i)/2$.

Case 2: $\sigma(u) < \sigma(V_i)/2$**.** First, we show that there must be at least one other vertex $w \in V_i$ that has at least $|V_i| - 2s$ neighbors in $G'[V_i]$. Suppose otherwise. Then it holds for the average degree of $G'[V_i]$ that

$$\bar{d} \overset{(*)}{\leq} \frac{\sigma(u) \cdot (\sigma(V_i) - 1) + (\sigma(V_i) - \sigma(u)) \cdot (\sigma(V_i) - 2s - 1)}{\sigma(V_i)}$$
$$= \frac{\sigma(V_i) \cdot (\sigma(V_i) - 1) - 2s \cdot (\sigma(V_i) - \sigma(u))}{\sigma(V_i)} \overset{(**)}{<} \frac{\sigma(V_i) \cdot (\sigma(V_i) - s)}{\sigma(V_i)}.$$

Inequality (*) follows from assuming the (maximum-density) case that u is adjacent to all vertices in V_i and that all edge and vertex weights have maximum possible values. Inequality (**) follows from $\sigma(V_i) - \sigma(u) > \sigma(V_i)/2$. We have thus shown that there is at least one vertex w that is adjacent to at least $|V_i|-2s$ vertices in G'. Since u has at least $|V_i| - s$ neighbors in $G'[V_i]$, there must be at least $|V_i| - 3s > 2k + 4s - 3s = 2k + s$ vertices in $G'[V_i]$ that are common neighbors of u and w. Clearly, more than $k+s$ of those vertices are common neighbors of u and w in G. This contradicts that G is reduced with respect to Rule 2.

We have thus shown that a reduced yes-instance contains at most $4k^2 + 8sk$ vertices. We obtain a fixed-parameter algorithm for WEIGHTED AVERAGE-s-PLEX EDITING as follows. First we exhaustively apply the reduction rules, which can clearly be done in polynomial time. If the reduced instance contains more than $4k^2 + 8sk$ vertices, then it is a no-instance. Otherwise, we can solve the problem with running time only depending on s and k, for example by brute-force generation of all possible partitions of the graph. The fixed-parameter tractability of AVERAGE-s-PLEX EDITING, then directly follows from the described reduction to WEIGHTED AVERAGE-s-PLEX EDITING. □

4 μ-Cliques

The main result of this section is that, in contrast to s-DEFECTIVE CLIQUE EDITING and AVERAGE-s-PLEX EDITING, μ-CLIQUE EDITING is fixed-parameter intractable with respect to the number k of allowed edge modifications.

Theorem 6. *For any fixed* $0 < \mu < 1$, μ-CLIQUE EDITING *is NP-complete and W[1]-hard with respect to the number k of allowed edge modifications.*

The reduction used in the proof of Theorem 6 does not work for the edge deletion case. However, we can establish the NP-hardness of μ-CLIQUE DELETION by a reduction from the NP-complete MAXIMUM TRIANGLE PACKING problem [9].

Theorem 7. *For any fixed* $0 < \mu < 1$, μ-CLIQUE DELETION *is NP-complete.*

5 Outlook

There are numerous topics for future research, we only point out some of them. For s-DEFECTIVE CLIQUE EDITING and AVERAGE-s-PLEX EDITING clearly further algorithmic improvements are necessary. For instance, s-DEFECTIVE CLIQUE EDITING is still missing a non-trivial kernelization algorithm, whereas for AVERAGE-s-PLEX EDITING a solution algorithm other than the brute-force one that can be applied to the reduced instance is needed. As a next step, experimental studies should then be undertaken regarding the running time of the algorithms and the quality of the produced clusterings. For μ-CLIQUE EDITING, other parameterizations should be studied. Finally, the parameterized complexity of μ-CLIQUE DELETION remains open.

References

[1] J. Abello, M. G. C. Resende, and S. Sudarsky. Massive quasi-clique detection. In *Proc. 5th LATIN*, volume 2286 of *LNCS*, pages 598–612. Springer, 2002.

[2] N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine Learning*, 56(1–3):89–113, 2004.

[3] S. Böcker, S. Briesemeister, Q. B. A. Bui, and A. Truß. Going weighted: Parameterized algorithms for cluster editing. *Theor. Comput. Sci.*, 2009. To appear.

[4] L. Cai. Fixed-parameter tractability of graph modification problems for hereditary properties. *Inf. Process. Lett.*, 58(4):171–176, 1996.

[5] E. J. Chesler, L. Lu, S. Shou, Y. Qu, J. Gu, J. Wang, H. C. Hsu, J. D. Mountz, N. E. Baldwin, M. A. Langston, D. W. Threadgill, K. F. Manly, and R. W. Williams. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nature Genetics*, 37(3):233–242, 2005.

[6] F. K. H. A. Dehne, M. A. Langston, X. Luo, S. Pitre, P. Shaw, and Y. Zhang. The cluster editing problem: Implementations and experiments. In *Proc. 2nd IWPEC*, volume 4169 of *LNCS*, pages 13–24. Springer, 2006.

[7] R. G. Downey and M. R. Fellows. *Parameterized Complexity*. Springer, 1999.

[8] M. R. Fellows, M. A. Langston, F. A. Rosamond, and P. Shaw. Efficient parameterized preprocessing for Cluster Editing. In *Proc. 16th FCT*, volume 4639 of *LNCS*, pages 312–321. Springer, 2007.

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

[10] D. L. Greenwell, R. L. Hemminger, and J. B. Klerlein. Forbidden subgraphs. In *Proceedings of the 4th Southeastern Conference on Combinatorics, Graph Theory and Computing*, pages 389–394, 1973.

[11] J. Guo. A more effective linear kernelization for Cluster Editing. *Theor. Comput. Sci.*, 410(8-10):718–726, 2009.

[12] J. Guo, C. Komusiewicz, R. Niedermeier, and J. Uhlmann. A more relaxed model for graph-based data clustering: s-plex editing. In *Proc. 5th AAIM*, volume 5564 of *LNCS*, pages 226–239. Springer, 2009.

[13] S. Kosub. Local density. In *Network Analysis*, volume 3418, pages 112–142. Springer, 2004.

[14] R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*. Oxford University Press, 2006.

[15] S. B. Seidman and B. L. Foster. A graph-theoretic generalization of the clique concept. *Journal of Mathematical Sociology*, 6:139–154, 1978.

[16] R. Shamir, R. Sharan, and D. Tsur. Cluster graph modification problems. *Discrete Appl. Math.*, 144(1–2):173–182, 2004.

[17] A. van Zuylen and D. P. Williamson. Deterministic algorithms for rank aggregation and other ranking and clustering problems. In *Proc. 5th WAOA*, volume 4927 of *LNCS*, pages 260–273. Springer, 2008.

[18] H. Yu, A. Paccanaro, V. Trifonov, and M. Gerstein. Predicting interactions in protein networks by completing defective cliques. *Bioinformatics*, 22(7):823–829, 2006.