# On the Complexity of Various Parameterizations of Common Induced Subgraph Isomorphism

Faisal N. Abu-Khzam[1], Édouard Bonnet[2] and Florian Sikora[2]

[1] Lebanese American University, Beirut, Lebanon
faisal.abukhzam@lau.edu.lb
[2] PSL, Université Paris-Dauphine, LAMSADE, UMR CNRS 7243, France
{edouard.bonnet,florian.sikora}@dauphine.fr

**Abstract.** MAXIMUM COMMON INDUCED SUBGRAPH (henceforth MCIS) is among the most studied classical NP-hard problems. MCIS remains NP-hard on many graph classes including bipartite graphs, planar graphs and $k$-trees. Little is known, however, about the parameterized complexity of the problem. When parameterized by the vertex cover number of the input graphs, the problem was recently shown to be fixed-parameter tractable. Capitalizing on this result, we show that the problem does not have a polynomial kernel when parameterized by vertex cover unless NP $\subseteq$ coNP/$poly$. We also show that MAXIMUM COMMON CONNECTED INDUCED SUBGRAPH (MCCIS), which is a variant where the solution must be connected, is also fixed-parameter tractable when parameterized by the vertex cover number of input graphs. Both problems are shown to be W[1]-complete on bipartite graphs and graphs of girth five and, unless P = NP, they do not to belong to the class XP when parameterized by a bound on the size of the minimum feedback vertex sets of the input graphs, that is solving them in polynomial time is very unlikely when this parameter is a constant.

## 1 Introduction

A common induced subgraph of two graphs $G_1$ and $G_2$ is a graph that is isomorphic to induced subgraphs of each. The problem of finding a common induced subgraph of maximum number of vertices (or edges) has many applications in a number of domains including bioinformatics and chemistry [10,11,16,12,15]. In the decision version of the problem, we are given an integer $k$ and the question is to decide if there is a solution with at least $k$ vertices. We say that $k$ is the natural parameter of the problem, that is the solution size.

Concerning its classical complexity, MAXIMUM COMMON INDUCED SUBGRAPH is NP-complete, and remains so on bipartite graphs and graphs with bounded treewidth. However, the problem is in P for trees [9] and graphs of (both) bounded treewidth and bounded degree [2].

A decision subproblem of MAXIMUM COMMON INDUCED SUBGRAPH is the well known INDUCED SUBGRAPH ISOMORPHISM (ISI) problem, which consists of deciding whether $G_1$ is isomorphic to an induced subgraph of $G_2$. In other words,

it is equivalent to MAXIMUM COMMON INDUCED SUBGRAPH where $k = |G_1|$. In this case $G_1$ is called the pattern graph while $G_2$ is the host graph. ISI is W[1]-hard in general, by a straightforward reduction from $k$-Clique. Therefore MCIS is also W[1]-hard. On the other hand, if ISI is in FPT on a certain graph class, then so is MCIS. To see this, note that an arbitrary instance $(G_1, G_2, k)$ of MCIS can be reduced in $O(f(k))$-time to two instances of ISI by enumerating all possible graphs on $k$ vertices and checking whether each is an induced subgraph of each of the two input graphs. This implies that ISI and MCIS have the same parameterized complexity when parameterized by the solution size, which we refer to as the natural parameter in this paper. Of course, the latter FPT-reduction takes time $O(2^{k^2})$ (multiplied by the time needed to solve ISI on the given graph class), which makes it prohibitively impractical. We shall provide a simpler reduction that takes $O(c^k)$-time on a class of graphs that includes $H$-minor free graphs and graphs of bounded degree.

Another way to deal with the hardness of a problem is to study its complexity with respect to auxiliary (or structural) parameters, to better understand the behavior of the problem (see for example [7]). MCIS is already hard on graphs with bounded treewidth, being NP-hard on forests, as we shall observe based on a classical result from Garey and Johnson [9]. Accordingly, the problem is W[1]-hard when parameterized by the treewidth of the input graphs. Therefore we need to look for bigger parameters. We shall study the problem with respect to the size of a (minimum) feedback vertex set that of a (minimum) vertex cover of input graphs. We observe that MCIS is not in XP when parameterized by the feedback vertex set number of the input graphs. This also implies that the problem is not in XP when parameterized by treewidth.

We observe that ISI remains W[1]-hard on graphs where $G_1$ has a $k$-vertex cover by a reduction from the W[1]-hard INDUCED BIPARTITE MATCHING problem [13]: if the pattern consists of $k$ disjoint edges, its vertex cover is $k$. Therefore, MCIS is W[1]-hard when the parameter is the vertex cover of one of the input graphs, even if the other graph is bipartite. However, if the parameter $k$ is the combination of the vertex cover of both input graphs, then the problem is in FPT, with a running time of $O((24k)^k)$ [1]. We shall prove in Section 3 that MCIS does not have a polynomial-size kernel in this case unless NP $\subseteq$ coNP/$poly$.

We also consider the MAXIMUM COMMON CONNECTED INDUCED SUBGRAPH problem. We observe that the problem is in FPT on graphs of bounded degree and show it to be W[1]-complete on the class of bipartite graphs, even if the input graph is $C_4$-free. Consequently, MCCIS is W[1]-complete on graphs of girth five. Finally, we show that MCCIS is fixed-parameter tractable when parameterized by a bound on the minimum vertex covers of the input graphs.

## 2 Preliminaries

Two finite graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are *isomorphic* if there is a bijection $\pi : V_1 \rightarrow V_2$ such that $\forall u, v \in V_1 : uv \in E_1 \Leftrightarrow \pi(u)\pi(v) \in E_2$. Given a graph $G = (V, E)$, a graph $G' = (V', E')$ is an *induced subgraph* of $G$

if $V' \subseteq V$ and $E' = E(V') = \{uv \in E \mid u, v \in V'\}$, i.e. $E'$ is the edge set with both extremities in $V'$. We also say that $G'$ is the subgraph of $G$ induced by $V'$.

The *girth* of a graph $G$ is the length of the shortest cycle contained in $G$. Contracting an edge $uv$ consists of deleting $uv$ and replacing the vertices $u$ and $v$ by a single vertex $w$ in the incidence relation (edges incident on $u$ or $v$ become incident on $w$). A graph $H$ is a *topological minor* of graph $G$ if $H$ is obtained from a subgraph of $G$ by applying zero or more edge contractions. Given a fixed graph $H$, a family $\mathcal{F}$ of graphs is said to be *$H$-minor free* if $H$ is not a minor of any element of $\mathcal{F}$.

The MAXIMUM COMMON INDUCED SUBGRAPH problem is defined formally as follows.

---

MAXIMUM COMMON INDUCED SUBGRAPH (MCIS):
• **Input**: Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$.
• **Output**: An induced subgraph $G_1'$ of $G_1$ isomorphic to an induced subgraph $G_2'$ of $G_2$ with a maximum number of vertices.

---

Another way of defining the problem would be the following: given two graphs, find a mapping between vertex subsets of the two graphs such that these subsets have a maximum size (the subgraph is maximum) and the following restrictions are respected: each vertex in one graph can be mapped at most to one vertex in the other graph (the subgraphs are isomorphic) and a pair of adjacent vertices in one of the graphs cannot be mapped to a pair of non-adjacent vertices in the other graph, and vice versa (the subgraph is induced in both graphs).

MAXIMUM COMMON CONNECTED INDUCED SUBGRAPH (MCCIS) is defined as MCIS with the additional restriction that the solution must be connected.

INDUCED SUBGRAPH ISOMORPHISM is defined similarly:

---

INDUCED SUBGRAPH ISOMORPHISM (ISI):
• **Input**: Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$.
• **Output**: An induced subgraph $G_1'$ of $G_1$ isomorphic to $G_2$.

---

*Parameterized complexity.* A parameterized problem $(I, k)$ is said *fixed-parameter tractable* (or in the class FPT) w.r.t. parameter $k$ if it can be solved in $f(k) \cdot |I|^c$ time, where $f$ is any computable function and $c$ is a constant (see [14,6] for more details about fixed-parameter tractability). The parameterized complexity hierarchy is composed of the classes FPT $\subseteq$ W[1] $\subseteq$ W[2] $\subseteq \cdots \subseteq$ XP. The class XP contains problems solvable in time $f(k) \cdot |I|^{g(k)}$, where $f$ and $g$ are unrestricted functions. A W[1]-hard problem is not fixed-parameter tractable (unless FPT $=$ W[1]) and one can prove W[1]-hardness by means of a *parameterized reduction* from a W[1]-hard problem. This is a mapping of an instance $(I, k)$ of a problem $A_1$ in $g(k) \cdot |I|^{O(1)}$ time (for any computable function $g$) into an instance $(I', k')$ for $A_2$ such that $(I, k) \in A_1 \Leftrightarrow (I', k') \in A_2$ and $k' \leq h(k)$ for some function $h$.

A powerful technique to design parameterized algorithms is *kernelization*. In short, kernelization is a polynomial-time self-reduction algorithm that takes an

instance $(I, k)$ of a parameterized problem $P$ as input and computes an equivalent instance $(I', k')$ of $P$ such that $|I'| \leqslant h(k)$ for some computable function $h$ and $k' \leqslant k$. The instance $(I', k')$ is called a *kernel* in this case. If the function $h$ is polynomial, we say that $(I', k')$ is a polynomial kernel. It is well known that a problem is in FPT iff it has a kernel, but this equivalence yields superpolynomial kernels (in general). To design efficient parameterized algorithms, a kernel of polynomial (or even linear) size in $k$ is important. However, some lower bounds on the size of the kernel can be shown unless some polynomial hierarchy collapses. To show this result, we will use the cross composition technique developed by Bodlaender et al. [3].

**Definition 1 (Polynomial equivalence relation [3]).** *An equivalence relation $\mathcal{R}$ on $\Sigma^*$ is said to be* polynomial *if the following two conditions hold: (i) There is an algorithm that given two strings $x, y \in \Sigma^*$ decides whether $x$ and $y$ belong to the same equivalence class in time $(|x| + |y|)^{O(1)}$. (ii) For any finite set $S \subseteq \Sigma^*$ the equivalence relation $\mathcal{R}$ partitions the elements of $S$ into at most $(\max_{x \in S} |x|)^{O(1)}$ classes.*

**Definition 2 (OR-cross-composition [3]).** *Let $L \subseteq \Sigma^*$ be a set and let $Q \subseteq \Sigma^* \times \mathbb{N}$ be a parameterized problem. We say that $L$ cross-composes into $Q$ if there is a polynomial equivalence relation $\mathcal{R}$ and an algorithm which, given $t$ strings $x_1, x_2, \ldots, x_t$ belonging to the same equivalence class of $\mathcal{R}$, computes an instance $(x^*, k^*) \in \Sigma^* \times \mathbb{N}$ in time polynomial in $\sum_{i=1}^{t} |x_i|$ such that: (i) $(x^*, k^*) \in Q \Leftrightarrow x_i \in L$ for some $1 \leqslant i \leqslant t$. (ii) $k^*$ is bounded by a polynomial in $\max_{i=1}^{t} |x_i| + \log t$.*

**Proposition 1 ([3]).** *Let $L \subseteq \Sigma^*$ be a set which is NP-hard under Karp reductions. If $L$ cross-composes into the parameterized problem $Q$, then $Q$ has no polynomial kernel unless NP $\subseteq$ coNP/poly.*

A parameterized problem is said to be *fixed-parameter enumerable* if all feasible solutions can be enumerated in $O(f(k)|I|^c)$ where $f$ is a computable function of the parameter $k$ only, and $c$ is a constant.

## 3 Structural parameterization of Maximum Common Induced Subgraph

Let us first recall that $tw(G) \leqslant fvs(G) \leqslant vc(G)$, where $tw(G)$ (resp. $fvs(G)$, $vc(G)$) represents the treewidth (resp. the feedback vertex set number, the vertex cover number) of $G$. As noted before, if the parameter is the combination of $tw(G_1)$ and $tw(G_2)$ then MCIS is known to be W[1]-hard. Even more, if the parameter is the combination of $fvs(G_1)$ and $fvs(G_2)$ (which is bigger than the combination of the treewidth), then the problem is not even in XP since MAX-IMUM COMMON INDUCED SUBGRAPH and INDUCED SUBGRAPH ISOMORPHISM are NP-hard on forests, a case where the parameter is equal to 0. Indeed, one can modify the reduction from 3-PARTITION done by Garey and Johnson in [9] for

SUBFOREST ISOMORPHISM to our problem, by building chains of $B + 3$ vertices instead of $B + 1$ in $G_2$ such that each chain of $G_1$ is separated by a vertex. The following theorem follows.

**Theorem 1.** *Unless* $\mathsf{P} = \mathsf{NP}$*,* MAXIMUM COMMON INDUCED SUBGRAPH *is not in* XP *when parameterized by a bound on the minimum feedback vertex sets of the pair of input graphs.*

Knowing that graphs with feedback vertex set $k$ also have treewidth $\leqslant k$, we obtain the following.

**Corollary 1.** *Unless* $\mathsf{P} = \mathsf{NP}$*,* MAXIMUM COMMON INDUCED SUBGRAPH *is not in* XP *when parameterized by the treewidth of the input graphs.*

It was shown in [1] that MCIS is in FPT if the parameter is the combination of $vc(G_1)$ and $vc(G_2)$. Accordingly, the problem has a kernel, but no polynomial bound is known on its size. We show that, in this case, the kernel cannot be polynomial unless $\mathsf{NP} \subseteq \mathsf{coNP}/poly$.

**Theorem 2.** *Unless* $\mathsf{NP} \subseteq \mathsf{coNP}/poly$*,* MAXIMUM COMMON INDUCED SUBGRAPH *has no polynomial kernel when parameterized by the sum of the sizes of vertex covers in the two input graphs.*

*Proof.* We will define an OR-cross-composition from the NP-complete CLIQUE, problem, where the given instance is a tuple $(G^c, l)$ and the question is whether the graph $G^c$ contains a clique on $l$ vertices.

Given $t$ instances, $(G_1^c, l_1), (G_2^c, l_2), \ldots, (G_t^c, l_t)$, of CLIQUE, where $G_i^c$ is a graph and $l_i \in \mathbb{N}, \forall 1 \leqslant i \leqslant t$, we define our equivalence relation $\mathcal{R}$ such that any strings that are not encoding valid instances are equivalent, and $(G_i^c, l_i), (G_j^c, l_j)$ are equivalent iff $|V(G_i^c)| = |V(G_j^c)|$, and $l_i = l_j$. Hereafter, we assume that $V(G_i^c) = [n]$ and $l_i = l$, for any $1 \leqslant i \leqslant t$. We will build an instance of MAXIMUM COMMON INDUCED SUBGRAPH parameterized by the vertex cover $(G_1, G_2, l', Z)$ where $G_1$ and $G_2$ are two graphs, $l' \in \mathbb{N}$ and $Z \subseteq V(G_2)$ is a vertex cover of $G_2$ computed in fpt-time, such that there is a solution of size $l'$ for MAXIMUM COMMON INDUCED SUBGRAPH iff there is an $i, 1 \leqslant i \leqslant t$ such that there is a solution of size $l$ in $G_i^c$. We will now describe how to build $G_1$ and $G_2$.

To build $G_2$ (see also Fig. 1):

- $V(G_2) = \{p, q, r\} \cup \{t_i \mid 1 \leqslant i \leqslant t\} \cup \{e_{uv} \mid 1 \leqslant u < v \leqslant n\} \cup \{v_i \mid 1 \leqslant i \leqslant n\}$,
- $E(G_2)_1 = \{pq, pr, qr\}$,
- $E(G_2)_2 = \{rt_i \mid 1 \leqslant i \leqslant t\}$,
- $E(G_2)_3 = \{t_i e_{uv} \mid uv \in E(G_i^c)\}$,
- $E(G_2)_4 = \{e_{uv}v_u, e_{uv}v_v \mid \forall 1 \leqslant u < v \leqslant n\}$,
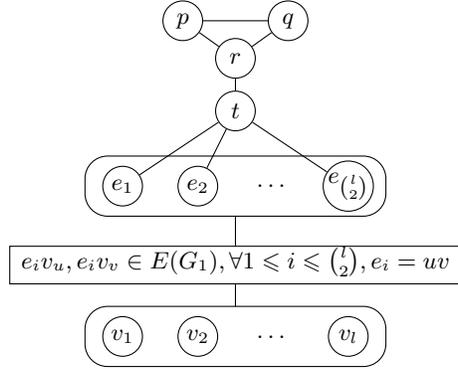- $E(G_2) = E(G_2)_1 \cup E(G_2)_2 \cup E(G_2)_3 \cup E(G_2)_4$.

To build $G_1$ (see also Fig. 2):

- $V(G_1) = \{p, q, r, t\} \cup \{e_i \mid 1 \leqslant i \leqslant \binom{l}{2}\} \cup \{v_i \mid 1 \leqslant i \leqslant l\}$,
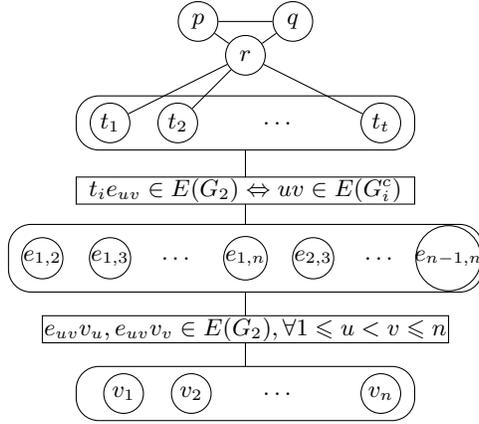
**Fig. 1.** Illustration of the construction of $G_2$.

- $E(G_1)_1 = \{pq, pr, qr, rt\}$,
- $E(G_1)_2 = \{te_i \mid 1 \leqslant i \leqslant \binom{l}{2}\}$,
- $E(G_1)_3 = \{e_i v_u, e_i v_v \mid \forall 1 \leqslant i \leqslant \binom{l}{2}, e_i = uv\}$,
- $E(G_1) = E(G_1)_1 \cup E(G_1)_2 \cup E(G_1)_3$.



**Fig. 2.** Illustration of the construction of $G_1$.

We set $l' = |V(G_1)|$, and $Z = \{p, r\} \cup \{e_{uv} | 1 \leqslant u < v \leqslant n\}$. It is easy to see that $Z$ is indeed a vertex cover for $G_2$ and that its size is equal to $\frac{n(n-1)}{2} + 2$, which is polynomial in $n$ and hence in the size of the largest instance. Note that the size of the graph $G_1$ does not depend on $t$ and is polynomial in $n$, so the size of its vertex cover is also polynomial in $n$ and independent in $t$.

Let us show that $G_1$ is an induced subgraph of $G_2$ iff at least one of the $G_i^c$'s has a clique of size $l$.

($\Leftarrow$) Suppose that $G_i^c$ has a clique of size $l$. We denote by $S \subseteq V(G_i^c)$ a clique of size exactly $l$ in $G_i^c$. We show that there is an induced subgraph $S'$ of $G_2$ of size $l'$, isomorphic to $G_1$. We set $V(S') = \{p, q, r\} \cup \{t_i\} \cup \{e_{uv} \mid \forall uv \in E(S)\} \cup \{v_u \mid u \in S\}$. One can easily check that this subgraph is isomorphic to $G_1$.

($\Rightarrow$) Assume now that $G_1$ is an induced subgraph of $G_2$. Denote by $S'$ the subgraph of $G_2$ isomorphic to $G_1$. Note that the only triangle in $G_2$ is $pqr$. Indeed, $T(V(G_2) \setminus \{p\})$ is bipartite. The triangle $pqr$ in $G_1$ has therefore to match $pqr$ in $G_2$. Moreover, $r$ in $G_1$ has to match $r$ in $G_2$ since $p$ and $q$ have no edges besides the clique $pqr$. The vertex $t$ in $G_1$ can only match a vertex $t_i$ for some $i \in [t]$. Then, $e_1$ up to $e_{\binom{l}{2}}$ in $G_1$ has to match $\binom{l}{2}$ vertices in $\{e_{uv} \mid 1 \leqslant u < v \leqslant n\}$ of $G_2$ which correspond to actual edges in $G_i^c$. Finally, $v_1$ up to $v_l$ in $G_1$ has to match $l$ vertices among the $v_j$'s in $G_2$. Note that the number of edges in $E(G_1)_3$ between the $e_j$'s and the $v_j$'s is exactly $2\binom{l}{2} = l(l-1)$. More precisely, each $e_j$ touches 2 edges in $E(G_1)_3$ and each $v_j$ touches $l-1$ edges in $E(G_1)_3$. In order to get a match in $G_2$, one should find a set of $\binom{l}{2}$ edges inducing exactly $l$ vertices. So, this set of $l$ vertices is a clique in $G_i^c$. $\qquad\square$

Note that the parameter of MCIS in this reduction is exactly the size of $G_1$. Therefore, this negative result holds for ISI too.

Despite the fact that ISI and MCIS have the same parameterized complexity when parameterized by the natural parameter, they exhibit different complexities with respect to structural parameters. In fact, the latter is not even in XP when parameterized by the vertex cover of only one of the two graphs while ISI is FPT when parameterized by the vertex cover of the second (host) graph. To see this, note that when the host graph has a $k$-vertex cover, the minimum size of a vertex cover in the pattern graph must be bounded by the parameter $k$, otherwise we have a no instance. The claim follows from the fixed-parameter tractability of MCIS in this case [1].

Although MCIS is not in XP w.r.t. some structural parameters such as treewidth and feedback vertex set number, it is, together with MCCIS and ISI, in W[1] w.r.t. the natural parameter.

**Theorem 3.** *MCIS, MCCIS and ISI are* W[1]*-complete w.r.t the natural parameter.*

*Proof.* By the previous discussion in the introduction, it suffices to show membership in W[1]. In [4], it is shown that if a problem can be reduced in FPT time to simulating a non-deterministic single-taped Turing Machine halting in at most $f(k)$ steps, for some function $f$, then it is in W[1]. The Turing Machine can have an alphabet and a set of states of size depending on the size of the input of the initial problem. In our case, we can design a Turing Machine that guesses in $2k$ steps the right $k$ vertices in $G_1$ (for ISI this part is not necessary) and the right $k$ vertices in $G_2$ (our alphabet being isomorphic to $V(G_1) \cup V(G_2)$) and then check in time, say, $k!k^2$ that the two induced subgraphs are isomorphic (and that they are connected for MCCIS). $\qquad\square$

We now turn our attention to the case where the MCIS is parameterized by a combination of the natural parameter and some structural parameter. For example, consider the case where the parameter is the sum of some bound $t$ on the feedback vertex set of the input graphs and the natural parameter $k$. The problem is FPT in this case since graphs of $t$-feedback vertex set are $H$-minor free (let $H$ be the "fixed" graph consisting of a disjoint union of $t+1$ triangles). Moreover, we know ISI is FPT in this case due to [8]. However, and as stated in Section 1, a solution to an instance of MCIS (in this case) is obtained via an exhaustive enumeration of $O(2^{k^2})$ instances of ISI. This can be improved on classes of graphs that are given with some fixed coloring $t$, such as $H$-minor free graphs and graphs of bounded maximum degree. In fact, if the input to MCIS is a pair of $c$-colored graphs, then a reduction algorithm would first check whether each of the two graphs has an (independent) color class of size $k$. If so, then both have an edgeless common subgraph of size $k$. Otherwise, the order of at least one of the two graphs, say $G_1$, is smaller than $ck$. In such case, the algorithm proceeds by running a (fixed-parameter) algorithm for ISI on each of the $O(2^{tk})$ induced subgraphs of $G_1$.

In [13] it was shown that INDUCED MATCHING is W[1]-hard on bipartite graphs. As mentioned earlier, this proves that MCIS is W[1]-hard in this case. We show that MCIS remains W[1]-hard on $C_4$-free bipartite graphs, which proves its W[1]-hardness on graphs of girth five.

**Theorem 4.** MAXIMUM COMMON INDUCED SUBGRAPH *is* W[1]-*complete w.r.t. size of the solution, as parameter, even on $C_4$-free bipartite graphs.*

*Proof.* Membership in W[1] comes from Theorem 3. For the hardness, consider the following reduction from the W[1]-hard problem CLIQUE. Given an instance $(G = (V, E), k)$ of CLIQUE, we build an instance $(G_1, G_2, k')$ of our problem as follows. The graph $G_2$ is the bipartite incidence graph of $G$ (the bipartition is between vertices representing $V$ and vertices representing $E$), the graph $G_1$ is the bipartite incidence graph of $K_k$, and $k' = k + \binom{k}{2} = |V(G_1)|$.

Note that a bipartite incidence graph is $C_4$-free since, in a simple graph, no two edges are incident on the same pair of vertices.

It is clear that $G_1$ occurs as a connected induced subgraph of $G_2$ iff there is a clique of size $k$ in $G$, because w.l.o.g. $k > 2$ and the vertices representing edges in $G_1$ and $G_2$ are of degree 2. $\qquad\square$

**Corollary 2.** MAXIMUM COMMON INDUCED SUBGRAPH *is* W[1]-*complete on graphs of girth five.*

## 4 Maximum Common Connected Induced Subgraph

MAXIMUM COMMON CONNECTED INDUCED SUBGRAPH is trivially FPT whenever INDUCED SUBGRAPH ISOMORPHISM is FPT, including $H$-minor free graphs, since the enumeration of all $O(2^{k^2})$ possible induced connected subgraphs can be

used as described before. The converse is also true. In fact, an instance $(G_1, G_2, k)$ of ISI can be reduced to an equivalent instance $(G_1', G_2', k+1)$ of MCCIS by letting $G_i'$ be the graph obtained by adding a single (universal) vertex to $G_i$ that is made adjacent to all other vertices of $G_i$. It follows that MCIS and MCCIS have the same parameterized complexity with respect to the natural parameter (i.e., solution size). In this section we study the complexity of MCCIS with respect to structural parameters.

**Lemma 1.** INDUCED CONNECTED SUBGRAPH ISOMORPHISM *is* NP-*hard even when both graphs have feedback vertex set number equal to one.*

*Proof.* Given an instance of INDUCED SUBGRAPH ISOMORPHISM on forests $G_1$ and $G_2$ (each with at least 2 trees), we build an instance of INDUCED CONNECTED SUBGRAPH ISOMORPHISM by adding a universal vertex (connected to every node) in $G_1$ and in $G_2$. One can see that these two universal vertices must be matched together since they are the only ones with sufficient high degree. Then, there is a solution for INDUCED SUBGRAPH ISOMORPHISM iff there is a solution for INDUCED CONNECTED SUBGRAPH ISOMORPHISM. The result of course holds for MCCIS too. □

**Corollary 3.** *Unless* P = NP, MAXIMUM COMMON CONNECTED INDUCED SUBGRAPH *is not in* XP *when parameterized by a bound of the minimum feedback vertex set number of the input graphs (and then when parameterized by a bound on the treewidth of each of the two input graphs).*

Given the above negative result, the next question is whether MCCIS is in FPT w.r.t. the parameter vertex cover. In [1], a parameterized algorithm is presented for MCIS when the parameter is a bound on the minimum vertex cover number of the input graphs. However, that algorithm cannot help us much for solving MCCIS since it relies on the existence of a feasible solution of size at least $\approx n - k$ which consists of mapping the two *big* independent sets of the two graphs with each other. Of course, this is not a feasible solution for MCCIS. In the following we prove that MCCIS is fixed-parameter tractable w.r.t $k = \max(vc(G_1), vc(G_2))$.

**Theorem 5.** MAXIMUM COMMON CONNECTED INDUCED SUBGRAPH *parameterized by a bound on the vertex covers of the input graphs is fixed-parameter tractable.*

*Proof.* In time $O^*(2^k)$ (even $O^*(1.2738^k)$ [5]), we can find minimum vertex covers $C_1$ and $C_2$ in $G_1$ and $G_2$ respectively. Let $I^{(j)}$ be the independent set $V(G_j) \setminus C_j$ for $j \in \{1, 2\}$. By assumption, our parameter $k$ is $\max(C_1, C_2)$, so we can enumerate all tripartitions of $C_1$ and $C_2$ in time $O^*(9^k)$. We denote by $C_{1,m}$, $C_{1,u}$ and $C_{1,i}$ (respectively $C_{2,m}$, $C_{2,u}$ and $C_{2,i}$) the three sets of a tripartition of $C_1$ (respectively $C_2$). For $j \in \{1, 2\}$, $C_{j,u}$ corresponds to the vertices of $C_j$ that are not matched, so they may be deleted. $C_{j,m}$ comprises the vertices matched to $C_{3-j,m}$ (that is, to the vertex cover of the other graph), and $C_{j,i}$ are the vertices matched to $I^{(3-j)}$, the independent set of the other graph. See Figure 3.

We observe that for $j \in \{1, 2\}$, $I^{(j)}$ can be partitioned into at most $2^k$ classes of twins: $I_1^{(j)}, I_2^{(j)}, \ldots I_{2^k}^{(j)}$. A class of twins in this context is a set of vertices with an identical neighborhood in the vertex cover and there are at most $2^k$ subsets of $C_j$. Potentially, some classes can be empty: they correspond to a subset of the vertex cover $C_j$ that is not the (exact) neighborhood of any vertex in $I^{(j)}$.

At this point, we can enumerate the mappings between $C_{1,m}$ and $C_{2,m}$ in time $O^*(k^k)$ and the mappings between $C_{j,i}$ and $I^{(3-j)}$ in time $O^*(k^{2^k})$. Indeed, to match a vertex $u$ with a vertex $v$ or a twin of $v$ is equivalent. Thus, in time $O^*((9k)^k k^{2^k})$ we can enumerate all the solutions of MCIS where only vertices of $I^{(1)}$ could still be matched to vertices of $I^{(2)}$. The optimal map of the independent sets can be done in linear time by matching the greatest number of vertices in each *equivalent* twin class (which is the size of the smaller of the two equivalent twin classes), where a twin class $I_r^{(j)}$ in $I^{(j)}$ is equivalent to a twin class $I_s^{(3-j)}$ in $I^{(3-j)}$ if the vertices of $N(I_r^{(j)}) \setminus C_{j,u}$ and $N(I_s^{(3-j)}) \setminus C_{3-j,u}$ are in one-to-one correspondence. $\qquad \square$
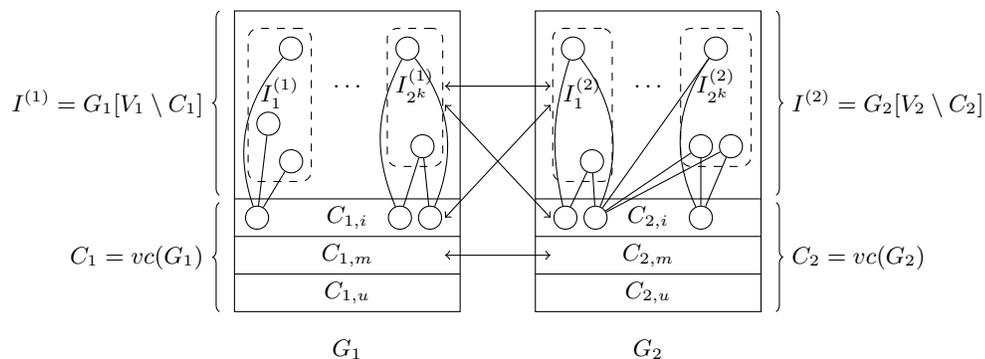


**Fig. 3.** Illustration of the proof of Theorem 5. Dashed boxes represent the classes inside the independent set. Arrows represent the matching between sets of vertices.

To find a solution for MCCIS, the algorithm described in the above proof enumerates all possible maximal common induced subgraphs in time $O^*((9k)^k k^{2^k})$. As such, it can be used as an enumeration algorithm for MCIS.

**Theorem 6.** MAXIMUM COMMON INDUCED SUBGRAPH *parameterized by vertex cover, is fixed-parameter enumerable.*

Finally, the following corollaries follow easily from the proofs of Theorem 2 and Theorem 4 since the graphs used in both proofs are connected.

**Corollary 4.** MAXIMUM COMMON CONNECTED INDUCED SUBGRAPH, *parameterized by a bound on the minimum vertex covers of input graphs, does not have a polynomial-size kernel unless* $\mathsf{NP} \subseteq \mathsf{coNP}/poly$.

**Corollary 5.** Maximum Common Connected Induced Subgraph *is* W[1]-*complete on bipartite graphs and graphs of girth five.*

Let $fvs + vc$ be the sum of the sizes of a feedback vertex set of $G_1$ and vertex cover of $G_2$. Then MCIS and MCCIS are in XP when parameterized by $fvs + vc$. In fact we can employ a method similar to the proof of Theorem 5 to show that both problems are in XP when one of the two graphs only has a $k$-vertex cover. Simply, we start by grouping the elements of $V(G_2) \setminus C_2$ into at most $2^k$ classes based on their (at most $2^k$) neighborhoods in $C_2$. The we enumerate in $O(n^k)$ all the possible subsets of $V(G_1)$ that can match exactly with the matched elements of the vertex cover $C_2$ of $G_2$. For each (of the $O(n^k)$) candidate subset $C_1$ of $V(G_1)$, we enumerate all $O(k^k)$ matchings with $C_2$, then we group the vertices of $V(G_1) \setminus C_1$ into $O(2^k)$ classes, then we exhaustively enumerate all the possible matches between the the complements of $C_1$ and $C_2$. This algorithm works for MCCIS, being based on exhaustive enumeration of all possible MCIS solutions.

In the following table we give a summary of some results obtained in this paper along with open questions. Note that $vc + fvs$ is not the same as $fvs + vc$. In the latter, the parameter is a bound on the vertex cover of $G_2$ (as well as the feedback vertex set of $G_1$) which makes ISI in FPT, while it remains open for $vc + fvs$. We also note that ISI is not in XP w.r.t. $vc(G_1)$ by a simple reduction from Independent Set (let $G_2$ be an edgeless graph on $k$ vertices, then its vertex cover number is 0).

|  | $vc + vc$ | $vc + fvs$ | $fvs + fvs$ | $vc$ |
|---|---|---|---|---|
| ISI | FPT; no Poly Kernel (Th. 2) | Open | $\notin$ XP | FPT for $vc(G_2)$, $\notin$ XP for $vc(G_1)$ |
| M(C)CIS | FPT ([1],Th. 5); no Poly Kernel | Open but $\in$ XP | $\notin$ XP (Cor. 3) | $\notin$ XP |

**Table 1.** Summary of different parameterized complexity results of ISI, MCIS and MCCIS for different structural parameters.

## 5 Conclusion

We studied the Maximum Common Induced Subgraph and Maximum Common Connected Induced Subgraph problems with respect to the solution size as natural parameter on special graphs classes, such as forests, bipartite graphs and graphs of girth five. The two problems are fixed-parameter tractable on $H$-minor free graphs, which include forests, but they are W[1]-complete on bipartite graphs and graphs of girth five.

We also considered the use of auxiliary parameters, such as a bound on the minimum vertex covers of the input graphs. Although both MCIS and MCCIS are in FPT in this case, we proved that no kernel of polynomial bound can be obtained unless NP $\subseteq$ coNP/*poly*. We noted that MCIS is not even in XP with

respect to other (smaller) auxiliary parameters, such as treewidth and feedback vertex set (see Table 1). A few corresponding open problems remain to be addressed. For example, are MCIS/MCCIS in FPT when parameterized by $vc + fvs$ or $vc + tw$? We know the problems are in XP in both cases, but we conjecture that they are both W[1]-hard.

# References

1. F. N. Abu-Khzam. Maximum common induced subgraph parameterized by vertex cover. *Inf. Process. Lett.*, 114(3):99–103, 2014.
2. T. Akutsu. A polynomial time algorithm for finding a largest common subgraph of almost trees of bounded degree. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 76(9):1488–1493, 1993.
3. H. L. Bodlaender, B. M. P. Jansen, and S. Kratsch. Cross-Composition: A New Technique for Kernelization Lower Bounds. In *STACS*, volume 9 of *LIPIcs*, pages 165–176, 2011.
4. M. Cesati. The turing way to parameterized complexity. *J. Comput. Syst. Sci.*, 67(4):654–685, 2003.
5. J. Chen, I. A. Kanj, and G. Xia. Improved upper bounds for vertex cover. *Theor. Comput. Sci.*, 411(40-42):3736–3756, Sept. 2010.
6. R. G. Downey and M. R. Fellows. *Fundamentals of Parameterized Complexity*. Texts in Computer Science. Springer, 2013.
7. M. R. Fellows, B. M. P. Jansen, and F. A. Rosamond. Towards fully multivariate algorithmics: Parameter ecology and the deconstruction of computational complexity. *Eur. J. Comb.*, 34(3):541–566, 2013.
8. J. Flum and M. Grohe. Fixed-parameter tractability, definability, and model-checking. *SIAM Journal on Computing*, 31(1):113–145, 2001.
9. M. R. Garey and D. S. Johnson. *Computers and Intractability: A guide to the theory of NP-completeness*. W.H. Freeman, San Francisco, 1979.
10. H. M. Grindley, P. J. Artymiuk, D. W. Rice, and P. Willett. Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *Journal of molecular biology*, 229(3):707–721, 1993.
11. I. Koch, T. Lengauer, and E. Wanke. An algorithm for finding maximal common subtopologies in a set of protein structures. *J. of comp. biology*, 3(2):289–306, 1996.
12. J. McGregor and P. Willett. Use of a maximal common subgraph algorithm in the automatic identification of the ostensible bond changes occurring in chemical reactions. *J. of Chemical Information and Computer Science*, 21:137–140, 1981.
13. H. Moser and S. Sikdar. The parameterized complexity of the induced matching problem. *Discrete Applied Mathematics*, 157(4):715–727, 2009.
14. R. Niedermeier. *Invitation to Fixed Parameter Algorithms*. Lecture Series in Mathematics and Its Applications. Oxford University Press, 2006.
15. J. W. Raymond and P. Willett. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *Journal of Computer-Aided Molecular Design*, 16:521–533, 2002.
16. A. Yamaguchi, K. F. Aoki, and H. Mamitsuka. Finding the maximum common subgraph of a partial $k$-tree and a graph with a polynomially bounded number of spanning trees. *Inf. Process. Lett.*, 92(2):57–63, 2004.