Facets for the Maximum Common Induced Subgraph Problem Polytope

Breno Piva, Cid de Souza — {bpiva, cid}@ic.unicamp.br

2 de setembro de 2011

Abstract: This paper presents some strong valid inequalities for the Maximum Common Induced Subgraph Problem (**MCIS**) and the proofs that the inequalities are facet-defining under certain conditions. The **MCIS** is an $\mathcal{NP}$-hard problem and, therefore, no polynomial time algorithm is known to solve it. In this context, the study of its polytope can help in the development of more efficient branch-and-bound and branch-and-cut algorithms.

Keywords: Combinatorial Optimization, Integer Programming, Polyhedron, Maximum Common Induced Subgraph

1 Introduction

Using graphs it is possible to model a big number of real world problems and a significant number of those is related to graph isomorphism. Particularly, the Maximum Common Induced Subgraph (**MCIS**) have several applications in areas like Computer Vision [6, 13], Video indexing [17], Chemistry and Biology [20, 5, 1]. This problem can be defined like follows:

Input:	two graphs G and H.
Goal:	find a induced subgraph G' in G isomorphic to an induced subgraph H' in H with a maximum number of vertices.

Here we should recall the concepts of **isomorphism** and of **induced subgraph**. Two graphs $G = (V, E)$ and $H = (W, F)$ with $|V| = |W|$ are called isomorphic if there is a bijection $\pi : V \to W$ such that $(i, j) \in E \Rightarrow (\pi(i), \pi(j)) \in F$ and $(i, j) \in F \Rightarrow (\pi^{-1}(i), \pi^{-1}(j)) \in E$. Besides, given a graph $G = (V, E)$, the graph $H = (W, F)$ is an induced subgraph of G if $W \subseteq V$ and $F = \{(i, j) \in E : i \in W \text{ and } j \in W\}$, i.e., F is the edge set with both extremities in W. In that case, we also say that H is the subgraph induced by W in G.

Another manner 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 observed: 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 (the subgraph is induced in both graphs).

The **MCIS** is a $\mathcal{NP}$-hard problem and can be found in the classic list in [4]. Therefore, unless $\mathcal{P}$=$\mathcal{NP}$, there are no polynomial algorithms to solve it. For this reason, several heuristics and approximation algorithms were proposed for the **MCIS** [7, 11, 16, 15, 14, 18, 19]. However, despite its complexity, it is still important to know exact solutions for instances, even if small ones, of the problem. The most used methods in the literature to solve this problem exactly are based in backtracking algorithms as in [10], [21] and [8], or in the reduction to another more studied $\mathcal{NP}$-hard problem, such as the Clique problem, as done in [3] and in two of the algorithms presented in [2].

In this work we present some strong valid inequalities for the **MCIS** polytope and the respective proofs that these inequalities are facet-defining under certain conditions. Such inequalities could then be used in branch-and-bound or branch-and-cut algorithms for the problem.

The rest of the paper is organized as follows. The next section presents a basic model for the **MCIS**. Section 3 presents some strong valid inequalities and the proofs that they are facet-defining and at last, Section 4 discuss some conclusions.

2 Integer Programming Model for the MCIS

In this section we present a very simple **IP** formulation for the **MCIS** problem. This model is based on the one given in [9]:

$$\max \quad \textstyle\sum_{i\in V(G)} \sum_{j\in V(H)} y_{ij} \tag{1}$$
$$\textstyle\sum_{i\in V(G)} y_{ij} \leq 1, \qquad \forall j \in V(H), \tag{2}$$
$$\textstyle\sum_{j\in V(H)} y_{ij} \leq 1, \qquad \forall i \in V(G), \tag{3}$$
$$y_{ij} + y_{kj} + y_{il} + y_{kl} \leq 1, \qquad \forall (i,k) \in E(G), \forall (l,j) \notin E(H), \tag{4}$$
$$y_{ij} + y_{kj} + y_{il} + y_{kl} \leq 1, \qquad \forall (i,k) \notin E(G), \forall (l,j) \in E(H), \tag{5}$$
$$y_{ij} \in \{0,1\} \qquad \forall i \in V(G), \forall j \in V(H). \tag{6}$$

In this formulation, the variables y_{ij}, for all $i \in V(G)$ and $j \in V(H)$, are set to 1 (one) if, and only if, the vertex i is mapped to the vertex j and set to 0 (zero) otherwise.

Constraints (7) and (3) guarantee, respectively, that every vertex in $V(H)$ is mapped to at most one vertex in $V(G)$ and that every vertex in $V(G)$ is mapped to, at most, a single vertex in $V(H)$. Constraints (4) and (5) though, guarantee, respectively, that two adjacent vertices in $V(G)$ are not mapped at the same time to two non-adjacent vertices of $V(H)$ and that two adjacent vertices in $V(H)$ are not mapped simultaneously to a pair of non-adjacent vertices in $V(G)$. It is easy a simple task to show that this formulation is correct.

3 Strong Valid Inequalities for the MCIS Polytope

In the present section we present proofs that some of the inequalities shown in Section 2 are facet-defining under certain conditions. We also present new inequalities and show that they are facet-defining.

The first thing we must observe is that the polytope is full-dimensional. We can prove this by showing that the vectors from the identity matrix plus the null vector are in the polytope and, therefore, there are $n + 1$ afine independent vectors (where n is the number of variables in the model).

3.1 Unique Mapping

We divide the valid inequalities for the **MCIS** in two categories. In the first category are the inequalities that guarantee that we can not map a vertex from one graph to more than a vertex in the other, e.g., (7) and (3). In the second one are those inequalities that determine that a pair of non-adjacent vertices from one graph can not be mapped to a pair of adjacent vertices in the other graph. This subsection deals with the first category while Subsection 3.2 deals with the second.

Throughout the text the reader should consider that whenever we mention a family of inequalities that refer to mappings of certain set types from G to H, there is always an equivalent family with mappings from H to G. Examples of this fact are (7) and (3) and also (4) and (5).

Now, we prove that the inequalities in (7) used in the model are facet-defining under certain conditions.

$$\sum_{i \in V(G)} y_{ij} \leq 1, \qquad \forall j \in V(H), \tag{7}$$

$$\tag{8}$$

If there is a universal vertex k in G, then, j must also be universal. (9)

If there is an isolated vertex k in G, then, j must also be isolated. (10)

Proposition 3.1 *If constraint* (7) *defines a facet* $F = \{x \in conv(P) : \sum_{i \in V(G)} y_{ij} = 1\}$*, in* $conv(P)$*, then, conditions* (9) *and* (10) *must be satisfied.*

Proof:

Assume that (7) is facet-defining and (9) is not satisfied, then, for, $k \in V(G)$, $j, l \in V(H)$, $(j, l) \notin E(H)$ and k universal, there is no $i \in V(G)$ such that $(i, k) \notin E(G)$, therefore, a lifting could be performed and (7) is not facet-defining, contradiction. Now, assume that (7) is facet-defining and (10) is not satisfied, then, for, $k \in V(G)$, $j, l \in V(H)$, $(j, l) \in E(H)$ and k isolated, there is no $i \in V(G)$ such that $(i, k) \in E(G)$, therefore, a lifting could be performed and (7) is not facet-defining, contradiction.

Ergo, if ($\neg$(9) or $\neg$(10)) then, the constraint (7) is not facet-defining for $conv(P)$. □

If conditions (9) and (10) are not satisfied and there is exactly one isolated or universal vertex, then the inequalities in the following family are valid and facet-defining as we shall see.

$$\sum_{i \in V(G)} y_{ij} + \sum_{l \in Adj(j)} y_{kl} + \sum_{l \notin Adj(j)} y_{pl} \leq 1, \quad \forall\ j \in V(H) \tag{11}$$

where k is the isolated vertex in G and p is the universal vertex in G (of course, only one of these can happen at once).

Just as we did for the other constraints, before proving that the inequality is facet-defining, we must first prove that it is valid for the **MCIS** polyhedron. So, we have:

Proposition 3.2 *The inequality* (11) *is valid for the* **MCIS** *polyhedron.*

Proof: For this proof, let us consider that there is an isolated vertex k in G. The case where there is an universal vertex in G is similar. Assume by contradiction that (11) is not valid for the **MCIS** polyhedron. Hence, there must exist some point in the polyhedron for which at least two of the variables in the **LHS** have value 1 at the same time. For this to happen there are three possible cases:

Case (1): Two variables have value 1 in the first sum. If this happens, it would mean that more than one vertex of G is been mapped to j. However, from the definition of the **MCIS** that is impossible.

Case (2): Two variables have value 1 in the second sum. That would mean that a single vertex, k, in G is been mapped to more than one vertex in H, which is forbidden by the definition of the **MCIS**.

Case (3): A variable from the first sum and another from the second sum have value 1 at the same time. That would mean that either one vertex of G is been mapped to two vertices in H (if $i = k$), or that a pair of non-adjacent vertices in G (k is isolated) is been mapped to a pair of adjacent vertices in H (l is in $Adj(j)$). However, from **MCIS**'s definition, that is also not possible.

Therefore, there is no point in the **MCIS** polyhedron for which the inequality is not satisfied, contradicting the hypothesis. So, (11) is valid for that polyhedron. □

Following we show a demonstration that (11) is facet-defining for $conv(P)$. Notice that for this proof we do not define any necessary condition for it to be true. This is so because from the definition of the constraint itself, the isolated and universal vertices sets are unitary and, under these conditions, the inequality is always facet-defining.

Proposition 3.3 *Constraint* (11) *defines a facet* $F = \{x \in conv(P) : \sum_{i \in V(G)} y_{ij} + \sum_{l \in Adj(j)} y_{kl} = 1\}$ *of* $conv(P)$.

Proof: The face defined by (11) is proper because the null vector is not in the face and is in $conv(P)$, also, any vector with 0 at every position except at position $ij : i \in V(G)$ is in the face. Let $F' = \{x \in conv(P) : \pi x = \pi_0\}$ e $F \subseteq F'$. Now, if (11) is facet-defining for $conv(P)$, then, there must be a positive constant $\alpha \in \mathbb{R}$ satisfying:

$(\pi, \pi_0) = \alpha(\gamma, 1)$ where $\gamma_{ij'} = 1$ if $j' = j\ \forall\ i \in V(G)$ or ($i = k$ and $j' \in Adj(j)$) and $\gamma_{ij'} = 0$, otherwise. That is, we must show that $\pi_0 = \alpha$ and $\pi_{ij'} = \alpha = \pi_0$ if $j' = j$ or ($i = k$ and $j' \in Adj(j)$) and $\pi_{ij'} = 0$, otherwise.

If the inequality (11) is facet-defining, than, it must be possible to find the same relations using the fact that $F \subseteq F'$. So we have:

(1) Clearly the vectors $(\ldots, 0, 1_{ij'}, 0, \ldots)$, where ($i \in V(G)$ and $j' = j$) or ($i = k$ and $j' \in Adj(j)$) are in F. As $F \subseteq F' \Rightarrow (\ldots, 0, 1_{ij'}, 0, \ldots) \in F' \Rightarrow \pi_{ij'} = \pi_0$.

(2) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{k'l},0,\ldots) \in F$, where $k' \in V(G)$, $i \in V(G)\setminus k$, $l \in V(H)\setminus j$, then, if $k' \neq k$, we can choose i so that if $(l,j) \in E(H)$, then, $(i,k') \in E(G)$ (which is always possible because k' is not isolated) and if $(l,j) \notin E(H)$, then, $(i,k') \notin E(G)$ (which is also always possible because k' is not universal). As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{k'l},0,\ldots) \in F'$ $\Rightarrow \pi_{ij} + \pi_{k'l} = \pi_0$. Therefore, from (1), we have $\pi_{k'l} = 0\ \forall\ k' \in V(G),\ k' \neq k,\ l \in V(H) \setminus j$.

(3) The vectors $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots)$, where $i \in V(G)\setminus k$, $l \in V(H)\setminus(j \cup Adj(j))$ are in F, since $(i,k) \notin E(G)$ and $(j,l) \notin E(H)$. As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij'},\ldots,0,\ldots,1_{kl},0,\ldots) \in F'$ $\Rightarrow \pi_{ij'} + \pi_{kl} = \pi_0$. Therefore,from (1), we have $\pi_{kl} = 0\ \forall\ l \in V(H) \setminus (j \cup Adj(j))$.

□

3.2 Edge Mapping

The family of inequalities (4) can be easily generalized to the following family of stronger inequalities. The inequalities in this new family are facet-defining under certain conditions as proven bellow.

$$\textstyle\sum_{i\in C}\sum_{j\in I} y_{ij} \leq 1, \qquad \forall \text{ maximal Clique } C \in G, \forall \text{ maximal IS } I \in H \tag{12}$$

$$\begin{gathered}\text{For every vertex } k \text{ in } G \text{ not in } C\\ \text{there is some vertex in } C \text{ adjacent to } k;\end{gathered} \tag{13}$$

$$\begin{gathered}\text{For every vertex } l \text{ in } H \text{ not in } I\\ \text{there is some vertex in } I \text{ non-adjacent to } l.\end{gathered} \tag{14}$$

Proposition 3.4 *Constraint* (12) *defines a facet* $F = \{x \in conv(P) : \sum_{i\in C}\sum_{j\in I} y_{ij} = 1\}$ *for* $conv(P)$ *is, and only if, either condition* (13) *or condition* (14) *is true.*

Proof: $(\Leftarrow)$
The face defined by (12) is proper because the null vector is not in the face and is in $conv(P)$, and any vector with 0 at every position except at position $ij : i \in C$ and $j \in I$ (notice that C and I are never empty unless G and H themselves are empty) is in the face. Let $F' = \{x \in conv(P) : \pi x = \pi_0\}$ and $F \subseteq F'$. Now, if (12) is facet-defining for$conv(P)$, then, there must exist some positive constant $\alpha \in \mathbb{R}$ satisfying:

$(\pi, \pi_0) = \alpha(\gamma, 1)$ where $\gamma_{ij} = 1$ if $i \in C$ e $j \in I$ and $\gamma_{ij} = 0$, otherwise. That is, we must prove that $\pi_0 = \alpha$, $\pi_{ij} = \alpha = \pi_0$ if $i \in C$ and $j \in I$ and $\Pi_{ij} = 0$, otherwise.

If (12) is facet-defining, then it must be possible to find the same relations as above by using the fact that $F \subseteq F'$. So, we have:

(1) The vectors $(\ldots,0,1_{ij},0,\ldots) \in F$, where $i \in C$ and $j \in I$ are clearly in F. As $F \subseteq F'$ $\Rightarrow (\ldots,0,1_{ij},0,\ldots) \in F' \Rightarrow \pi_{ij} = \pi_0$.

(2) Let $(\dots,0,1_{ij},\dots,0,\dots,1_{kl},0,\dots) \in F$, where $i \in C$, $k \in C$, $j \in I$, $l \notin I$. We can choose j so that $(j,l) \in E(H)$, which is always possible since I is maximal. As $F \subseteq F'$ $\Rightarrow (\dots,0,1_{ij},\dots,0,\dots,1_{kl},0,\dots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall\ k \in C$ and $l \notin I$.

(3) Let $(\dots,0,1_{ij},\dots,0,\dots,1_{kl},0,\dots) \in F$, where $i \in C$, $k \notin C$, $j \in I$ and $l \in I$. We can choose i so that $(i,k) \notin E(G)$, which is always possible because C is maximal. As $F \subseteq F'$ $\Rightarrow (\dots,0,1_{ij},\dots,0,\dots,1_{kl},0,\dots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall\ k \notin C$ and $l \in I$.

(4) Let $(\dots,0,1_{ij},\dots,0,\dots,1_{kl},0,\dots) \in F$, where $i \in C$, $k \notin C$, $j \in I$, $l \notin I$. We can choose i and j so that either $(i,k) \in E(G)$ and $(j,l) \in E(H)$ (which is alway possible if (13) is satisfied, because I is maximal) or, $(i,k) \notin E(G)$ and $(j,l) \notin E(H)$ (which is always possible if (14) is satisfied, because C is maximal). As $F \subseteq F' \Rightarrow (\dots,0,1_{ij},\dots,0,\dots,1_{kl},0,\dots) \in F'$ $\Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall\ k \notin C$ and $l \notin I$.

$(\Rightarrow)$

If (13) and (14) are not satisfied, then in (4) for some k and l it would not be possible to chose an i such that $(i,k) \in E(G)$ and $(j,l) \in E(H)$ or $(i,k) \notin E(G)$ and $(j,l) \notin E(H)$. Therefore, there is no vector in F having value 1 at position kl which is equivalent to say that in the polyhedron defined by F there is an equality $y_{kl} = 0$ and, so, its dimension is smaller than $|V(G)| \times |V(H)| - 1$, hence, it can not be a facet of $conv(P)$.

Thus, if $\neg$(13) and $\neg$(14) the inequality (12) is not facet-defining for $conv(P)$. $\square$

Using a lifting procedure [12] it is possible to strengthen (12) and find the following family of strong valid inequalities, which are also facet-defining under certain conditions.

$$\sum_{h=0\ :\ h \notin mci(G)\ \wedge\ h \notin mii(H)}^{min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1} \sum_{i \in C(G_h)} \sum_{j \in I(H_h)} y_{ij} + \sum_{k \in MC(G)} \sum_{l \in MI(H)} y_{kl} + \sum_{k \in MC(G)} \sum_{l \in Cl(R^H)} y_{kl} +$$
$$+ \sum_{k \in Is(R^G)} \sum_{l \in MI(H)} y_{kl} + \sum_{h \in mci(G)\ :\ \{k\}=C(G_h)} \sum_{l \in F(H_h)} y_{kl} + \sum_{h \in mii(H)\ :\ \{l\}=I(H_h)} \sum_{k \in F(G_h)} y_{kl} \leq 1, \quad (15)$$
$$\forall\ \mathcal{SC}(G), \forall\ \mathcal{SI}(H),$$

Proposition 3.5 *The inequality* (15) *is valid for the* **MCIS** *polyhedron.*

Proof:

First of all, we must notice that the third and forth terms of the sum are mutually exclusive ($R^H \neq \emptyset \Rightarrow R^G = \emptyset$ and vice-versa). Let us consider that the number of ISs in the sequence is bigger than the number of cliques, hence, the forth term of the sum is empty. Now, assume by contradiction that (15) is not valid for the **MCIS** polyhedron. So, there must be some point in that polyhedron that does not satisfy the inequality. We are not going to consider the cases in which some unique mapping restriction could be violated:

Case (1): Two variables have value 1 with different values for h in the first sum. If this happens, then, a pair of non-adjacent vertices from G (by construction, the cliques in the sequence are independent from each other) would map to a pair of adjacent vertices in H (by construction, the ISs are totally connected to each other). However, such situation can not happen in the **MCIS**.

Case (2): Two variables have value 1 with different values for h in the fifth or sixth sum. If this happens, just as in the first case, then a pair of non-adjacent vertices in G (the cliques with size 1 are non-adjacent and so are the vertices in two sets $F(G_r)$ and $F(G_s)$ with $r \neq s$) would map to a pair of adjacent vertices from H (the ISs of size 1 are adjacent, as are the vertices in two sets $F(H_r)$ and $F(H_s)$ with $r \neq s$). Therefore, this situation can not occur for it violates the **MCIS** definition.

Case (3): Two variables have value 1 in any of the other sums. The set $MC(G)$ is an IS and the set $MI(H)$ is a clique and in all the sums we make mappings of these sets, respectively, to sets that are totally connected and sets that are totally disconnected, therefore, a point in the **MCIS** polyhedron in the specified conditions would have an edge mapped to a non-edge, but from the **MCIS** definition, this point is not in the polyhedron, a contradiction.

Case (4): More than a variable have value 1 in different sums. Any two variables belonging to different sums correspond to non-adjacent vertices from G and adjacent vertices from H [1], therefore, any point in the specified conditions maps an edge to a non-edge and, thus, it is not in the **MCIS** polyhedron.

Therefore, there is no point in the **MCIS** polyhedron that does not satisfy the inequality, thus, it is valid. □

Now, consider the following conditions:

$$|MC(G)| > 1 \tag{16}$$

$$|MI(H)| > 1 \tag{17}$$

Proposition 3.6 *If conditions* (16) *and* (17) *are satisfied, then,* (15) *defines a facet* $F = \{x \in conv(P) : \sum_{h=0 \,:\, h \notin mci(G) \,\wedge\, h \notin mii(H)}^{min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1} \sum_{i \in C(G_h)} \sum_{j \in I(H_h)} y_{ij} + \sum_{k \in MC(G)} \sum_{l \in MI(H)} y_{kl} + \sum_{k \in MC(G)} \sum_{l \in Cl(R^H)} y_{kl} + \sum_{k \in Is(R^G)} \sum_{l \in MI(H)} y_{kl} + \sum_{h \in mci(G) \,:\, \{k\}=C(G_h)} \sum_{l \in F(H_h)} y_{kl} + \sum_{h \in mii(H) \,:\, \{l\}=I(H_h)} \sum_{k \in F(G_h)} y_{kl} = 1\}$ *for* $conv(P)$.

Proof: In this proof we consider that $|\mathcal{SC}(G)| < |\mathcal{SI}(H)|$ (In this case the forth sum disappears). For the other case the proof is similar. The face defined by (15) is proper since the null vector is not in the face and is in $conv(P)$, also, any vector having 0 at every position except at position $ij : i \in C(G_h)$ and $j \in I(H_h)$, for some $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$ is in the face, therefore, the face is proper.

[1] Except, possibly, variables corresponding to vertices in some $F(G_r)$ and $C(G_r)$ which can be adjacent. However, these variables only have coefficient 1 when $|I(H_r)| = 1$ and, in this case, if both variables have value 1 simultaneously, there would be two vertices from G mapping to a single vertex in H. A similar situation can occur with some $F(H_r)$ and $I(H_r)$, but, in this case, the vertices would be non-adjacent.

Let $F' = \{x \in conv(P) : \pi x = \pi_0\}$ and $F \subseteq F'$. Now, if the inequality (15) is facet-defining for $conv(P)$, then, there must exist a positive constant $\alpha \in \mathbb{R}$ satisfying:

$(\pi, \pi_0) = \alpha(\gamma, 1)$ where $\gamma_{ij} = 1$ is the variable y_{ij} have coefficient 1 in F and $\gamma_{ij} = 0$, otherwise. That is, we must show that $\pi_0 = \alpha$ and $\pi_{ij} = \alpha = \pi_0$ is the variable y_{ij} have coefficient 1 in F and $\pi_{ij} = 0$, otherwise.

To facilitate the understanding of the proof, Figure 1 and Figure 2 presents diagrams with the types of mappings contemplated in each case of the proof. So, e.g., a line linking clique $C(G_h)$ to IS $I(H_h)$ represents a mapping of a vertex from that clique to a vertex in that IS. Also, the different types of lines used represent the different cases of the proof according to the legend in the figures. For this proof, the figure has been divided in two to improve visualization and show all the cases. Figure 1 shows the cases of mappings of vertices of a clique having more than one vertex to sets of the sequence in the other graph, whilst Figure 2 shows cases of mappings to cliques with a single vertex.

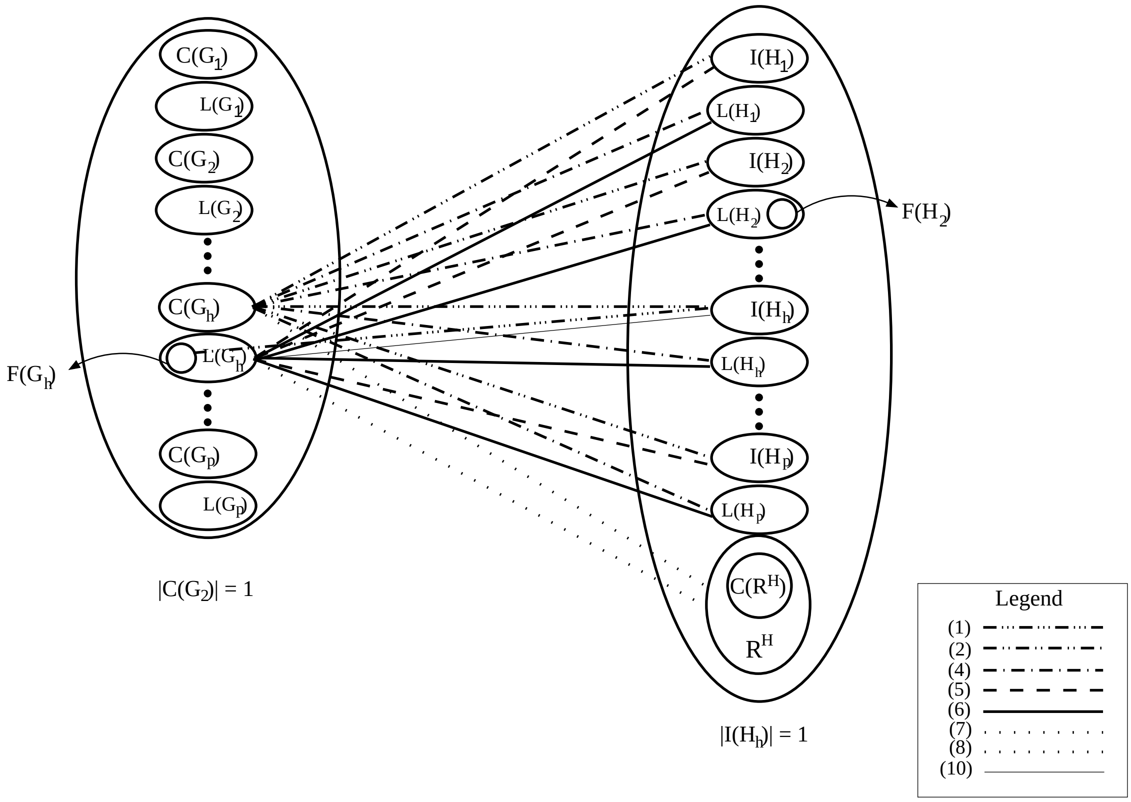

Figura 1: Partial diagram of Proposition 3.6 proof.

If (15) defines a facet, then, it should be possible to find the same relations as above using the fact that $F \subseteq F'$, so, we have:

(1) Clearly the vectors $(\ldots, 0, 1_{ij}, 0, \ldots)$, where ij corresponds to some position in which the variable y_{ij} have coefficient 1 in F are in F. As $F \subseteq F' \Rightarrow (\ldots, 0, 1_{ij}, 0, \ldots) \in F' \Rightarrow \pi_{ij} = \pi_0$.

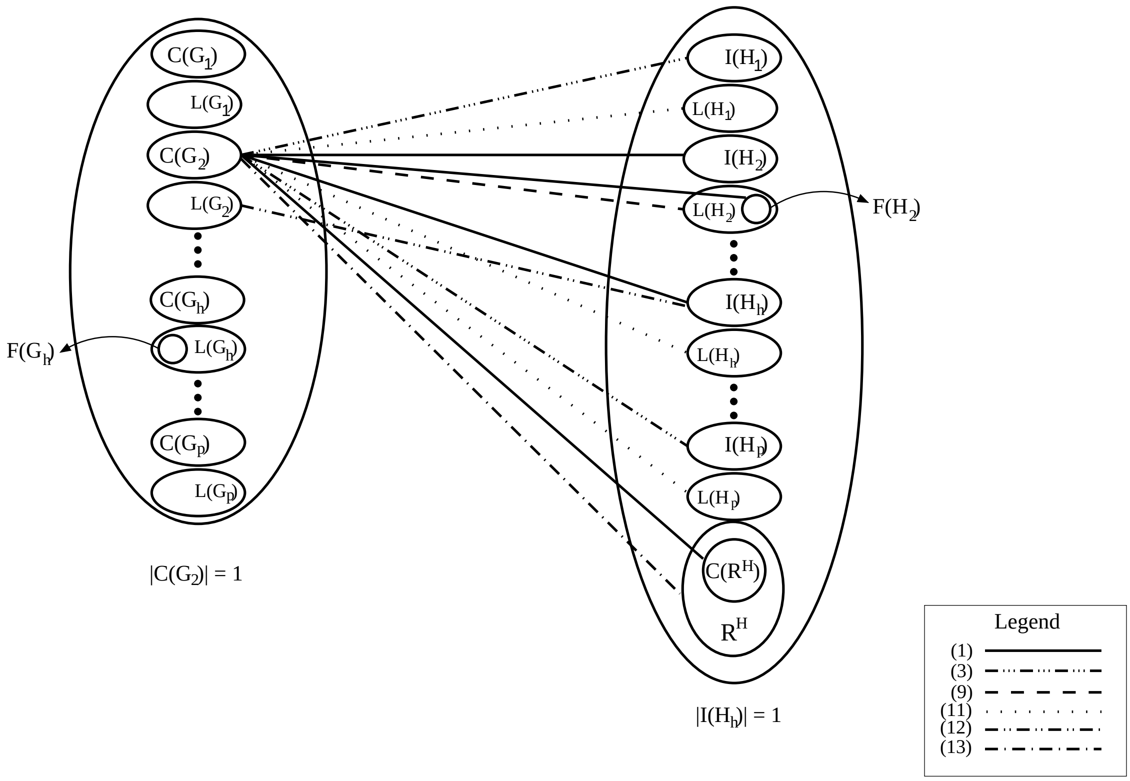

Figura 2: Partial diagram of Proposition 3.6 proof.

(2) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in C(G_h)$, $k \notin MC(G)$, $l \in I(H_r)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$ and $h \neq r$. We can choose i in $C(G_h)\backslash\{k\}$ and j in $I(H_h)$ and, therefore, $(i,k) \in E(G)$ and $(j,l) \in E(H)$. As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F'$ $\Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0\ \forall\ k \in C(G_h)$, $k \notin MC(G)$, $l \in I(H_r)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$, $h \neq r$.

(3) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in C(G_h)$, $l \in I(H_r)$, $l \notin MI(H)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$ and $h \neq r$. We can choose i in $C(G_r)$ and j in $I(H_r)\backslash\{l\}$ and, therefore, $(i,k) \notin E(G)$ and $(j,l) \notin E(H)$. As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F' \Rightarrow \pi_{ij}+\pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0\ \forall\ k \in C(G_h)$, $l \in I(H_r)$, $l \notin MI(H)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$, $h \neq r$.

(4) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in C(G_h)$, $k \notin MC(G)$, $l \in L(H_r)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$, then, if $r \geq h$, we can choose j in $I(H_h)$ so that $(j,l) \in E(H)$ (which is always possible since $I(H_h)$ is maximal) and we can choose i in $C(G_h) \setminus \{k\}$ and, therefore, $(i,k) \in E(G)$ (because both vertices are in the same clique). If $r < h$, then we can choose j in $I(H_r)$ so that $(j,l) \notin E(H)$ (which is always possible because, from the definition of $L(H_r)$, such j must exist) and we can choose i in $C(G_r)$ and, hence, $(i,k) \notin E(G)$. As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0\ \forall\ k \in C(G_h)$, $k \notin MC(G)$, $l \in L(H_r)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$.

(5) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in L(G_r)$, $l \in I(H_h)$, $l \notin MI(H)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$, then, if $r \geq h$, we can choose i in $C(G_h)$ so that $(i,k) \notin E(G)$ (which is always possible since $C(G_h)$ is maximal) and we can choose j in $I(H_h) \setminus \{l\}$ and, therefore, $(j,l) \notin E(H)$ (because both vertices are in the same IS). If $r < h$, then we can choose i in $C(G_r)$ so that $(i,k) \in E(G)$ (which is always possible because, by the definition of $L(G_r)$, such i must exists) and we can choose j in $I(H_r)$ and, hence, $(j,l) \in E(H)$. As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0\ \forall\ k \in L(G_r)$, $l \in I(H_h)$, $l \notin MI(H)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$.

(6) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in L(G_h)$, $l \in L(H_r)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$. If $h \leq r$, we can choose i in $C(G_h)$ so that $(i,k) \in E(G)$ (which is always possible because, from the definition of $L(G_h)$, such i always exists) and we can choose j in $I(H_h)$ so that $(j,l) \in E(H)$ (which is always possible since $I(H_h)$ is maximal). If $h > r$, then we can choose j in $I(H_r)$ so that $(j,l) \notin E(H)$ (which is always possible because, by the definition of $L(H_r)$, such j always exists), and we can choose i in $C(G_r)$ so that $(i,k) \notin E(G)$ (which is possible since $C(G_r)$ is maximal). As $F \subseteq F'$ $\Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0\ \forall\ k \in L(G_h)$, $l \in L(H_r)$, $h,r = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$.

(7) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in C(G_h)$, $k \notin MC(G)$, $h = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$, $l \in R^H$, then, we can choose i in $C(G_h) \setminus \{k\}$ and, therefore, $(i,k) \in E(G)$ and we can choose j in $I(H_h)$ and, hence, $(j,l) \in E(H)$ (otherwise, l would be in $I(H_h)$ or in $L(H_h)$). As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F'$ $\Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0\ \forall\ k \in C(G_h)$, $k \notin MC(G)$, $h = 0\ldots min(|\mathcal{SC}(G)|,|\mathcal{SI}(H)|)-1$, $l \in C(R^H)$.

(8) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in L(G_h)$, $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, $l \in R^H$, then we can choose i in $C(G_h)$ so that $(i,k) \in E(G)$ and we can choose j in $I(H_h)$ and, therefore, $(j,l) \in E(H)$ (otherwise, l would be in $I(H_h)$ or in $L(H_h)$). As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall \ k \in L(G_h)$, $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, $l \in R^H$.

(9) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in C(G_h)$, $k \in MC(G)$, $l \in L(H_h) \setminus F(H_h)$, $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, then we can choose j in $I(H_s)$, $s > h$, or in $F(H_t)$, $t \in mci(G)$ and $t < h$ so that $(j,l) \notin E(H)$ and we can choose i in $C(G_s)$ or in $C(G_t)$ (depending on the j chosen) and, hence, $(i,k) \notin E(G)$. As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F'$ $\Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall \ k \in C(G_h)$, $k \in MC(G)$, $l \in L(H_h)$, $l \notin F(H_h)$, $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$.

(10) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in L(G_h \setminus F(G_h))$, $l \in I(H_h)$, $l \in MI(H)$, $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, then we can choose i in $C(G_s)$, $s > h$, or in $F(H_t)$, $t \in mii(H)$ and $t < h$ so that $(i,k) \in E(G)$ and we can choose j in $I(H_s)$ or in $I(H_t)$ (depending on the i chosen) and, therefore, $(j,l) \in E(H)$. As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F'$ $\Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall \ k \in L(G_h)$, $k \notin F(G_h)$, $l \in I(H_h)$, $l \in MI(H)$, $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$.

(11) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in C(G_h)$, $k \in MC(G)$, $l \in L(H_r)$, $h,r = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, $h \neq r$, then we can choose j in $I(H_r)$ so that $(j,l) \notin E(H)$ (which is always possible because, from the definition of $L(H_r)$, there is always such j), and we can choose i in $C(G_r)$ and, therefore, $(i,k) \notin E(G)$. As $F \subseteq F' \Rightarrow$ $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall \ k \in C(G_h)$, $k \in MC(G)$, $l \in L(H_r)$, $h,r = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, $h \neq r$.

(12) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in L(G_r)$, $l \in I(H_h)$, $l \in MI(H)$, $h,r = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, $h \neq r$, then we can choose i in $C(G_r)$ so that $(i,k) \in E(G)$ (which is always possible because, from the definition of $L(G_r)$, such i exists), and we can choose j in $I(H_r)$ and, therefore, $(j,l) \in E(H)$. As $F \subseteq F' \Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall \ k \in L(G_r)$, $l \in I(H_h)$, $l \in MI(H)$, $h,r = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, $h \neq r$.

(13) Let $(\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F$, where $k \in C(G_h)$, $k \in MC(G)$, $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, $l \in R^H \setminus Cl(R^H)$, then we can choose j in $Cl(R^H)$ or $F(H_r)$ so that $(j,l) \notin E(H)$ and we can choose i in $C(G_r)$, $r \in mci(G)$ (which is always possible because of condition (16)) and, therefore, $(i,k) \notin E(G)$. As $F \subseteq F'$ $\Rightarrow (\ldots,0,1_{ij},\ldots,0,\ldots,1_{kl},0,\ldots) \in F' \Rightarrow \pi_{ij} + \pi_{kl} = \pi_0$. Therefore, from (1), we have $\pi_{kl} = 0 \ \forall \ k \in C(G_h)$, $k \in MC(G)$, $h = 0 \ldots min(|\mathcal{SC}(G)|, |\mathcal{SI}(H)|) - 1$, $l \in R^H \setminus Cl(R^H)$.

□

Whenever $|MC(G)| = 1, |MI(H)| = 1, |MI(G)| = 1$ or $|MC(H)| = 1$, the inequalities (15) are modified so the third and forth terms of the sum are adequately replaced by:

$$\sum_{l \in R^H \,:\, MC(G)=\{k\}} y_{kl} \tag{18}$$

$$\sum_{k \in R^G \,:\, MI(H)=\{l\}} y_{kl} \tag{19}$$

The proof of validity for these inequalities is identical to the one for Proposition 3.5 and the proof that these inequalities are facet-defining is very similar to the one of Proposition 3.6, except for case (13) that would cease to exist because it would be contemplated in case (1).

4 Conclusions

The inequalities and proofs presented in this paper are as far as we are concerned, novel. Understanding the **MCIS** polytope is important not only for possibly helping in the development of efficient algorithms for the **MCIS** but as well for potentially been useful in the development of algorithms for problems resulting from adding side constraints to the **MCIS**.

Referências

[1] L. Chen and W. Robien. Application of the maximal common substructure algorithm to automatic interpretation of 13C NMR spectra. *Journal of Chemical Information and Computer Sciences*, 34:934–941, 1994.

[2] D. Conte, C. Guidobaldi, and C. Sansone. A comparison of three maximum common subgraph algorithms on a large database of labeled graphs. In *Lecture Notes in Computer Science*, volume 2726, pages 130–141. Springer-Verlag, 2003.

[3] B. Falkenhainer, K. Forbus, and D. Gentner. The structure-mapping engine: algorithms and examples. *Artificial Intelligence*, 34:1–63, 1989/90.

[4] M. R. Garey and D. S. Johnson. *Computer and intractability: A guide to the theory of NP–completeness*. Freeman, San Francisco, 1979.

[5] E. Gifford, M. Johnson, D. Smith, and C. Tsai. Structure-reactivity maps as a tool for visualizing xenobiotic structure-reactivity. *Network Science*, 2:1–33, 1996.

[6] R. Horaud and T. Skordas. Stereo correspondence through feature grouping and maximal cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1168–1180, 1989.

[7] V. Kann. On the approximability of the maximum common subgraph problem. In *Lecture Notes in Computer Science*, volume 577, pages 377–388. Springer-Verlag, 1992.

[8] E. B. Krissinel and K. Henrick. Common subgraph isomorphism detection by backtracking search. *Softw. Pract. Exper.*, 34(6):591–607, 2004.

[9] G. Manić. Modelagem matemática e aplicações de problemas de otimização relativos à busca de subgrafos com estruturas comuns. FAPESP's First Scientific Report, 2007. Post-doctoral grant # 2006/01817-7 (in Portuguese, unpublished).

[10] J. J. Mcgregor. Backtrack search algorithms and the maximal common subgraph problem. *SOFTWARE-PRACT. AND EXPER.*, 12(1):23–34, 1982.

[11] B. T. Messmer and H. Bunke. Decision tree approach to graph and subgraph isomorphism detection. *Pattern Recognition*, 32(12):1979–1998, 1999.

[12] G. Nemhauser and L. Wolsey. *Integer and combinatorial optimization.* Wiley New York, 1988.

[13] M. Pelillo, K. Siddiqi, and S. W. Zucker. Matching hierarchical structures using association graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:1105–1120, 1999.

[14] J. W. Raymond, E. J. Gardiner, and P. Willett. Heuristics for similarity searching of chemical graphs using a maximum common edge subgraph algorithm. *Journal of Chemical Information and Computer Sciences*, 42(2):305–316, 2002.

[15] J. W. Raymond, E. J. Gardiner, and P. Willett. RASCAL: Calculation of graph similarity using maximum common edge subgraphs. *The Computer Journal*, 45(6):631–644, 2002.

[16] J. W. Raymond and P. Willett. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *Journal of Computer-Aided Molecular Design*, 16(7):521–533, 2002.

[17] K. Shearer, H. Bunke, and S. Venkatesh. Video indexing and similarity retrieval by largest common subgraph detection using decision trees. *Pattern Recognition*, 34:1075–1091, 2001.

[18] W. H. Suters, F. N. Abu-Khzam, Y. Zhang, C. T. Symons, N. F. Samatova, and M. A. Langston. A new approach and faster exact methods for the maximum common subgraph problem. In *Computing and combinatorics*, volume 3595 of *Lecture Notes in Computer Science*, pages 717–727. Springer-Verlag, 2005.

[19] Y. Wang and C. Maple. A novel efficient algorithm for determining maximum common subgraphs. In *International Conference on Information Visualisation (IV'05)*, pages 657–663, London, UK, July 2005. IEEE Computer Society. http://doi.ieeecomputersociety.org/10.1109/IV.2005.11.

[20] P. Willet. Matching of chemical and biological structures using subgraph and maximal common subgraph isomorphism algorithms. *IMA Vol. Math. Appl.*, 108:11–38, 1999.

[21] A. K. C. Wong and F. A. Akinniyi. An algorithm for the largest common subgraph isomorphism using the implicit net. In *Proc. Int. Conf. Systems, Man and Cybernetics*, volume 1, pages 197–201. Inst. Electrical & Electronics Engineers, 1983.