

Solving VLSI design and DNA sequencing problems using bipartization of graphs

Pierre Fouilhoux · A. Ridha Mahjoub

Received: 4 February 2009 / Published online: 29 September 2010
© Springer Science+Business Media, LLC 2010

Abstract In this paper we consider the 2-layer constrained via minimization problem and the SNP haplotype assembly problem. The former problem arises in the design of integrated and printed circuit boards, and the latter comes up in the DNA sequencing process for diploid organisms. We show that, for any maximum junction degree, the problem can be reduced to the maximum bipartite induced subgraph problem. Moreover we show that the SNP haplotype assembly problem can also be reduced to the maximum bipartite induced subgraph problem for the so-called minimum error correction criterion. We give a partial characterization of the bipartite induced subgraph polytope. Using this, we devise a branch-and-cut algorithm and report some experimental results. This algorithm has been used to solve real and large instances.

Keywords Electronics · Genomics · Bipartite induced subgraph problem · Polyhedral approach · Branch-and-cut algorithm

1 Introduction

For the past few decades, combinatorial optimization techniques have shown to be powerful tools for formulating, analysing and solving optimization problems arising from practical decision situations. In particular many problems from VLSI (very

P. Fouilhoux
Laboratoire LIP6, CNRS UMR 7626, Université Pierre et Marie Curie, 4 place Jussieu, 75005 Paris, France
e-mail: Pierre.Fouilhoux@lip6.fr

A.R. Mahjoub (✉)
Université Paris-Dauphine, LAMSADE, CNRS, Place du Maréchal de Lattre de Tassigny, 75775 Paris CEDEX 16, France
e-mail: mahjoub@lamsade.dauphine.fr

large scale integrated) circuit design and molecular biology have been formulated as combinatorial optimization models. A big amount of research has also been done for designing algorithmic approaches for these problems [6, 16, 24, 26, 32]. However, for some optimization problems from these areas, for which efficient algorithmic and computational methods are needed, even an adequate model is still unknown. In this paper, we give two further applications of combinatorial optimization to VLSI circuit design and DNA sequencing. We show that the constrained via minimization problem and the SNP haplotype assembly problem, with respect to the so-called minimum error correction criterion, can be reduced to the maximum bipartite induced subgraph problem. We also discuss the bipartite induced subgraph polytope and devise a branch-and-cut algorithm for solving these problems. In order to present these problems we first need to give some definitions.

Let $G = (V, E)$ be a graph. If $W \subset V$, then $E(W)$ denotes the set of all edges of G with both endnodes in W . The graph $H = (W, E(W))$ is the subgraph of G induced by W . A graph is called *bipartite* if its node set can be partitioned into two non empty disjoint sets V_1 and V_2 such that no two nodes in V_1 and no two nodes in V_2 are linked by an edge. Given a weight function $c : V \rightarrow \mathbb{R}$ that associates with every node v a weight $c(v)$, the *bipartite induced subgraph problem* (BISP for short) is to find a bipartite induced subgraph $(W, E(W))$ of G such that $c(W) = \sum_{v \in W} c(v)$ is as large as possible.

A *stable set* of a graph is a set of pairwise non adjacent nodes and the stable set problem consists of finding a stable of maximum weight. The BISP is a generalization of the maximum stable set problem. In fact, if $H = (W, F)$ is a graph, then the maximum stable set problem in H can be reduced to the BISP in the graph $G = (V, E)$ obtained from H by adding for every edge uv of H , a node w with weight M , where M is a big positive value, and the edges wu and wv . This implies that the BISP is NP-hard. The BISP has been shown to be NP-hard even in graphs with maximum degree three and in planar graphs when the maximum degree is ≥ 4 [10]. The BISP is solvable in polynomial time in series-parallel graphs [3] and in planar graphs when the maximum node degree is limited to three [10]. (A graph is called series-parallel if it can be obtained from a graph consisting of one edge by subdivisions and duplications of edges.)

Let us now describe two applications in which the maximum bipartite subgraph problem arises in a rather natural way.

One important problem in VLSI design is to reduce the number of *vias* (holes in a printed circuit board, contact cut on a chip) of a 2-layer electronic circuit. This problem, called the *constrained via minimization problem* (CVMP), has been extensively investigated in the case where the so-called maximum junction degree d is limited to three. Hashimoto and Stevens [20] are the first who studied this problem when $d \leq 2$. This case has been shown later to be polynomially solvable by [22]. Chen et al. [9] and Pinter [29] have independently shown that the CVMP with $d \leq 3$ can be reduced to the max-cut problem in planar graphs, and can then be solved in polynomial time. Barahona et al. [5] described a cutting plane algorithm for the problem in that case. Choi et al. [10] showed that the CMVP with $d \geq 4$ is NP-complete. However, to the best of our knowledge, no exact method has been developed for the CVMP when $d \geq 4$.