Consecutive Ones and a Betweenness Problem in Computational Biology

Thomas Christof, Marcus Oswald, and Gerhard Reinelt

Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 293/294
D-69120 Heidelberg, Germany
{Thomas.Christof, Marcus.Oswald, Gerhard.Reinelt}@@IWR.Uni-Heidelberg.De

Abstract. In this paper we consider a variant of the betweenness problem occurring in computational biology. We present a new polyhedral approach which incorporates the solution of consecutive ones problems and show that it supersedes an earlier one. A particular feature of this new branch-and-cut algorithm is that it is not based on an explicit integer programming formulation of the problem and makes use of automatically generated facet-defining inequalities.

1 Introduction

The general *Betweenness Problem* is the following combinatorial optimization problem. We are given a set of n objects $1, 2, \ldots, n$, a set $\mathcal{B}$ of betweenness conditions, and a set $\overline{\mathcal{B}}$ of non-betweenness conditions. Every element of $\mathcal{B}$ (of $\overline{\mathcal{B}}$) is a triple (i, j, k) (a triple $\overline{(i, j, k)}$) requesting that object j should be placed (should not be placed) between objects i and k. The task is to find a linear order of all objects such that as few betweenness and non-betweenness conditions as possible are violated, resp. to characterize all orders that achieve this minimum. If violations are penalized by weights, we call the problem of finding a linear order minimizing the sum of weights of violations *Weighted Betweenness Problem*. This problem is $\mathcal{NP}$-hard in general.

In this paper we consider a special variant of this problem occurring in computational molecular biology, namely in the physical mapping problem with end probes. For the purpose of this paper we do not elaborate on the biological background, but refer to [4]. We define the problem simply as follows. We are given a set of m so-called *clones* $i \in \{1, 2, \ldots, m\}$ to each of which two so-called *end probes* i^t and i^h are associated. These $n = 2m$ probes are numbered such that $i^t = 2i - 1$ and $i^h = 2i$. Depending on the data, we have for every pair of a clone i and a probe $j \in \{1, 2, \ldots, n\} \setminus \{i^t, i^h\}$ either a betweenness condition (i^t, j, i^h) or a non-betweenness condition $\overline{(i^t, j, i^h)}$. Violation of a betweenness condition is penalized with cost c_ρ, and violation of a non-betweenness constraint receives a

R. E. Bixby, E. A. Boyd, and R. Z. Ríos-Mercado (Eds.): IPCO VI
LNCS 1412, pp. 213–228, 1998.