A Mixed Integer Linear Programming Approach for Metabolic Network Completion Problem

Kerian Thuillier[1], Anne Siegel[1], Rumen Andonov[1]

Univ Rennes, Inria, CNRS, IRISA, Rennes, France,
kerian.thuillier@ens-rennes.fr, {anne.siegel, rumen.andonov}@irisa.fr

Mots-clés : MILP, *flow network, gap filling, metabolic network completion problem*

1 Introduction

Among all the tools used to study cells, metabolic networks are a way to represent all the chemical reactions used by them. Although we have a lot of information for well-studied species, which therefore are associated with high-quality metabolic networks, poorly studied organisms suffer from a data incompleteness. To fulfill this lack of knowledge, computational methods were developed to suggest putative reactions allowing the completion of less studied species metabolic networks thanks to reactions appearing in networks of related well-known species. There are many different network completion problems, which can be classified in three classes [4]. The first class of problem is usually solved with linear programming approaches. The second class is formulated as combinatorial problems currently processed with logical programming. The third class of problems is a hybrid version merging combinatorial and linear constraints currently processed with hybrid solvers combining linear programming and boolean satisfiability kind of constraints [2]. Our objective is to model with linear constraints the topological metabolic network completion problem in order to solve the hybrid problem with a Mixed Integer Linear Programming approach (MILP). This problem belongs to the second class of problems and is currently solved with combinatorial approach. Since the hybrid completion is naturally deduced from the topological completion, our contribution allows us to derive a new approach to the problem of hybrid metabolic network completion [2] fully based on MILP.

2 Metabolic Network Completion Problem

Metabolic Networks are often modeled as bipartite graphs G such that $G = (R \cup M, E, s)$ with respectively R and M the reaction and metabolite sets, E the edges with weights s corresponding to stoichiometric coefficients. What distinguishes metabolic networks from conventional bipartite graphs is the presence of precise rules for topological activation. Before defining these rules, let us start by explaining the notion of topological activation. A node (metabolite or reaction) t is topologically activated from a set of input source nodes S if there is a path from a subset of the elements of S to t such that, for all reaction node along the path, all predecessors to each reaction are themselves connected to the source nodes S according to the topological activation semantics. This corresponds to build a Boolean dynamical system from the bipartite graphs, based on conjonction rules for all reaction nodes and disjunction rules for all metabolic nodes, and to compute all nodes activated by the dynamics from the input sources. In order to formalize this notion, let us note δ_v^- all the predecessors of the node v. Thus in the case of metabolic networks, we have a metabolite node $m \in A_t(S)$ (*i.e.* m is topologically activated from a set of nodes S) *iff* $\exists r \in \delta_m^-$, $r \in A_t(S)$; and a reaction node $r \in A_t(S)$ *iff* $\forall m \in \delta_r^-$, $m \in A_t(S)$.

Given two metabolic networks – a network $G_1 = (R_1 \cup M_1, E_1, s_1)$ to be completed and a network $G_2 = (R_2 \cup M_2, E_2, s_2)$ used as reference – and two sets – targets reactions $T \subseteq R_1$

and sources metabolites $S \subseteq M_1$ –, the topological completion problem consists in determining the minimum number of reactions of R_2 to be added to the network G_1 in order to activate all targets reactions T from the source metabolites S according to the topological activation semantics.

Another kind of completion problem is stoichiometric completion which belongs to the first class of problem, and aims to ensure that all elements of T have a strictly positive flow which is spread from S according to metabolic flux conservation. The hybrid completion problem consists in solving the topological completion and a stoichiometric completion.

3 Our approach

Here we proposed to treat the topological completion problem as a flow problem. To determine the topological activation of a set of target nodes T from a set of source nodes S, we emit a flow from S to T. Each node with a non-null flow value is then topologically activated from S. Denote by a_v a binary variable set to 1 if the node v is topologically activated from S. Our specific constraints to metabolic networks are then formulated : $\forall m \in M$, $a_m \leq \sum_{r \in \delta_m^-} a_r \leq a_m \times |R \cup M|$, where $|R \cup M|$ is used as big-M, and $\forall r \in R$, $\forall m \in \delta_r^-, a_r \leq a_m$.

However, these conditions are not enough, because metabolic networks have self-activating loops. The elements of these loops are therefore not guaranteed to be topologically activated from S. To remove these self-activated loops, we use a method inspired by the Miller-Tucker-Zemlin technique (MTZ) [1]. Based on this method, the hybrid problem is straightforward by adding a new flow that we aim to maximize.

4 Results

This model is implemented with AMPL and use Python *NetworkX* library to extract data from *sbml* file (standard file format used to store metabolic networks). This implementation was tested on the *Meneco* benchmark of *Escherichia Coli IJR904* [4]. This benchmark is composed of several metabolic networks obtained by degrading the initial metabolic network of *IJR904* by 10, 20, 30 and 40 %. On 360 instances tested (180 degraded by 10% and 180 degraded by 40%), all instances degraded by 10% finished in less than 1 000 seconds, and 15 of instances degraded by 40%. These results are better than the results currently obtained by *Fluto* [2], a tool that performs hybrid completion in logical programming with linear overlay. Note that instances degraded by 40% that have not finished in less than 1 000 seconds still return solutions. However, these solutions are not guaranteed to be the minimal solutions. The correction of these solutions is verified by two tools : *MeneTools* an ASP-based software checking the topological activation of reactions [4], and *CobraPy* a python module computing metabolic flux [3].

Références

[1] C. E. Miller, A. W. Tucker and R. A. Zemlin. Integer programming formulation of the traveling salesman problems. *Journal of the ACM*, 7 Issue 4 :326–329, 1960.

[2] Clémence Frioux et al. Technical note hybrid metabolic network completion. *Cambridge University Press*, 2018.

[3] Ali Ebrahim, Joshua A. Lerman, Bernhard O. Palsson, and Daniel R. Hyduke. Cobrapy : Constraints-based reconstruction and analysis for python. *BMC Systems Biology*, 7(1) :74, Aug 2013.

[4] Sylvain Prigent et al. Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLOS Computational Biology*, 2017.