ERCIM "ALAIN BENSOUSSAN" FELLOWSHIP PROGRAMME

ERCIM
European Research Consortium for Informatics and Mathematics

Scientific Report

First name / Family name	Arne / Reimers
Nationality	German
Name of the *Host Organisation*	Centrum Wiskunde & Informatica
First Name / family name of the *Scientific Coordinator*	Gunnar / Klau
Period of the fellowship	01/10/2014 to 31/08/2015

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

During my fellowship I analysed the applicability of branch-decompositions to metabolic networks (networks of chemical reactions in living organisms) and to polytope-theory in general. During my PhD I had shown that extreme pathways in metabolic networks (vertices of polytopes) can be enumerated in total polynomial time if the branch-width of the network is bounded by a constant. Hence, I tried during my fellowship to

- extend the computational result also to other algorithmic problems in the fields of metabolic network analysis and polytope-theory,
- find good branch-decompositions of metabolic networks and to apply the enumeration algorithm also in practice.

Furthermore, I compared the existing concepts of flux modules and I joined a problem-solving group where we investigated variants of the Maximum Weight Connected Subgraph (MWCS) problem.

Extension to other algorithmic problems

Every decision problem on polyhedra that I investigated also remained hard even if I restricted the problem to instances of bounded branch-width. For example, I showed that it remains NP-hard to decide if a polyhedron contains a vertex that does not lie on a given hyperplane even if the polyhedron has branch-width 3. In metabolic networks, this problem corresponds to finding an extreme pathway that contains two given reactions, which

is an important problem for engineering organisms that produce a specific desired by-product.
It surprised me that I only found decision problems that remained hard w.r.t. bounded branch-width, since in graph theory many NP-hard problems are easy for instances of bounded branch-width (typically this has been shown for bounded tree-width, which is equivalent to bounded branch-width).

A talk by Daniel Dadush, a researcher at CWI, then gave me the essential ingredient to resolve this apparent paradox. Together with Nicolai Hähnle, he showed that linear programs can be solved in strongly polynomial time if the polytope of feasible solutions has a specific roundness property. This roundness property is related to the largest absolute value of a subderminant of the coefficient matrix.

This parameter together with branch-width (based on a slightly different connectivity function, which I developed during my REP in Lyon, France) now also allows me to get tractability results for the decision problems that were still hard with bounded branch-width only. This closes the gap to the tractability on graphs, because the coefficient matrix of directed graphs is totally unimodular, i.e. the value of every subdeterminant is -1,0 or 1.

A manuscript containing these results is in preparation.

Branch-decompositions of metabolic networks in practice
To compute branch-decompositions of metabolic networks, I tried out the heuristics developed by Ma et al. [1] and an algorithm by Poolman et al. [2]. The method by Ma et al. [1] was directly developed as a branch-width heuristic targeted at instances arising in integer linear programming. On the other hand, the method by Poolman et al. [2] only coincidentally turned out to be a branch-width heuristic. However, the method by Poolman et al. [2] was developed for instances arising from metabolic networks.

It turned out that the method by Ma et al. [1] was enourmously outperformed by the method of Poolman et al. [2]. However, also the method of Poolman et al. [2] (and variations thereof) also only produced branch-decompositions of relatively large branch-width. While these results could still be artifacts of the heuristic, I consider these results as strong indications that metabolic networks actually do have a rather large branch-width and are not very modular.

A manuscript with these results is submitted to *Biochemical Society Transactions* for a special issue on the Metabolic Pathway Analysis 2015 conference.

Never the less, I also implemented the extreme pathway enumeration algorithm together with the heuristics based on the method of Poolman et al. [2] in the cbmpy toolbox, which is maintained by Brett Olivier at VU Amsterdam. For small networks (E. coli core), it is possible to enumerate all pathways with this method, if some side constraint like 95%-optimality is enforced.

Further activities
To compare the existing concepts of flux modules, I built a unifying theory and showed

that the differences are mostly caused by the fact that the methods analyzed different flux spaces. Together with Timo Maarleveld I worked on a manuscript that illustrates the differences.

Together with Gunnar Klau, Leen Stougie, Leo van Iersel, and Thomas Hume I analysed the maximum weight cross-connected subgraph problem (MWCCS), which has applications in biological network comparison. For a special case, we could reduce the problem to MWCS with budget constraints (B-MWCS). We showed that we can solve B-MWCS in pseudo-polynomial time for instances of bounded tree-width. Also for this work, a manuscript is in preparation.

I also worked on publishing some results that I obtained during my PhD.

[1] Ma J, Margulies S, Hicks IV, and Goins E. Branch decomposition heuristics for linear matroids. *Discrete Optimization* 10:102-109, 2013
[2] Poolman MG, Sebu C, Pidcock MK, Fell DA. Modular decomposition of metabolic systems via null-space analysis. *Journal of Theoretical Biology* 240:691-705, 2007

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

- Arne C. Reimers, Frank J. Bruggeman, Brett G. Olivier, and Leen Stougie. Fast flux module detection using matroid theory. *Journal of Computational Biology*. May 2015, 22(5): 414-424.

  Abstract: Flux balance analysis (FBA) is one of the most often applied methods on genome-scale metabolic networks. Although FBA uniquely determines the optimal yield, the pathway that achieves this is usually not unique. The analysis of the optimal-yield flux space has been an open challenge. Flux variability analysis is only capturing some properties of the flux space, while elementary mode analysis is intractable due to the enormous number of elementary modes. However, it has been found by Kelk et al. (2012) that the space of optimal-yield fluxes decomposes into flux modules. These decompositions allow a much easier but still comprehensive analysis of the optimal-yield flux space. Using the mathematical definition of module introduced by Müller and Bockmayr (2013b), we discovered useful connections to matroid theory, through which efficient algorithms enable us to compute the decomposition into modules in a few seconds for genome- scale networks. Using that every module can be represented by one reaction that represents its function, in this article, we also present a method that uses this decomposition to visualize the interplay of modules. We expect the new method to replace flux variability analysis in the pipelines for metabolic networks.

- Arne C. Reimers., Yaron Goldstein, and Alexander Bockmayr. Generic flux coupling analysis. Mathematical BioSciences. April 2015, 262:28-35

  Abstract: Flux coupling analysis (FCA) has become a useful tool for aiding metabolic reconstructions and guiding genetic manipulations. Originally, it was introduced for constraint-based models of metabolic networks that are based on the steady-state assumption. Recently, we have shown that the steady-state assumption can be replaced by a weaker lattice-theoretic property related to the supports of metabolic fluxes. In this paper, we further extend our

approach and develop an efficient algorithm for generic flux coupling analysis that works with any kind of qualitative pathway model. We illustrate our method by thermodynamic flux coupling analysis (tFCA), which allows studying steady-state metabolic models with loop-law thermodynamic constraints. These models do not satisfy the lattice-theoretic properties required in our previous work. For a selection of genome-scale metabolic network reconstructions, we discuss both theoretically and practically, how thermodynamic constraints strengthen the coupling results that can be obtained with classical FCA. A prototype implementation of tFCA is available at http://hoverboard.io/L4FC.

Papers prepared during the fellowship period:

- Arne C. Reimers. Hierarchical decomposition of metabolic networks using k-modules. *Biochemical Society Transactions*. In press
- Arne C. Reimers. Obstructions to sampling qualitative properties. *PLOS ONE*. In press
- Arne C. Reimers, Alexandra M. Reimers, Yaron Goldstein. Minimal equivalent subgraphs containing a given set of arcs (submitted to Theoretical Computer Science)

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

- 40th Conference on the Mathematics of Operations Research, January 13-15, 2015 in Lunteren, Netherlands
- Metabolic Pathway Analysis, June 8-12, 2015 in Braga, Portugal

IV – RESEARCH EXCHANGE PROGRAMME (REP)

In my research exchange programme, I visited the BAMBOO/ERABLE team of Marie-France Sagot at INRIA Grenoble Rhone-Alpes in Lyon, France. Together with Marie-France Sagot and Murray Patterson, I worked there on modularization approaches to simplify metabolic networks. In this context we developed a more topological form of modularization than the one that I developed during my PhD-thesis.

This alternative form also gives rise to a submodular connectivity function that can be used to define branch-decompositions. In particular it gives better (theoretical) bounds on the complexity of subnetworks than my old connectivity measure. Therefore, the REP played an important part for one of my main results of this fellowship programme.