Determining MinSpan vectors of COBRA model

Aarash Bordbar

Background

In this tutorial, we show how to calculate MinSpan vectors [1] for a COBRA model. COBRA models are predominantly studied under steady-state conditions, thus the null space of the **S** matrix is of high interest. MinSpan vectors are the sparsest linear basis of the null space of **S** matrix while still maintaining the biological and thermodynamic constraints of the COBRA model. The **S** matrix has dimensions (**m** x **n**) where **m** is the number of metabolites and **n** is the number of reactions. The linear basis of the null space (**N**) has dimensions (**n** x **n-r**) where **r** is the rank of **S**. Thus this algorithm calculates **n-r** vectors that are linearly independent of each other and also are minimal. For further info on MinSpan, it's derivation, implementation, and uses, see Bordbar et al. 2014 [1].

Implementation

MinSpan vectors are calculated by mixed integer linear programming (MILP) using the COBRA function `detMinSpan`. The algorithm requires a robust MILP solver such as Gurobi v5+. This tutorial and the algorithm has been rigorously tested and support Gurobi v5+, which is freely available for academics. `detMinSpan` will not work with GLPK and other solvers are untested.

For MinSpan vectors to be calculated, the model must 1) consist of only reactions that are able to carry flux under that particular condition, 2) allow for the trivial solution (**v** = 0) to be feasible, and 3) have the biomass function removed. `detMinSpan` will automatically check and complete the first two modifications, but the biomass must be removed manually.

The algorithm is an iterative pruning of null space basis vectors to the sparsest possible matrix. The problem is NP hard meaning that for large COBRA models, an optimal solution is not guaranteed and an approximate solution is found by setting a time limit on the MILP calculation.

Example

In this example, we will calculate the MinSpan vectors for the *E. coli* core network. We begin by initializing the COBRA toolbox in Matlab and loading the core model.

```
initCobraToolbox()
load e_coli_core
```

The biomass function is then removed from the model using the COBRA function `removeRxns`.

```
bmName = {'BIOMASS_Ecoli_core_w_GAM'};
e_coli_core = removeRxns(e_coli_core, bmName);
```

The algorithm is then run.

```
minSpanVectors = detMinSpan(e_coli_core);
```

The `minSpanVectors` matrix consists of 23 linearly independent vectors. A further description of these vectors is provided and their comparison to Extreme Pathways [2] is provided in the Supplementary Material of Bordbar et al 2014 [1] (Section 1 and Figure S2).

References

1.) Bordbar A, Nagarajan H, Lewis NE, Latif H, Ebrahim A, Federowicz S, Schellenberger J, Palsson BO. "Minimal metabolic pathway structure is consistent with associated biomolecular interactions" *Mol Syst Biol* **10:**737 (2014).

2) Schilling CH, Letscher D, Palsson BO. "Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* **203:**229-248 (2000).