Reverse engineering of metabolic networks, a critical assessment

Diana M. Hendrickx, Huub C.J. Hoefsloot, Margriet M.W.B. Hendriks, Age K. Smilde

Biosystems data analysis, University of Amsterdam
Metabolic and Endocrine Diseases, UMC Utrecht

D.M.Hendrickx@uva.nl

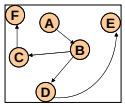
Netherlands Metabolomics Centre

Overview

Inferring metabolic networks from metabolite concentration data is a topic in systems biology that can help to improve our understanding of the functioning of cellular systems. Mathematical techniques to extract information about the network from data have been proposed in literature. However, if full inference of a real world metabolic network is the goal there exists a large discrepancy between the requirements of the inference methods with regard to the sampling frequency and noise levels on the one hand and the contemporary measurement practice on the other.

Networks from measured data

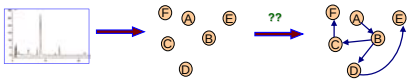
1. What?



Metabolic network = network composed of metabolites and their biochemical interactions in the cell of an organism.

nodes = metabolites
edges = interactions

2. How?



A — B : topology
1A — 1B : stoichiometry
A → B : directionality
A —k→ B : kinetics

3. This study

NETWORK TOPOLOGY

DIRECTIONS

4. Problems

NOISE

MISSING METABOLITES

HUGE AMOUNT OF POSSIBLE NETWORK STRUCTURES

5. Research questions

Many network inference methods exist, but:

- Do these methods work for real world networks?
- Is the sampling frequency large enough to apply the methods?
- Influence of noise on the performance of the methods?

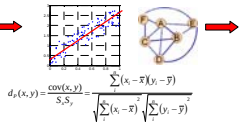
Network inference methods

1. Partial Pearson correlations

(Cakir et al., 2009)

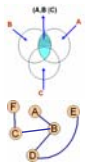
1. DATA COLLECTION: Steady state data (biological variation)

2. SIMILARITY MEASURE: Pearson correlation (linear)



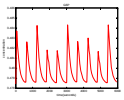
$$d_p(x,y) = \frac{\mathrm{cov}(x,y)}{S_x S_y} = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2}\sqrt{\sum_i (y_i-\bar{y})^2}}$$

3. CONDITIONING

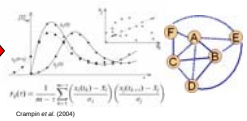


2. Time-lagged correlations

1. DATA COLLECTION: Time series



2. SIMILARITY MEASURE: Time lagged correlations

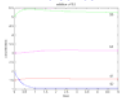


Crampin et al. (2004)

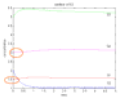
3. PRUNING

3. Zero slopes method

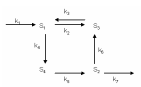
1. DATA COLLECTION: Time series



2. ZERO SLOPES



3. NETWORK

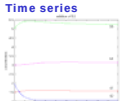


Example:
perturbing metabolite S2

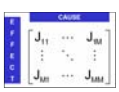
- decreasing graph of perturbed metabolite.
- zero slope = indirect interaction between S2 and the metabolite which concentration profile is presented by the graph.
- non-zero slope = direct interaction
- constant graph = no interaction

4. Jacobian method

1. DATA COLLECTION: Time series



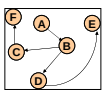
2. JACOBIAN



$$\frac{dS(t)}{dt} = F(S(t)) = J \cdot S$$

$$J = \frac{\partial F}{\partial S}$$

3. NETWORK

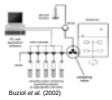


J: Jacobian
S: concentrations of the metabolites
F: function that determines the time evolution of the concentration of each metabolite

Example:
J_{M1} describes how changes in metabolite 1 affect changes in metabolite M.

Current measurement practice

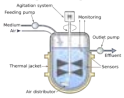
Bioreactor + rapid sampling device
(seconds / subsecond)



Buziol et al. (2002)

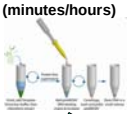
Micro-organisms

Only bioreactor
(hours)



Analytical method
Noise level = 5 – 25%

Plant tissue samples
(minutes/hours)



Human blood samples
(minutes / hours)

Animal blood samples
(minutes / hours)

Analytical method
Noise level = 5 – 25%

Important: - sampling frequency
- noise

Results

1. Problems with reverse engineering

- Weak interactions → missing edges
- Spurious edges → indirect interactions
- Large influence of noise
- All metabolites have to be measured

2. Requirements of the network inference methods

- Very fast sampling frequencies (often much faster than possible in practice)
- Very low noise levels (<< 3%, not possible with contemporary experiments)

Conclusions

- If full inference of a real world metabolic network is the goal, then the requirements for the sampling frequency are not consistent with contemporary practice.
- We do not need to estimate the whole network from the data alone. There exists already a lot of biological information in databases.
- Future research: incorporate *a priori* knowledge from databases to improve network inference.