Inferring metabolic networks from time series data

Emily S.C. Ching and Chumin Sun

Department of Physics, The Chinese University of Hong Kong, Shatin, Hong Kong

Introduction

Metabolic networks comprise the chemical reactions of metabolism and the regulatory interactions that guide these reactions. They are often modeled by a set of differential equations which govern the time evolution of the concentration of the metabolites. Our aim is to reconstruct metabolic networks from the concentration time series of the metabolites.

Method and Results

1. A simple model [1]

$$\frac{dx_i}{dt} = e_i - \omega_i x_i - \sum_{j=1}^{N} g_{ij} A_{ij} x_i x_j + \eta_i$$

where $x_i(t)$ is concentration of metabolite i ; e_i is the generation rate of metabolite i ; $\omega_i x_i$ accounts for the degradation process. When the production of metabolite i depends on the concentration of metabolite j, there is a link from node j to node i in the network $A_{ij} = 1$, where g_{ij} is the coupling strength. Otherwise $A_{ij} = 0$, especially $A_{ii} = 0$. η_i is external disturbance described by a white noise with zero mean and variance σ^2.

When η_i =0, the system attains a steady solution $\bar{x}_i$. In the presence of weak noise, denote $\delta x_i(t) \equiv x_i(t) - \bar{x}_i$, we obtain

$$\frac{d\delta x_i}{dt} \approx \sum_{j=1}^{N} Q_{ij} \delta x_j + \eta_i$$

where

$$Q_{ij} = \begin{cases} -g_{ij} A_{ij} \bar{x}_i & j \neq i \\ -\omega_i - \sum_{k=1}^{N} g_{ik} A_{ik} \bar{x}_k & j = i \end{cases}$$

when $j \neq i$, Q_{ij} contains information of the network structure A_{ij}. Following [2], we obtain

$$\boldsymbol{K}_\tau \approx e^{\tau \boldsymbol{Q}} \boldsymbol{K}_0$$

where $\boldsymbol{K}_\tau$ denotes the time-lagged covariance matrix of the measurements taken at two times separated by a time interval τ: $(\boldsymbol{K}_\tau)_{ij} = \langle (x_i(t+\tau) - \langle x_i(t+\tau) \rangle)(x_j(t) - \langle x_j(t) \rangle) \rangle$ and $\boldsymbol{K}_0$ denotes the covariance matrix of measurements taken at the same time: $(\boldsymbol{K}_0)_{ij} = \langle (x_i(t) - \langle x_i(t) \rangle)(x_j(t) - \langle x_j(t) \rangle) \rangle$. $\langle \cdots \rangle$ denotes a time average. Using $x_i(t)$, we calculate $\boldsymbol{K}_\tau$ and $\boldsymbol{K}_0$ and thus reconstruct $\boldsymbol{Q}$ by

$$\boldsymbol{M} \equiv \frac{1}{\tau} \log \boldsymbol{K}_\tau \boldsymbol{K}_0^{-1} \approx \boldsymbol{Q} \quad \text{for small } \tau$$

By clustering the off-diagonal elements of $\boldsymbol{M}$, we obtain the reconstructed $A_{ij}^{(e)}$. Moreover, we can reconstruct $g_{ij}^{(e)}$ by

$$g_{ij}^{(e)} \equiv -\frac{M_{ij} A_{ij}^{(e)}}{\langle x_i(t) \rangle} \approx -\frac{Q_{ij} A_{ij}}{\bar{x}_i} = g_{ij}$$

Results

Here we verify our method using a random network of 100 nodes with the connected probability 0.2.

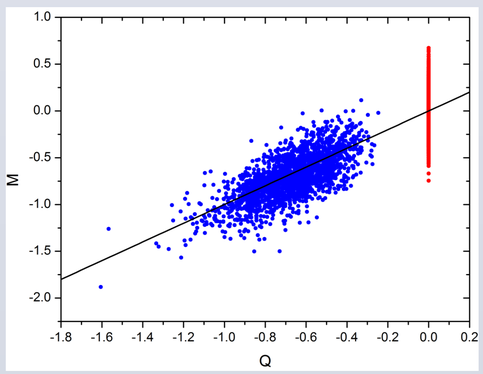

Fig. 1 M_{ij} versus Q_{ij}. Red points stand for unconnected group and blue points stand for connected group. Solid line: y=x.

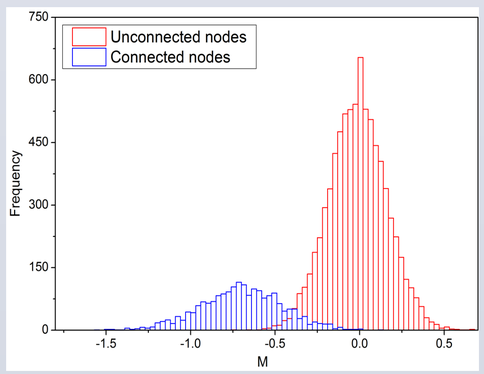

Fig. 2 Histogram of M_{ij} of connected group and unconnected group.

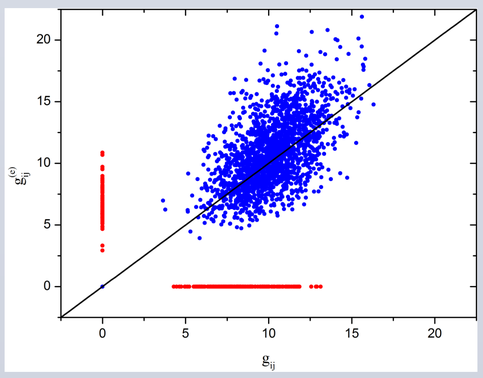

Fig. 3 $g_{ij}^{(e)}$ versus g_{ij}. Red points stand for incorrect inferences. Solid line: y=x.

In Fig. 1, $M_{ij} \approx Q_{ij}$, so clustering analysis of M_{ij} gives an accurate reconstruction of A_{ij} with sensitivity 90.32%, specificity 98.82% and average error rate of $g_{ij}^{(e)}$ 19.23% (see Fig. 2 and Fig. 3).

2. Model by biochemical systems theory [1]

$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^{N} x_j^{g_{ij}} - \beta_i \prod_{j=1}^{N} x_j^{h_{ij}} + \eta_i$$

where α_i, β_i are rate constants; η_i is external disturbance described by a white noise; g_{ij}, h_{ij} are kinetic orders. For $g_{ij} \neq 0$ ($h_{ij} \neq 0$), there is an influx (efflux) link from x_j to x_i (see Fig. 4). In other words, network structure information is contained in g_{ij} and h_{ij}. The system again approaches a fixed point in the noise-free limit and we obtain

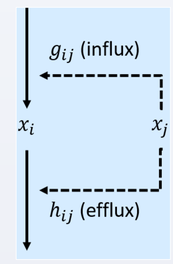

Fig. 4 Influx link and efflux link from node j to node i.

$$\frac{d\delta x_i}{dt} \approx \sum_{j=1}^{N} Q_{ij} \delta x_j + \eta_i, \quad Q_{ij} = \beta_i \prod_{k=1}^{N} \bar{x}_k^{h_{ik}} \frac{g_{ij} - h_{ij}}{\bar{x}_j}$$

If $g_{ij} \neq h_{ij}$, node j is linked to node i (influx or efflux) when $Q_{ij} \neq 0$. By clustering off-diagonal elements of $\boldsymbol{M}$ into two groups, network link information can be extracted.

Results

A. Generic inhibition and activation model

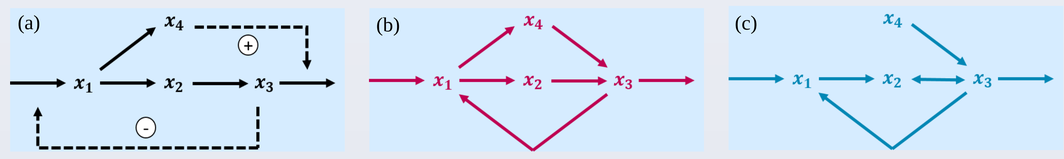

Fig. 5 actual network pathway of generic inhibition and activation model (a), reconstruction results of Granger Causality Test (b) and our method (c).

	Our method	Granger Causality
Missed links	0	1
Incorrectly inferred links	0	1

In Fig. 5, a known metabolic reaction network, the generic inhibition and activation model, has been widely employed as a metabolic case study. We compare our results with that obtained by **Granger Causality Test**, which is widely-used method [3].

B. Lactate model

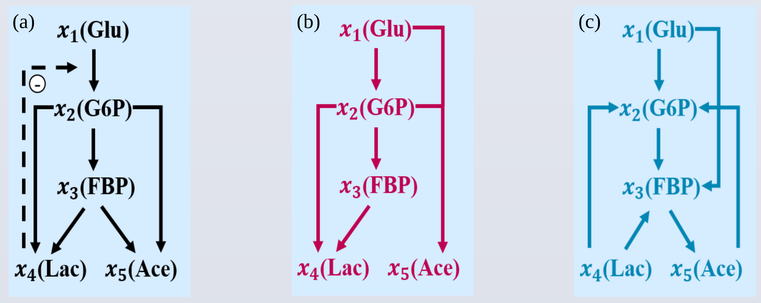

Fig. 6 actual network pathway of Lactate model (a), reconstruction results of Granger Causality Test (b) and our method (c).

	Our method	Granger Causality
Missed links	3	4
Incorrectly inferred links	1	3

To infer the metabolic pathway, we can use the glycolysis pathway in *Lactococcus lactis* [3] (Fig. 6).

C. Aspartate model

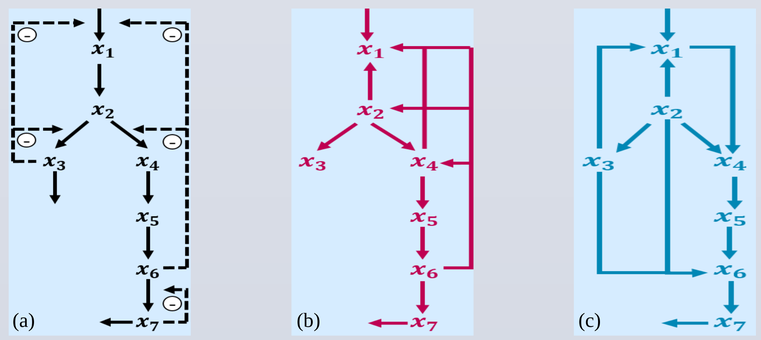

Fig. 7 actual network pathway of Aspartate model (a), reconstruction results of Granger Causality Test (b) and our method (c).

	Our method	Granger Causality
Missed links	4	6
Incorrectly inferred links	2	4

The aspartate-family amino acid biosynthesis pathway in *Arabidopsis thaliana* [3] can also be used to test our method (Fig. 7).

Conclusion

We propose a general method for metabolic network reconstruction. The method only utilizes time series data of metabolites concentration. Moreover, we compare the clustering method with widely-used Granger Causality Test method and obtain better results.

Further work: we are currently studying how to distinguish g_{ij} from h_{ij}, namely find which g_{ij} and h_{ij} are nonzero, some interesting phenomenon show that the clustering method is scalable, more results are to be found.

References

[1] E.O. Voit, ISRN Biomathematics Vol. 2013, 897658 (2013).

[2] Emily S.C. Ching and H.C. Tam, arXiv:1604.02224 (2016).

[3] Sriyudthsak K, Mejia R F, Arita M, et al. Nucleic acids research, gkw415 (2016).