

A Network Analysis of Expression Time Series

Gregory T. Dewey
Keck Graduate Institute
535 Willow Drive
Claremont, CA 91711
Greg_Dewey@kgi.edu

Ashish Bhan
Department of Mathematics
Claremont Graduate University
Claremont, CA 91711
Ashish.bhan@cgu.edu

ABSTRACT

A common approach to tackling complex phenomena is to first establish the extent and limitations of the linear response domain of the system. While there is no *a priori* reason to assume that the time series of a gene expression profile can be well related *via* a simple linear model, it is a reasonable starting point and can provide important clues to the origin of nonlinearity in the system. Linear models, in themselves, can often lead to surprisingly complicated responses, especially for multivariate systems. This provides a strong motivation for an in-depth exploration of such models. In this work, a dynamic, linear response model for analyzing time series of whole genome expression is presented. The simplest assumption about expression profile data is that the expression state represented in the data from one time point determines the expression state seen at the next time point. This assumption is equivalent to modeling the data by a first-order Markov process. The linear version of this model is described by a single transition matrix, Λ , defining the transitions from one state to the next. The expression levels of each of m genes at two points in time can be described by column vectors, $a(t)$ and $a(t-1)$, each of length m . The transition between the two states is

modeled by: $a_i(t) = \sum_{j=1}^m \lambda_{ij} a_j(t-1)$ where $a_i(t)$ is the

expression level of the i th gene at time t after some exposure or treatment. The transition coefficients are λ_{ij} which are the respective elements of the $m \times m$ transition matrix. The matrix elements represent the influence of the expression level of the j th gene on that of the i th gene. Using this model, we calculate a transition matrix for both cell-cycle and diauxic shift data in yeast. These models are statistically robust and lead naturally to a network of interactions reflected in the data. A graphical network can be readily derived from these results by thresholding the parameter λ_{ij} and replacing it with a 1 above a certain threshold

and a 0 below the threshold. This procedure results in a sparse matrix that gives a network representation of the data. This approach can provide a direct method of classifying genes according to their place in the resulting network and offers an alternative to traditional clustering approaches. These network groupings compare favorably with previously used methods like cluster analysis. The network derived by this method shows a hierarchical structure that is dominated by a collection of central hubs. These hubs are interconnected and have a cascade of tree-like structures attached to them. The statistical properties of these resulting networks were determined for a number of different time series data sets for yeast in the public domain. These results consistently show networks that have “small world” characteristics and show scale free distributions of connectivities. A general class of network growth models has been derived that show behavior consistent with the experimental results. Non-linear and higher order Markov behavior of the network can also be included by a self-consistent method. Networks derived from these more sophisticated models show similar behavior.

REFERENCES

- [1] Barabasi, Albert-Laszlo and Reka Albert, Emergence of Scaling in Random Networks, Science, Oct '99, 509-512
- [2] Cormen, T.H., Leiserson, C.E. and R.L. Rivest, Introduction to Algorithms, 1997, MIT Press, Cambridge MA
- [3] Derisi, J.L., Iyer, V. R and P.O. Brown, 1997, Science 278, 680-6
- [4] Dewey, T.G. and D. Galas, 2001, Functional and Integrative Genomics 1, 269-278