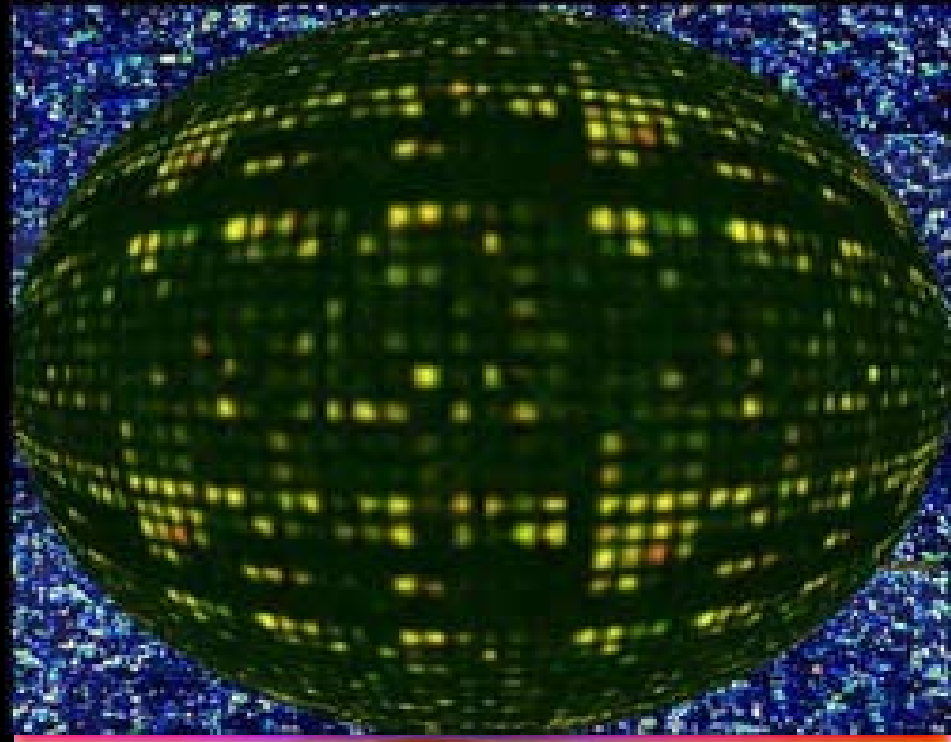
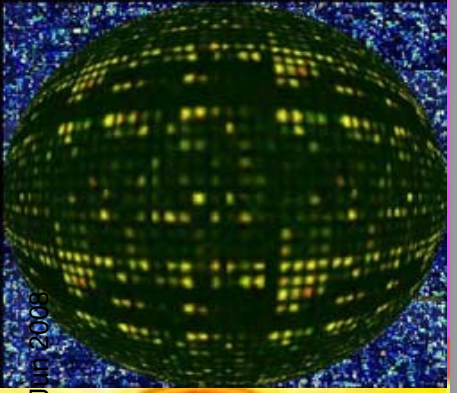
Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

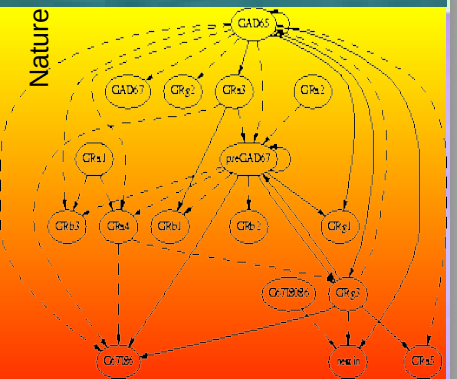
Temporal Gene Expression Analysis



Dissecting the Biological Motherboard

System Biology and Beyond



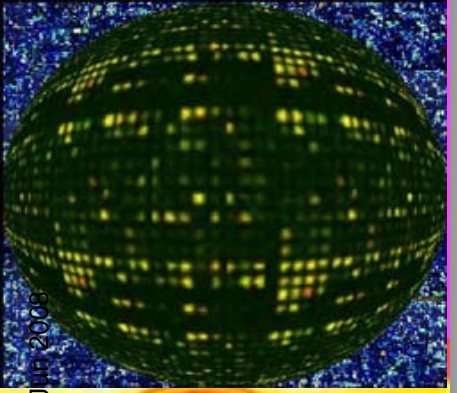


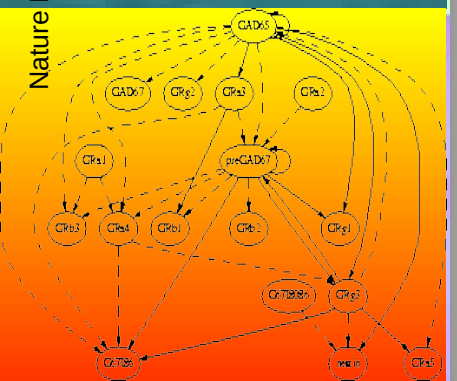
Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

Temporal Organisation

- Research Background (Past) and hypothesis (Past-Present interface)[5min]
- Neural Net System Biology Approach with gene silencing simulations (Present)[15min]
- Conclusions Present-Future interface)[5min]
- Questions/Comments? Future [5min]

Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008





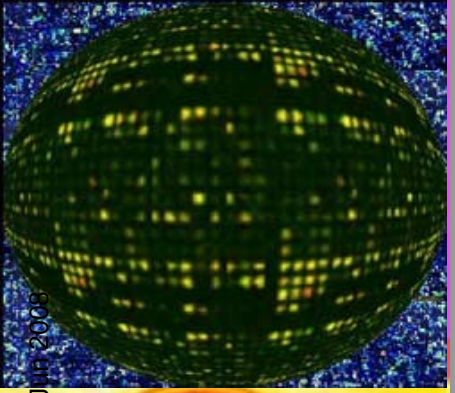
Why Temporal gene expression analysis is important?

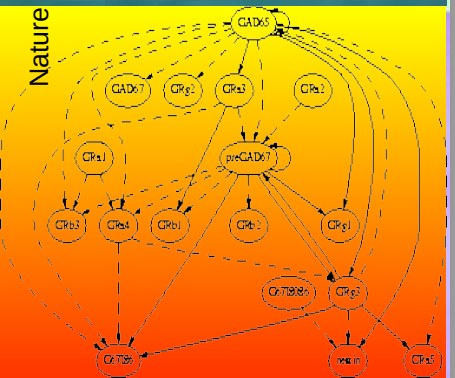
Temporal gene expression data is of particular interest to researchers as it can be used to create regulatory gene networks. Such gene networks represent the regulatory relationships between genes over time and provide insight into how genes up- and down-regulate each other from one time-point to the next (the Biological Motherboard).

Harbinger of hope for the vision of diagnosis, prognosis and medicine being personalised.

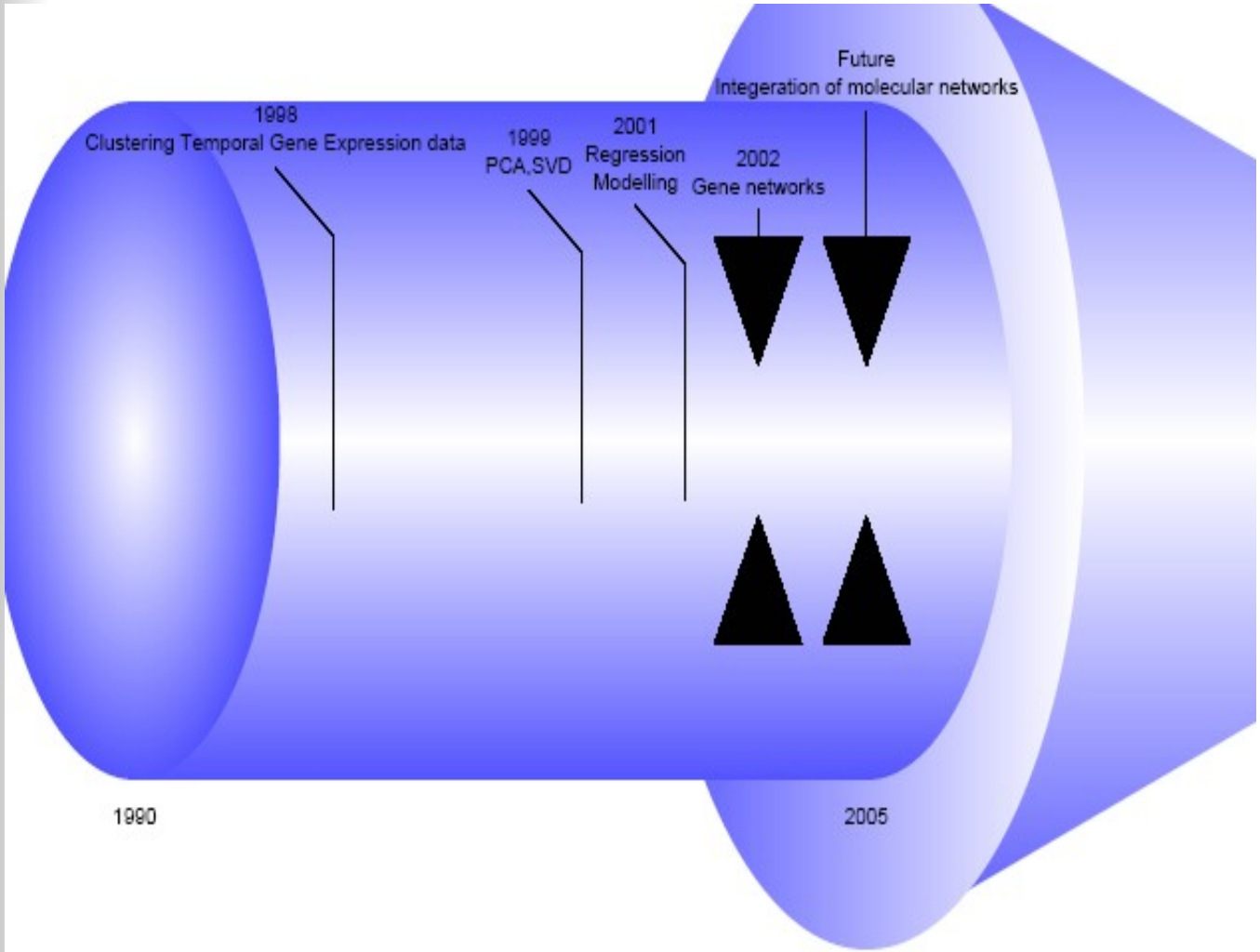
Researchers believe that temporal gene expression models the biological system more accurately and closely as dynamic living systems

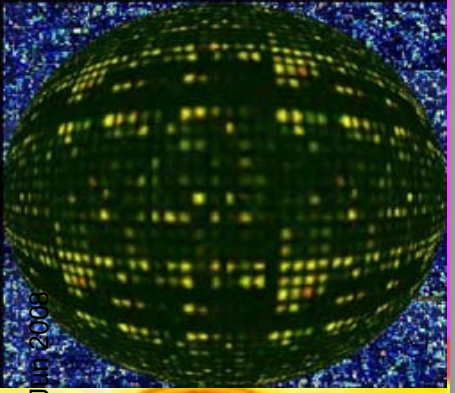
Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

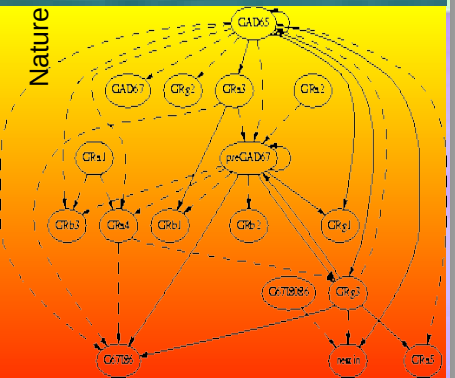




Temporal Gene Expression Analysis Research







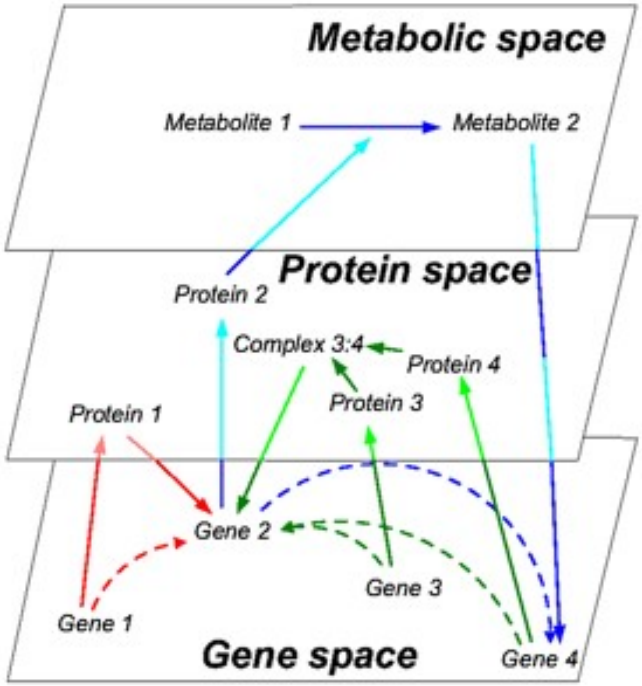
Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

Unique Research Propositions

- *One of the problems in the application of artificial neural networks in this research area concerns to the difficulty in creating complex regulatory networks from the analysis of high-dimensional weight matrices that represent the individual connections between genes over time.*

- *This work introduces an ANN based novel approach by introducing sensitivity analysis as a central theme of the NNSBAGSS methodology discussed here.*

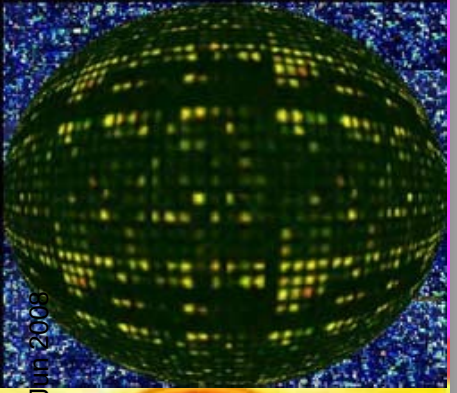
- *Sensitivity analysis is more amenable than weight analysis and can be extended to more complex neural networks, such as those that result from the analysis of large temporal gene expression datasets*

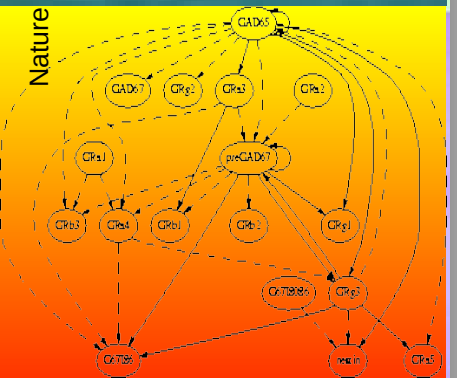


Projections over Phenomenological gene network models, of the whole biochemical system.

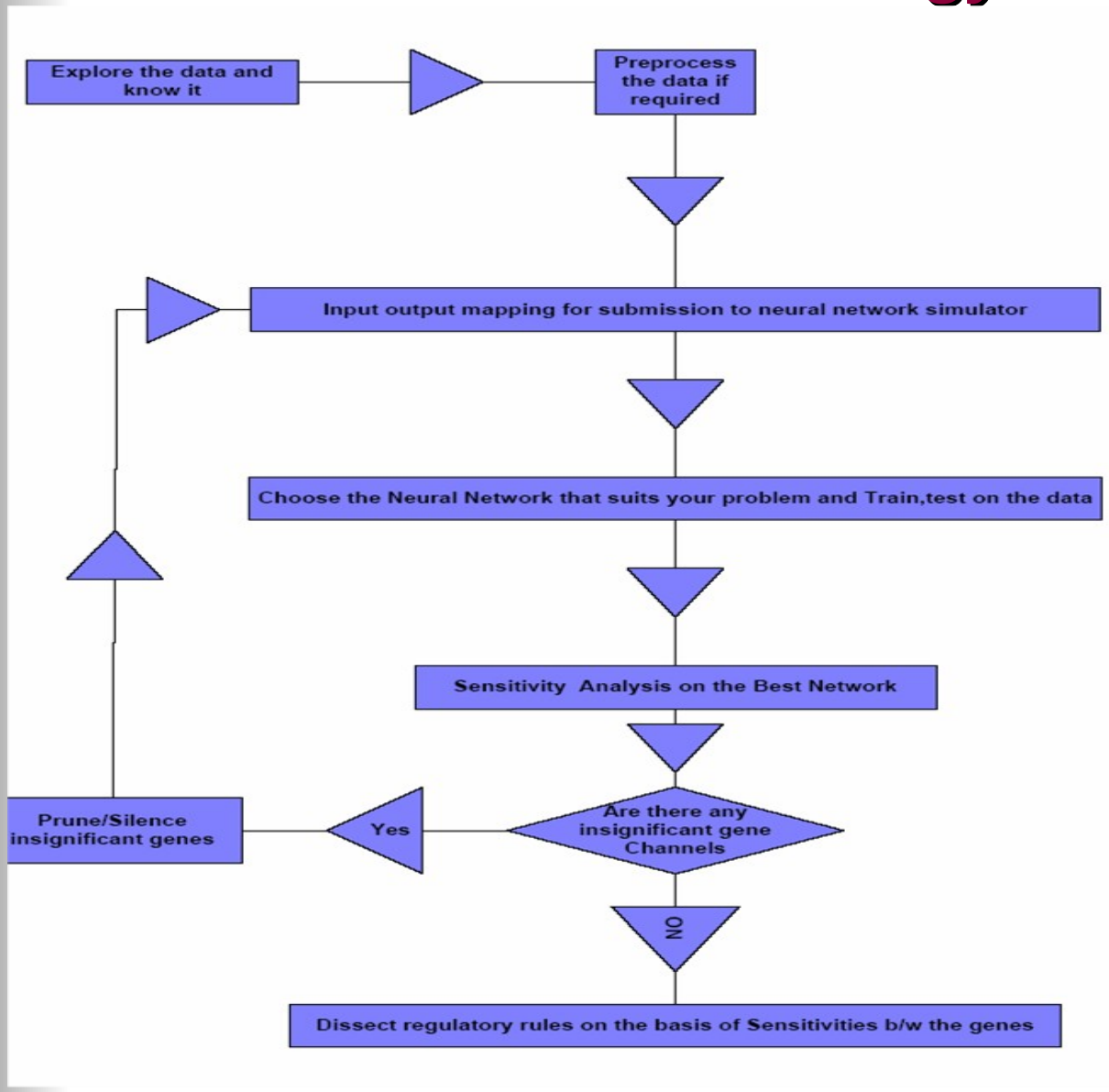
Source: Brazhnik et al, 2002 TIBS

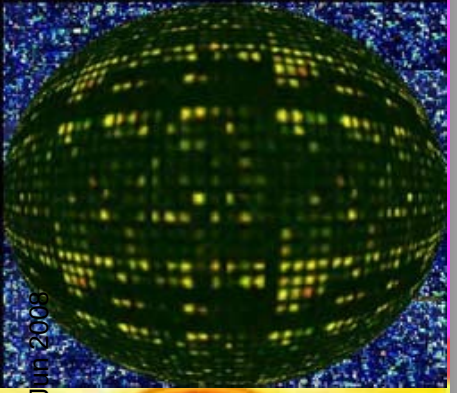
Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

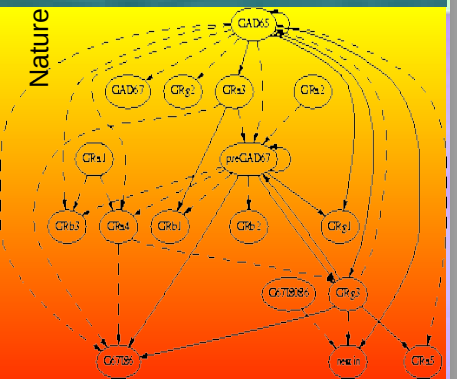




NNSBAGSS methodology







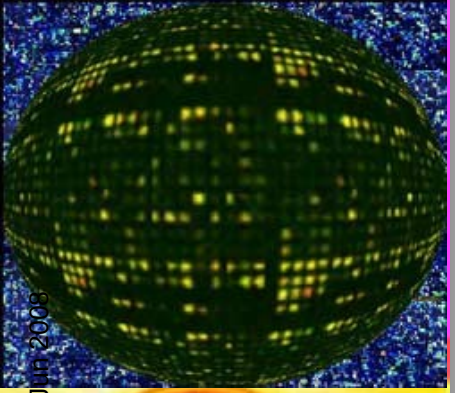
Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

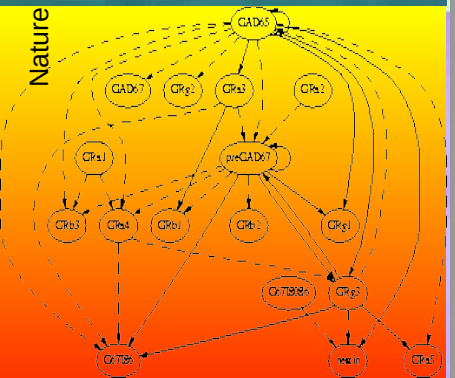
NNSBAGSS,what's in the name?

NeuralNetwork System Biology Approach with Gene Silencing Simulations.

Sensitivity analysis has an analogy with the system biology approach described by Leroy Hood ,Ideker etal in which they see the effectofGenetic and environmentalperturbations in a system and then analysing the effecton other system components thus gaining novelpredictions and hypotheses about the system .

A particular temporalGene expression experiment being regarded as a system we perturb the Gene expression value ofa Gene atprevious time step to see the effecton Gene expression value ofother Genes atsubsequenttime steps thus revealing the inherentcause-effectrelations in our system (temporalGene expression system).





Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

NNSBAGSS methodology : Artificial temporal data

- *The data involved in this experiment is an artificial gene expression dataset forward engineered by a C++ program which creates temporal Boolean Liang networks.*

- *The data contain 10 genes derived from Boolean rules with k values of 2 and 3. The data consists of 2^{10} (1024) records needed to specify every possible truth assignment for each dataset. The data consists of 1024 *20 expression values This data was generated using a C++ programme and had the following 20 inherent rules*

Table 1 - Rules embedded in the artificial dataset

Rule	Arguments	Rule	Arguments
1	3→1	11	8→6
2	4 &10→1	12	5→6
3	6 &10→2	13	8 &7→7
4	4 &9→2	14	6 &10→7
5	5 &4→3	15	9 &3→8
6	9→3	16	4 &1→8
7	2→4	17	7→9
8	9→4	18	5 &4→9
9	6→5	19	5 &10→10
10	9 &6→5	20	1→10

between values at time t and the subsequent value at time t+1. A gene goes off at a time-step unless it is specifically switched on by some gene at the previous time-step.

Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

NNSBAGSS methodology : Artificial temporal data

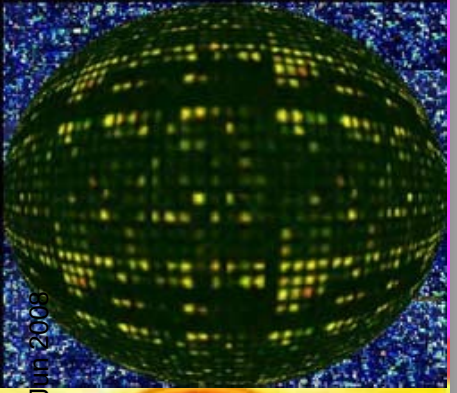
Choice of desired input-output mapping

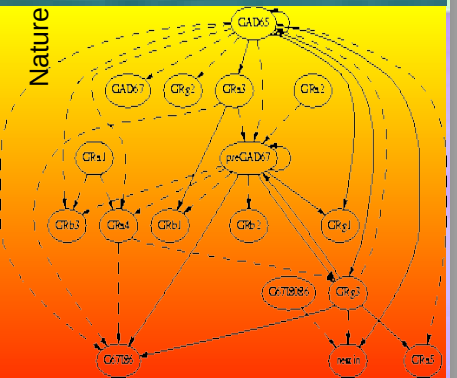
The task that the ANN has to solve is a linear problem ie. the artificial neural network has to learn the input (gene expression value at time t) to output (gene expression value at time t+1) mapping, where no non-linear relationships exist between input values and output value. There are 1024 exemplars of this input-output mapping.

Choice of neural architecture

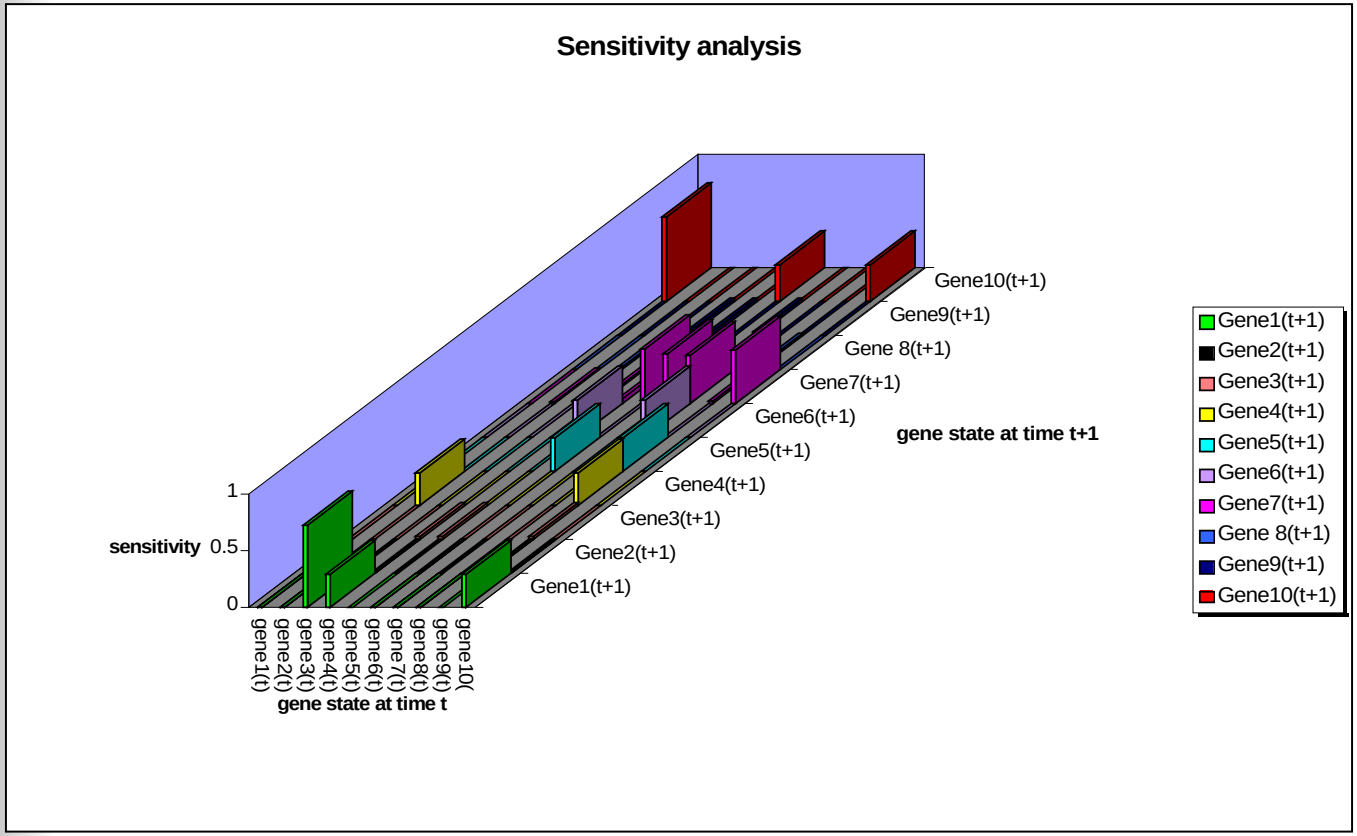
Neural networks can be very powerful learning systems. However, it is very important to match the neural architecture to the problem. For the artificial temporal data we tried a number of single layer perceptrons (that is, full connectivity between the input nodes and output nodes, with no hidden layers), adopting different learning rule parameters and transfer functions. We followed the standard training, cross-validation and testing routine for perceptrons with different parameters. Out of 1024 randomised exemplars we used 60% (614) for training, 15% (154) for cross-validation and 25% (256) for testing. The best network with minimum Mean Squared Error for training as well as cross validation set was chosen. The best neural network architecture was a perceptron with the activation function as the hyperbolic Tanh function. This will squash the range of each neuron in the layer to between -1 and 1.

Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

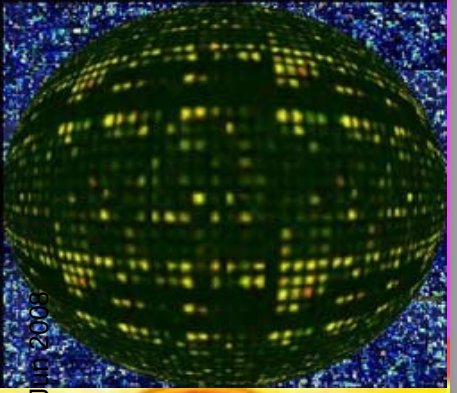


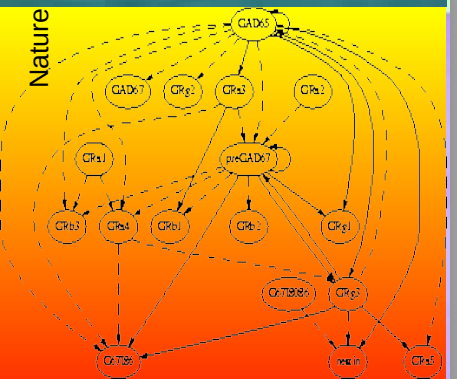


NNSBAGSS methodology : Artificial temporal data



Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008





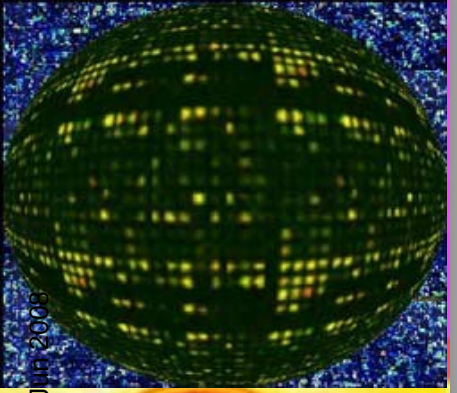
NNSBAGSS methodology : Artificial temporal data

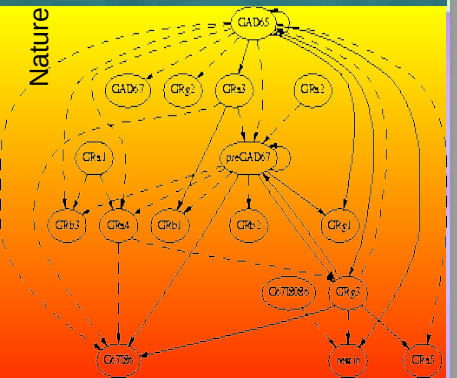
- *By analysing Figure 3 and the sensitivity matrix (not shown) all the embedded rules were reverse engineered by the approach. This can be seen most clearly in the first column of "Gene1(t-1)", where gene 3 has a large affect and genes 4 and 10 have a less marked affect. This corroborates the rules seen earlier as gene 3 can activate gene 1 by itself (Rule 1), whereas it requires the combined effort of gene 4 and gene 10 (Rule 2) to accomplish this. The sensitivity analysis therefore not only defines regulatory behaviour, but also the relative strengths of the regulation. Therefore, this demonstrates that our method is able to reverse engineer the embedded rules from artificial gene expression data.*

Table 1 - Rules embedded in the artificial dataset

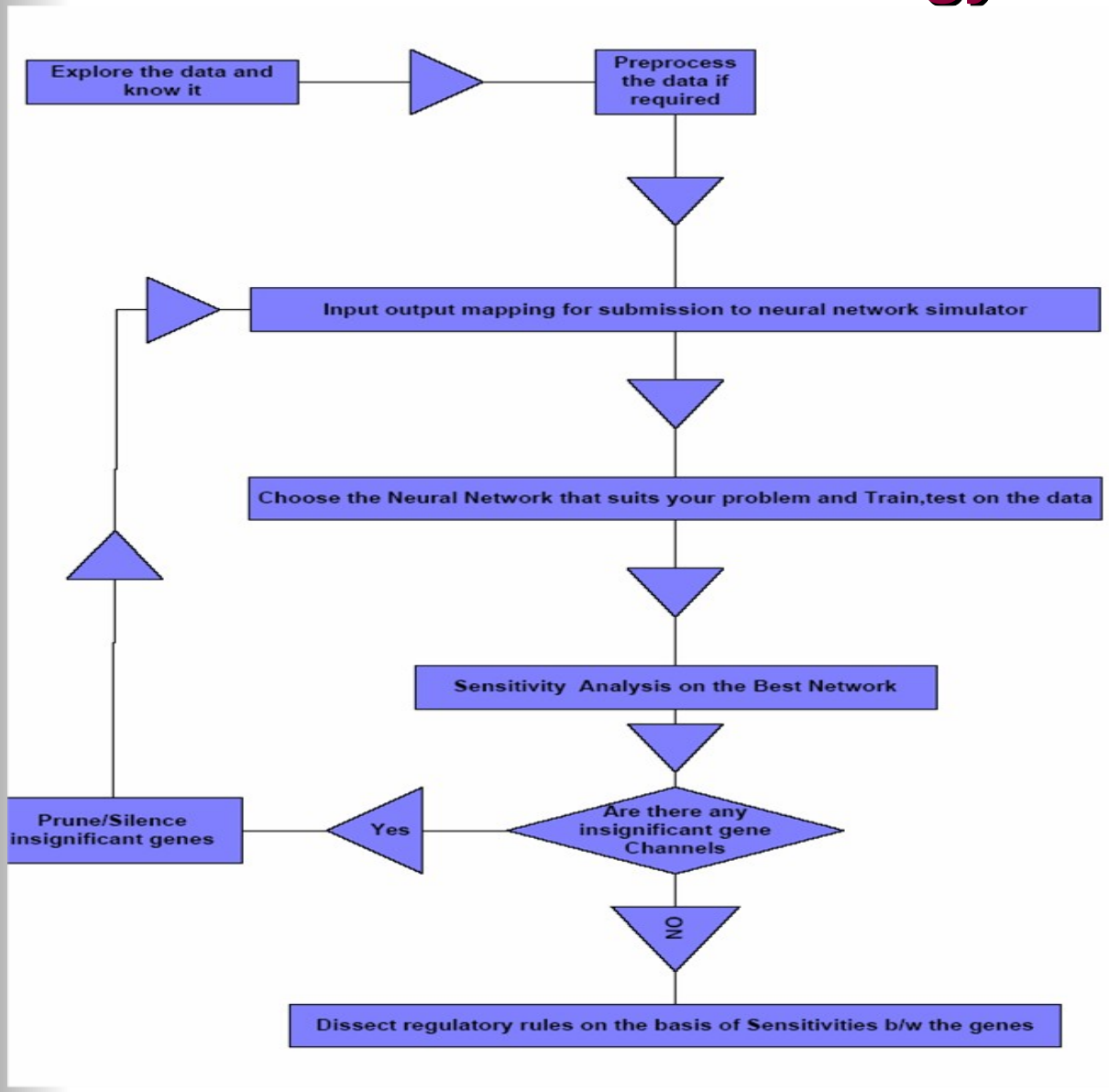
Rule	Arguments	Rule	Arguments
1	3→1	11	8→6
2	4 &10→1	12	5→6
3	6 &10→2	13	8 &7→7
4	4 &9→2	14	6 &10→7
5	5 &4⇒3	15	9 &3→8
6	9→3	16	4 &1→8
7	2→4	17	7→9
8	9→4	18	5 &4→9
9	6→5	19	5 &10→10
10	9 &6⇒5	20	1→10

Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

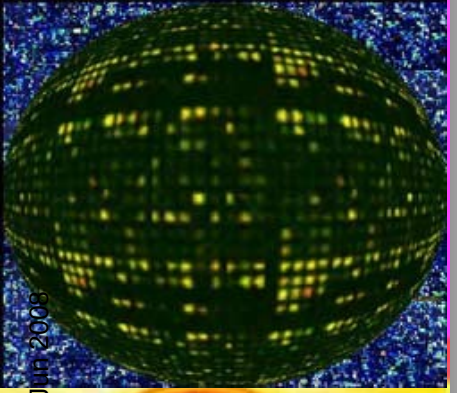


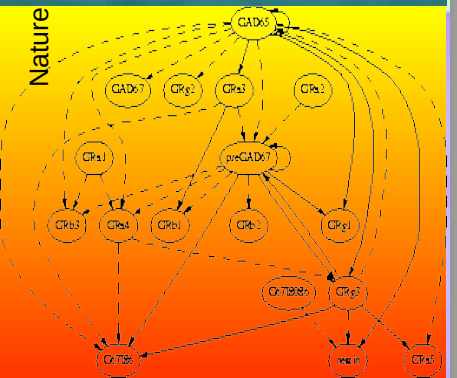


NNSBAGSS methodology



Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008





NNSBAGSS methodology : Rat C.N.S development time-course data

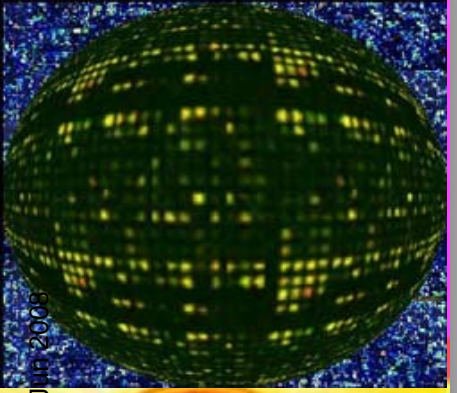
The Rat (Sprague-Dawley albino rat) data is an RT-PCR study of 112 genes each measured on cervical spinal cord tissue in triplicate at nine different time points during the development of the rat central nervous system [4]. This gene expression data is accepted to be non-noisy, small and accurate, and is ideal for testing a new strategy because of previous work in literature.

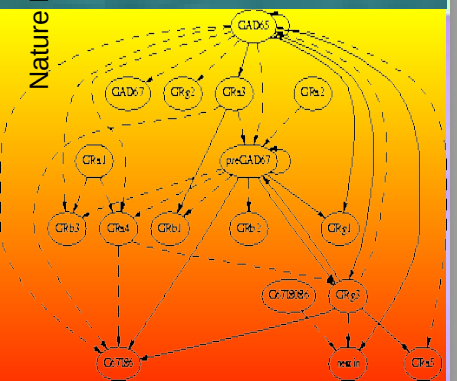
There are nine different time points in the CNS development study namely E11, E13, E15, E18, E21, P0, P7, P14, A. The input output mapping were E11 input-E13 output, E13 Input-E15 output, E15 input-E18 output, E18 input-E21 output, E21 input-P0 output, P0 input-P7 output, P7 input-P14 output, P14 input-A output.

There are eight exemplars for network to learn from. Each exemplar has 112 input gene expression values of the prior time step and 112 desired gene expression output values of the subsequent time step.

The training, testing and cross validation regime was followed to choose the best network. The sensitivity analysis was performed on the perceptron trained on all eight exemplar pairs (9 time steps).

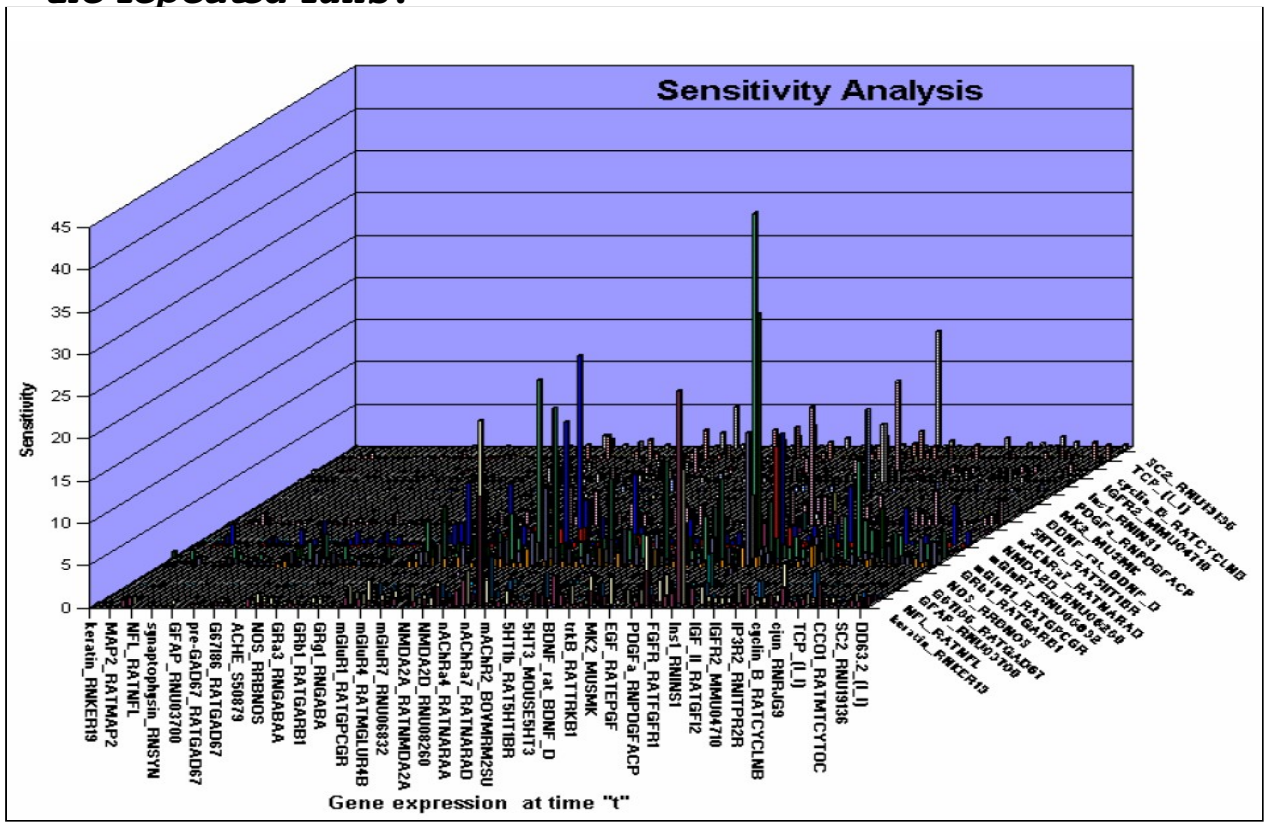
Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008



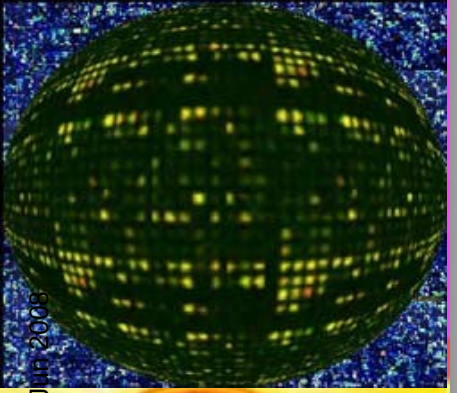


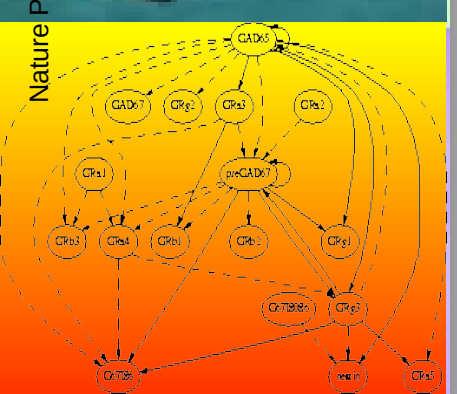
NNSBAGSS methodology : Rat C.N.S development time-course data

The graph in Figure below is the representation of sensitivity matrix showing the global sensitivity of all 112 genes at the previous time step affecting 112 genes at the next time step. Again these remain similar over the repeated runs.

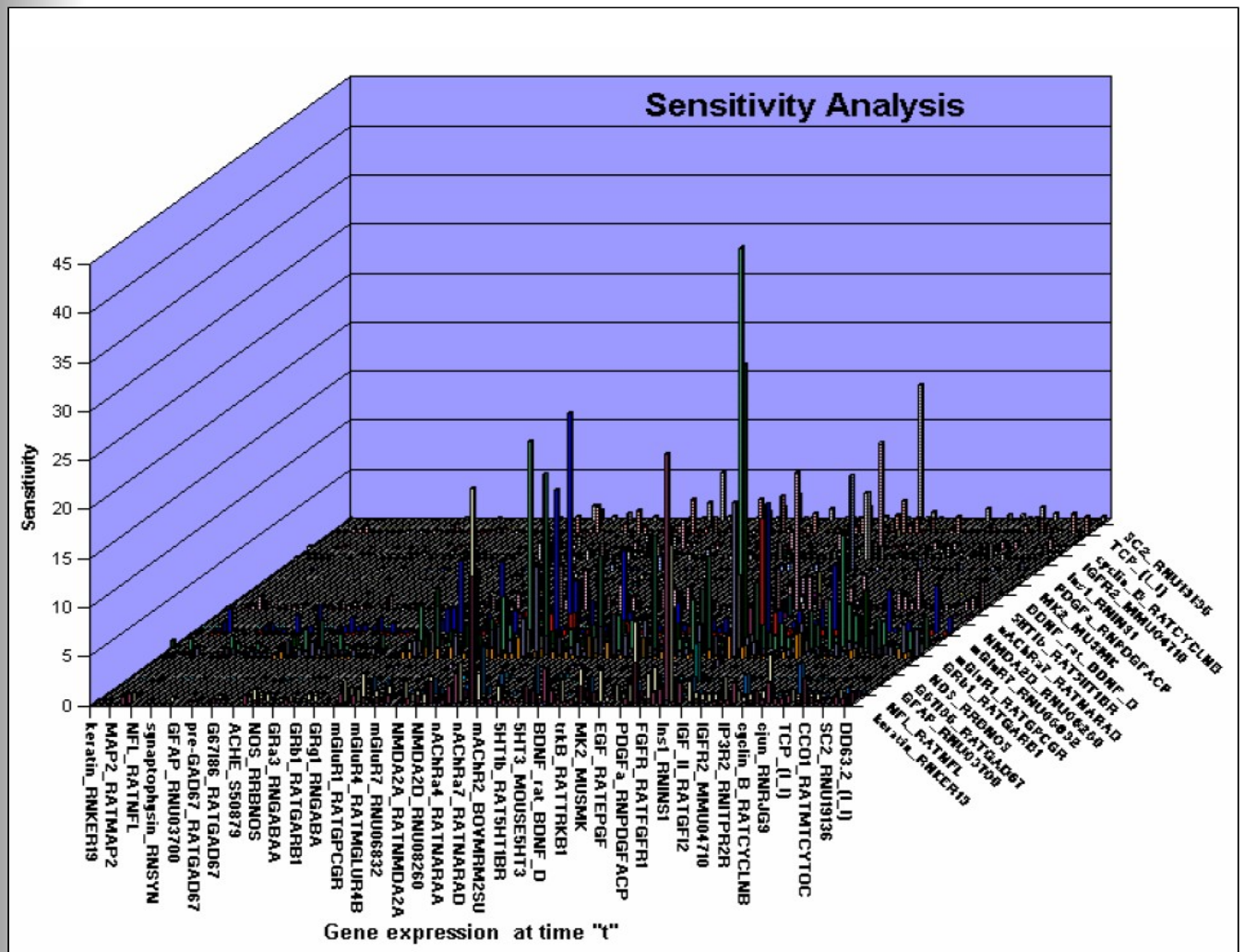


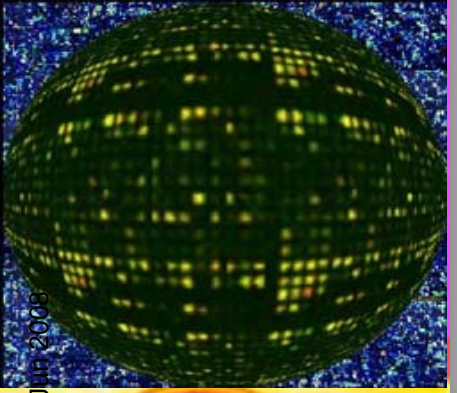
Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

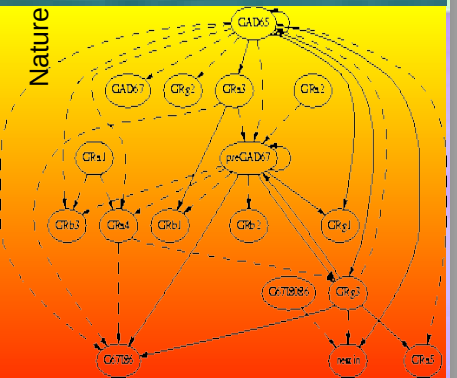




NNSBAGSS methodology : Rat C.N.S development time-course data







Nature Precedings : doi:10.1038/npre.2008.2003.1 : Posted 23 Jun 2008

Conclusions/Future directions

The reverse engineering of gene networks and causal relationships between genes is the major obstacle in extracting system knowledge from gene expression data. These experiments demonstrate that a neural network approach, combined with sensitivity analysis, can reverse engineer both actual and biologically plausible rules from artificial and real-world data, respectively. The results presented here provide evidence of a novel, alternative approach to reverse engineering that can lead to automatic extraction of rules from temporal gene expression data. The process is simple and repeatable, with clear visualisations resulting from the method.

The application of this methodology with more complex networks on large gene expression data sets is currently on-going and results of our experiments on one such data set (I've retalmentioned in the abstract), I will be presenting in Y B F on 20/10/2004 for the first time.