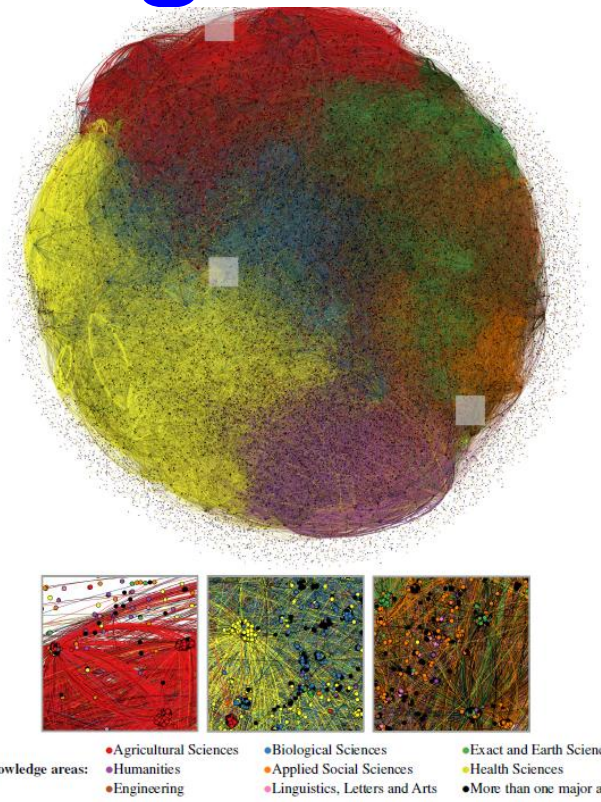
Enabling Multi-scale Science: Mathematical and Computational Modeling



Roberto M. Cesar-Jr

IME – USP
http://www.ime.usp.br/~cesar/

eScience-USP
FAPESP – CNPq - Capes

Summary

- Which “scale”?
- Multi-scale HOWTO
- Concluding remarks?

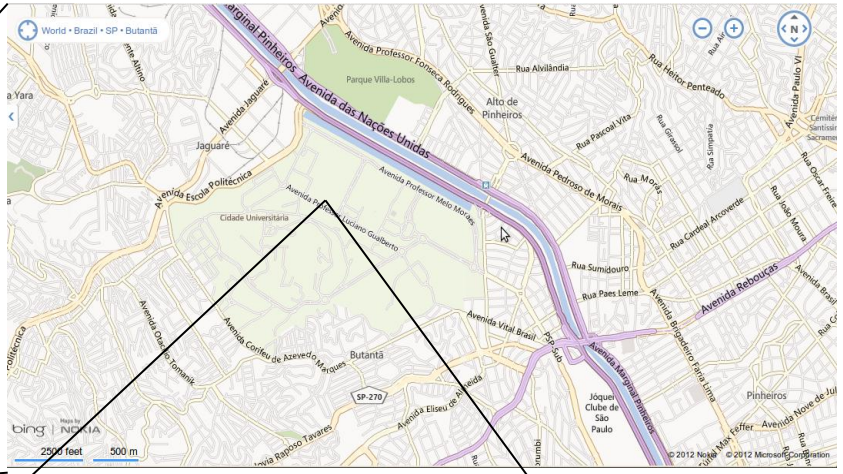
Which “scale”?

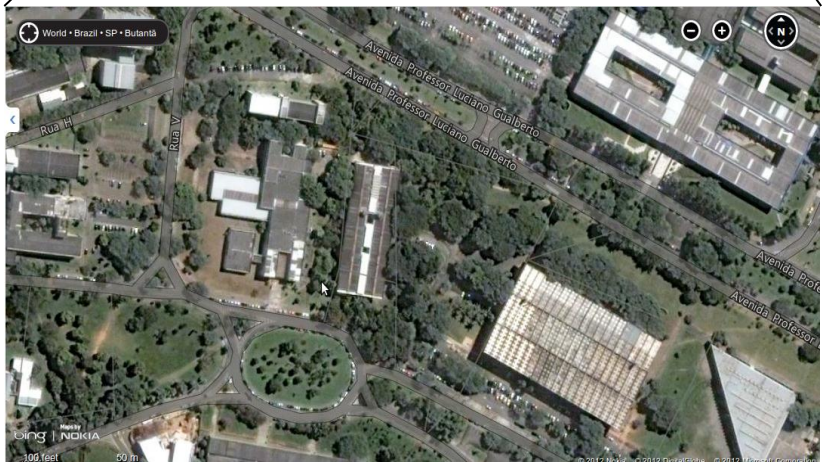
Which “scale”...

…are we talking about?

- Spatial
- Temporal
- Conceptual
- ?

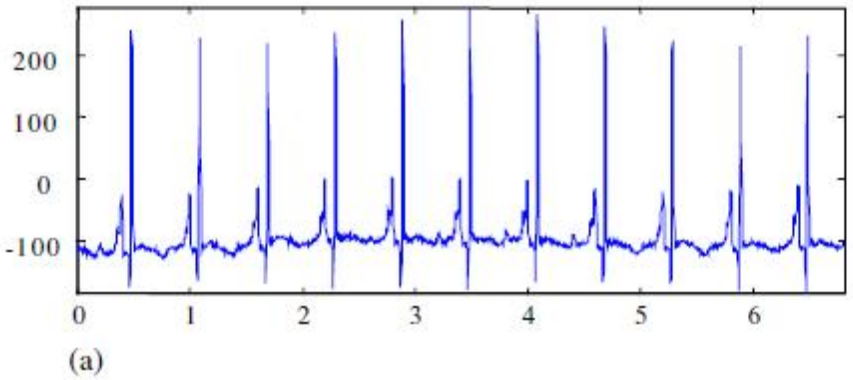
Which “scale”...

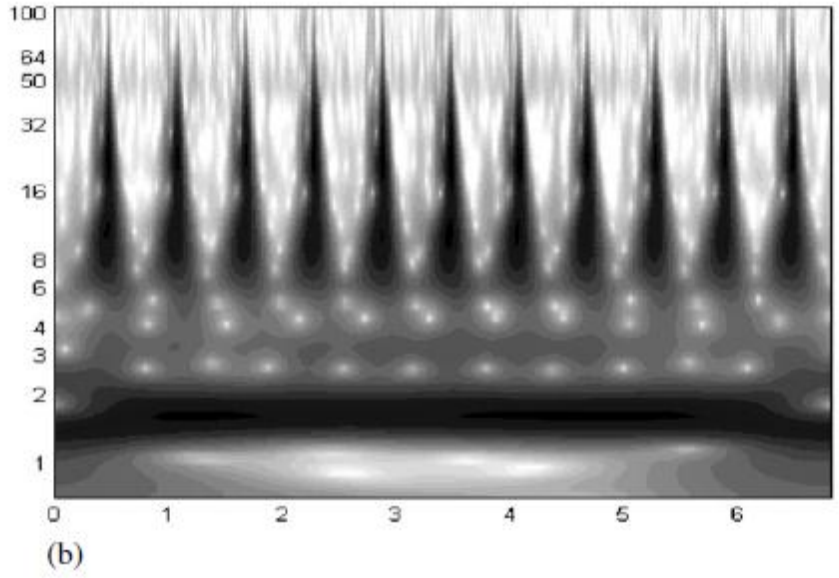




Spatial scales

Which “scale”...





Temporal scales

INSTITUTE OF PHYSICS PUBLISHING PHYSIOLOGICAL MEASUREMENT
Physiol. Meas. 26 (2005) R155–R199 doi:10.1088/0967-3334/26/5/R01

TOPICAL REVIEW

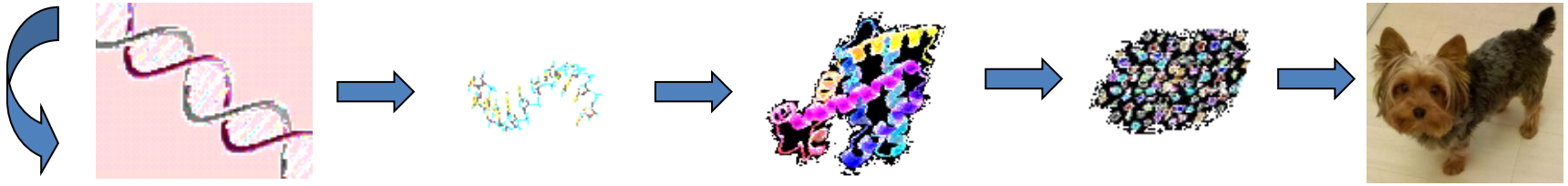
Wavelet transforms and the ECG: a review

Paul S Addison
CardioDigital Ltd, Elvingston Science Centre, East Lothian, EH33 1EH, UK
E-mail: p.addison@cardiodigital.com

Data-intensive knowledge discovery and systems biology: Conceptual scales

Central dogma of biology

Replication Transcription Translation



DNA mRNA Proteins Metabolites Organisms

Genomics Transcriptmics Proteomics Metabolomics Various analysis

Heterogeneous, complex, distributed databases

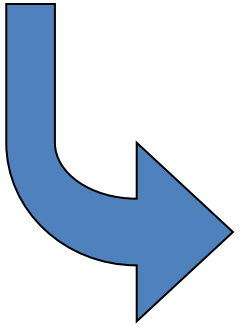
Data-intensive knowledge discovery and systems biology

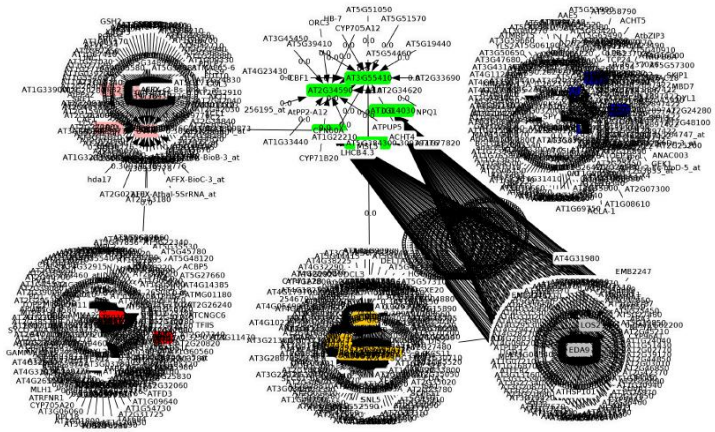
Central dogma of biology

Replication Transcription Translation

DNA mRNA Proteins Metabolites Organisms

Genomics Transcriptmics Proteomics Metabolomics





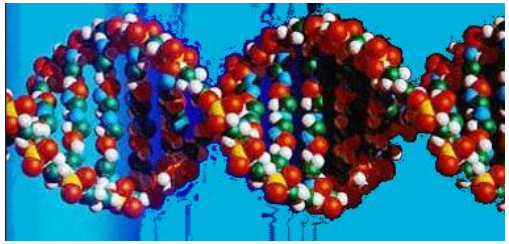
Second order conceptual scales

Multi-scale HOWTO

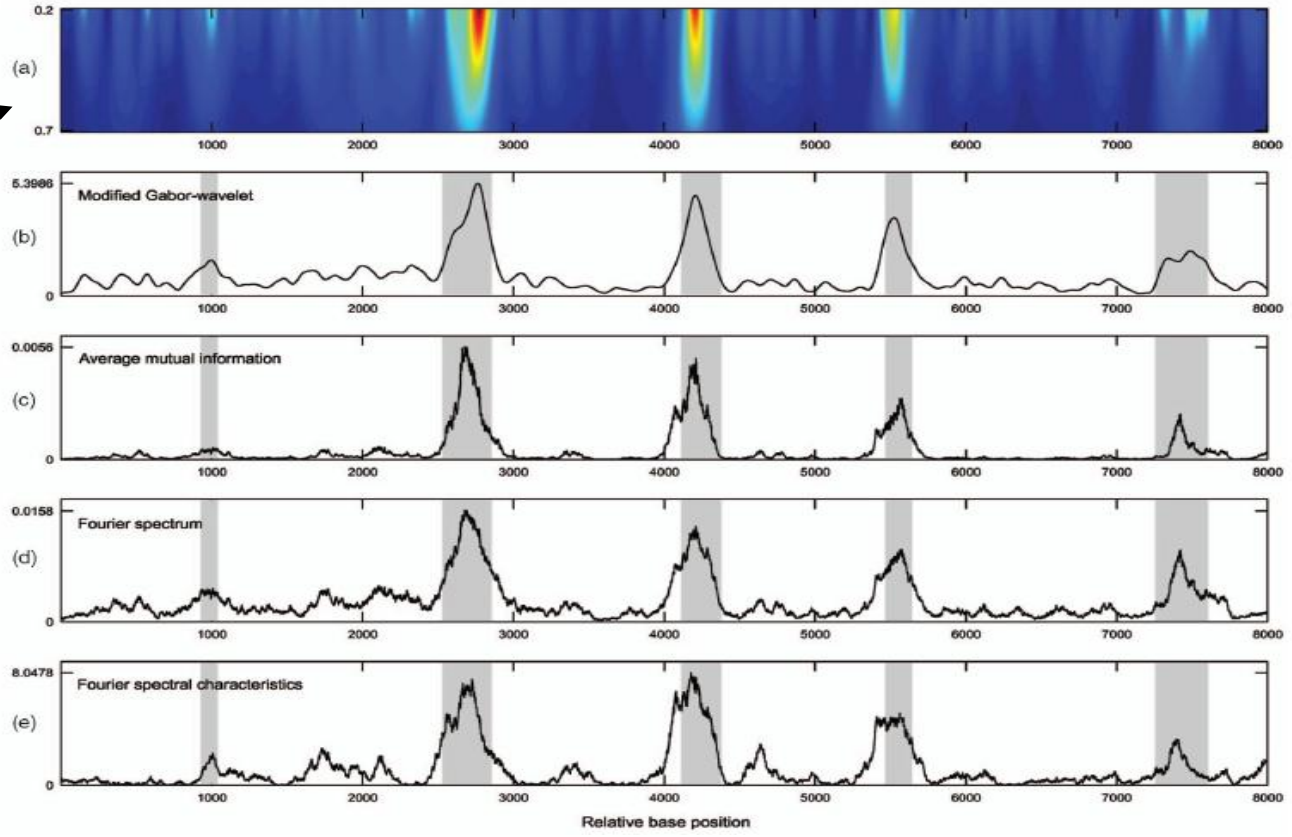
HOWTO

- Problems with or without inherent scales
- Mathematical approaches available (continuous, discrete, multidimensional):
  - Wavelets
  - Pyramids and Resolution
  - Graphs
  - ....
- Computational methods and structures

HOWTO



Information:
- coding
- unfolding
- folding



Identification of Protein Coding Regions Using the Modified Gabor-Wavelet Transform

Jesús P. Mena-Chalco, Helaine Carrer, Yossi Zana, and Roberto M. Cesar Jr.

HOWTO

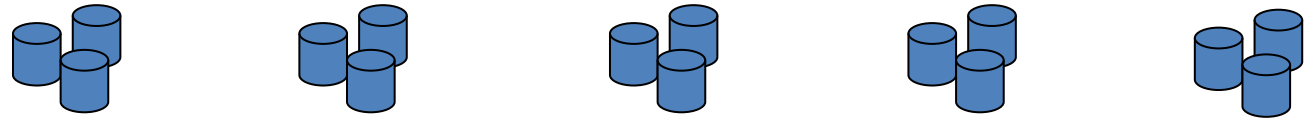
Data-intensive knowledge discovery and systems biology

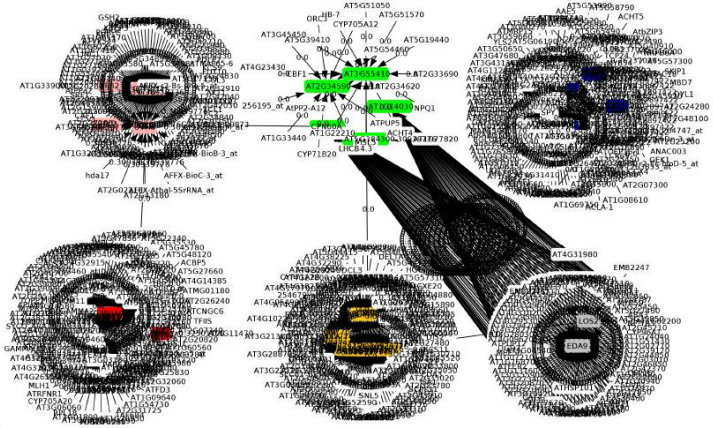
Central dogma of biology

Replication Transcription Translation

DNA mRNA Proteins Metabolites Organisms

Genomics Transcriptmics Proteomics Metabolomics





Second order conceptual scales

HOWTO

Expression of gene i at time t: $x_i[t] \in \{-1,0,+1\}$

State of the regulatory network at time t: $x[t] = \begin{bmatrix} x_1[t] \\ x_2[t] \\ . \\ . \\ x_n[t] \end{bmatrix}$

Network dynamics: $x[t+1] = \phi(x[t])$

HOWTO

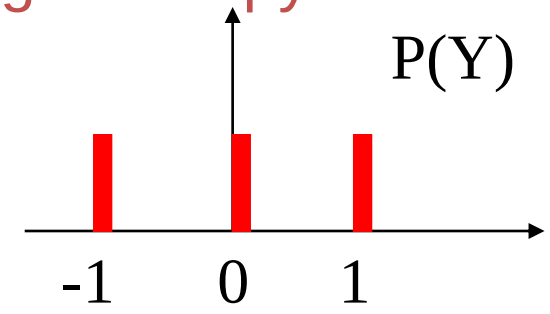
Entropy

$$H(Y) = -\sum_{y \in \{-1,0,1\}} P(y) \log P(y)$$

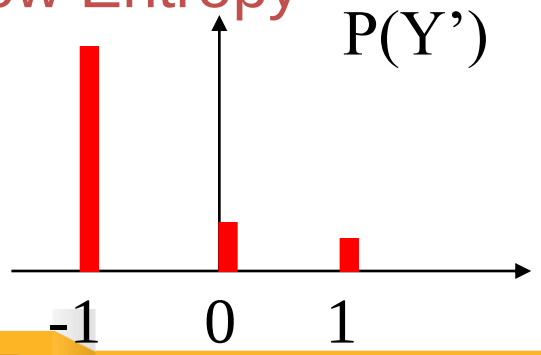
Mean conditional entropy

$$E[H(Y \mid X)] = -\sum P(X) \sum P(Y \mid X).\log(P(Y \mid X))$$

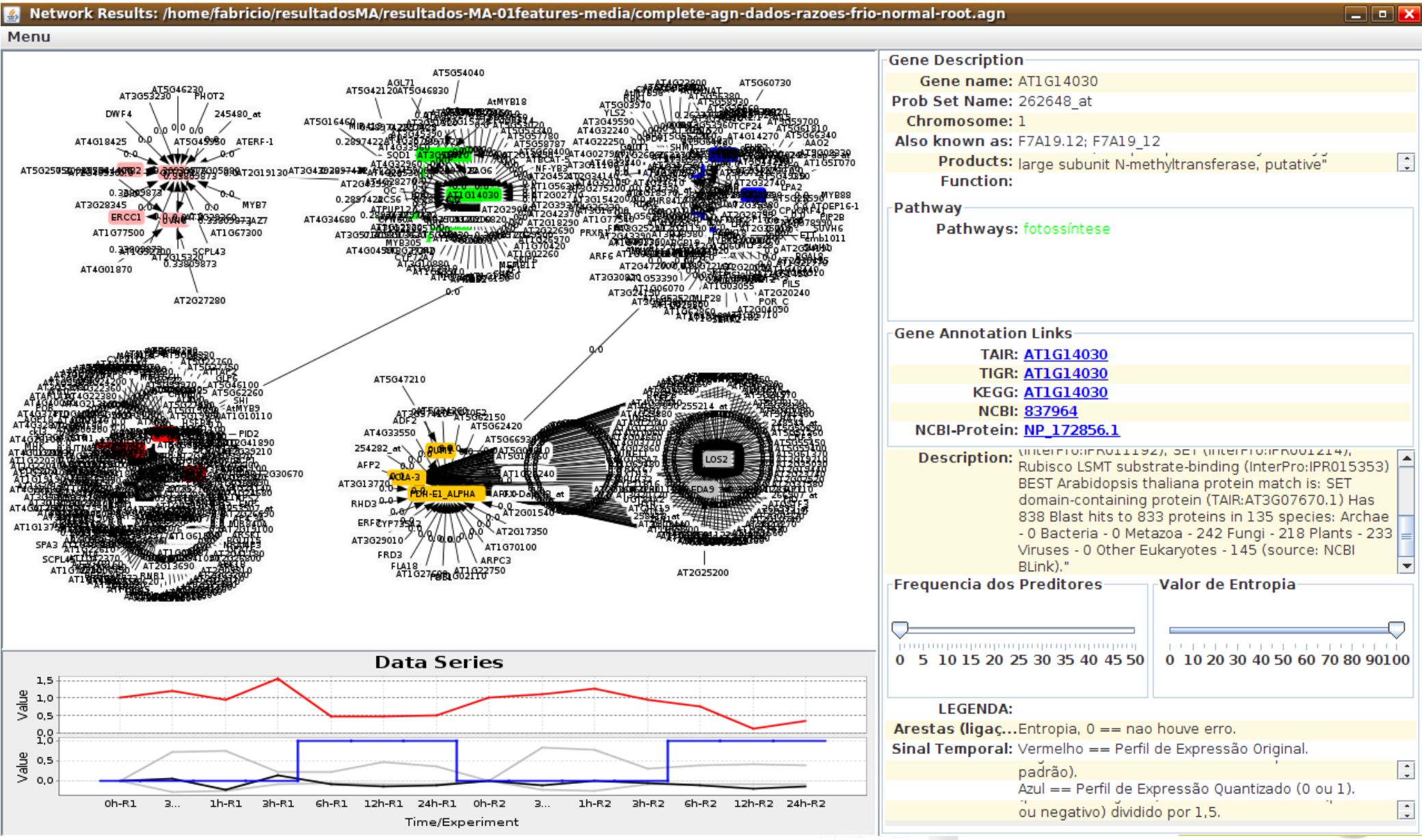
High Entropy



Low Entropy



HOWTO


Network Results: /home/fabricio/resultadosMA/resultados-MA-01features-media/complete-agn-dados-razoes-frio-normal-root.agn
Menu
AT1G14030
Gene Description
Gene name: AT1G14030
Prob Set Name: 262648_at
Chromosome: 1
Also known as: F7A19.12; F7A19_12
Products: large subunit N-methyltransferase, putative"
Function:
Pathway
Pathways: fotossíntese
Gene Annotation Links
TAIR: AT1G14030
TIGR: AT1G14030
KEGG: AT1G14030
NCBI: 837964
NCBI-Protein: NP_172856.1
Description: Rubisco LSMT substrate-binding (InterPro:IPR015353) BEST Arabidopsis thaliana protein match is: SET domain-containing protein (TAIR:AT3G07670.1) Has 838 Blast hits to 833 proteins in 135 species: Archae - 0 Bacteria - 0 Metazoa - 242 Fungi - 218 Plants - 233 Viruses - 0 Other Eukaryotes - 145 (source: NCBI BLink)."
Frequencia dos Preditores
0 5 10 15 20 25 30 35 40 45 50
Valor de Entropia
0 10 20 30 40 50 60 70 80 90100
LEGENDA:
Arestas (ligaç... Entropia, 0 == nao houve erro.
Sinal Temporal: Vermelho == Perfil de Expressão Original.
padrão).
Azul == Perfil de Expressão Quantizado (0 ou 1).
ou negativo) dividido por 1,5.
Data Series
Value
1,5
1,0
0,5
0,0
0h-R1
3...
1h-R1
3h-R1
6h-R1
12h-R1
24h-R1
0h-R2
3...
1h-R2
3h-R2
6h-R2
12h-R2
24h-R2
Time/Experiment

HOWTO

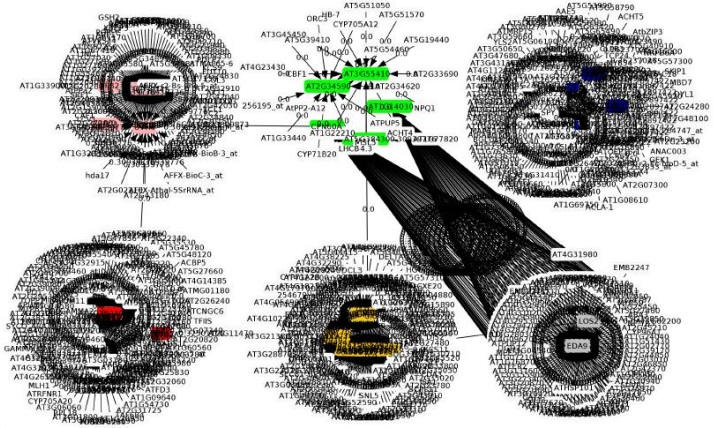
Data-intensive knowledge discovery and systems biology

Central dogma of biology

Replication Transcription Translation

DNA mRNA Proteins Metabolites Organisms

Genomics Transcriptmics Proteomics Metabolomics



Mappings between second order conceptual scales

Concluding remarks

HOWTO

- Different “types” of scales
- Problems with inherent scales
- Unfolding and folding the scales
- Mathematical models
- Computational models and data-structures
- Software

Multi-scale for eScience Approaches

- Multi-scale as a way to cope with complexity
- Multiple scales for exploratory data analysis, browing data, grasping initial understanding,...
- Data integration: mathematical and data models
- Network matching, searching, comparison, motifs