Organisation of RDF Representation

Storage and maintenance of meta-data are a basic requirement for presentation of semantic web resources. In order to efficiently maintain meta-data associated with instance data, a hierarchical ordering of RDF is adopted where provenance and meta-data is associated with sub-graphs ordered by some meaningful relationship.

Knowledge Discovery and Network Analysis

The known biochemical interaction network is very small to answer the biological questions. The complete network is therefore inferred using new Machine learning and stochastic methods by fusing networks reconstructed from heterogeneous data sources.

The networks are first inferred using Gene Expression data, Protein Localization and Phylogenetic profiles datasets and then integrated together to form the final network.

The network produced through these “knowledge discovery” methods are made available by “knowledge representation”. The network can be evaluated for its usefulness in an array of already developed network analysis techniques such as, Markov Clustering and Gene Rank, as well as for development of new methods computationally utilizing information about relationships between genes (and their products) to analyze data and generate hypothesis relevant to biomedical research.

BioMANTA Queries

The aim of the BioMANTA project is to create a semantic web representation of biomedical and biomolecular data that can be queried to uncover links between biological systems, networks and disease. The following example illustrates the types of queries that the BioMANTA framework will address:

A recent study (Ray et al.) demonstrated that blood plasma levels of 18 signaling proteins can be used to predict Alzheimer’s progression in presymptomatic patients.

• How are these biomarkers linked to Alzheimer’s Disease?
• What signaling pathways and biological functions are these proteins associated with?
• Are these genes or pathways significantly linked to knockout mouse phenotypes related to memory or cognition?
• Can a causal link between these pathways and Alzheimer’s be hypothesized?
• Is this gene set significantly linked to other diseases?
• Are there existing therapies for these diseases that could be viable Alzheimer’s treatments?

BioMANTA Output and Visualization

After executing queries across the BioMANTA resource, one significant challenge will be to visualize the returned results.

One key challenge will be visualising data at a higher resolution than gene level. While network analysis is frequently carried out at this level of resolution, the recent discovery of extensive variability in the structure and function of the protein products of genes indicates that this diversity must be accommodated in future network analysis. This additional complexity, along with data from tissue specific and time-course expression profiling generates challenges for the task of visualising the output of queries over the semantic interactome.