Modelling and comparing protein interaction networks using subgraph counts

Charlotte M. Deane*
Department of Statistics, University of Oxford, Oxford, UK deane@stats.ox.ac.uk
Tiago Rito
Department of Statistics, University of Oxford, Oxford, UK rito@stats.ox.ac.uk
Gesine Reinert
Department of Statistics, University of Oxford, Oxford, UK reinert@stats.ox.ac.uk

The astonishing progress of molecular biology, engineering and computer science has resulted in mature technologies capable of examining multiple cellular components at a genome-wide scale. Protein-protein interactions are one example of such growing data. These data are often organised as networks with proteins as nodes and interactions as edges. Albeit still incomplete, there is now a substantial amount of data available and there is a need for biologically meaningful methods to analyse and interpret these interactions.
This presentation will focus on the relationship between network architecture and the biological characteristics of proteins. We will demonstrate how the content of small subgraphs (small interaction patterns between 2-5 proteins) can be used to compare networks and bring to light biological and evolutionary information.
In one case study we consider the protein age patterns found in the edge and triangle subgraphs of the yeast protein interaction network. We assess their statistical significance both according to what would be expected by chance given the node frequencies found in the yeast network, and also, for the case of triangles, given the age frequencies observed in the currently available pairwise data. We find that pairwise interactions between Old proteins are over-represented even when controlling for high degree, and triangle interactions between Old proteins are over-represented even when controlling for pairwise interaction frequencies. There is evidence for negative selection of interactions between Middle-aged and Old proteins within triangles, despite pairwise Middle-Old interactions being common. Most triangles consist solely of vertices with high degree.
Our analyses show that protein interaction networks are highly heterogeneous and therefore no single model is likely to fit the network and that there are specific structural and functional patterns within subgraphs.

Key Words: ego networks, protein function, protein structure