

# SEMS: RESEARCH PROJECT DESCRIPTION

## 1. Project Background and Description

### Soft Computing Techniques for Interpretable and Reliable Gene Network Inference

Genes do not function in isolation but form highly interconnected networks of >100,000 interaction. These networks display a special structure, containing highly interacting genes ('hubs') and the nodes showed a power law distribution. However, the best network inference software ('random forest') shows a high precision when applied to small network with many replicates but returns low success rates in larger genomes with only a few replicates, a situation more common in practice. When noise is added to the genes expression data the success rate further decreases.

In an attempt to improve on this low precision, recent studies employed soft computing techniques, e.g. Artificial Neural Networks, Fuzzy Cognitive Maps, Fuzzy Inference Systems (FISs), Evolutionary Algorithms (EAs) and their hybridisations. Although these approaches improved the inference accuracy for small to medium networks due to their nonlinear approximation capability, they generally suffer from the following problems.

- (1) They could not handle large-scale networks.
- (2) They require excessive computational time.
- (3) The inferred results were less interpretable due to the employed optimisation mechanisms.
- (4) It was difficult to incorporate prior/existing knowledge into these data-driven approaches.

In light of the above problems, this project will extend on this line of thinking using advanced soft computing techniques (e.g. multi-objective optimisation and clustering algorithms, interpretable FISs and transfer learning) to effectively integrate multi-omics data, decompose the problem space and elicit FISs to achieve interpretable and reliable gene network inference.

## 2. Project Scope

**Objective 1:** Develop an effective data integration mechanism based on transfer learning to integrate multi-omics data.

**Objective 2:** Develop a multi-objective fuzzy clustering algorithm to decompose the problem space. Based on the reduced search space, interpretable FISs will be developed taking into account the inference accuracy and comprehensibility as the two objectives at the same time. Known gene interactions (prior knowledge) will be converted into fuzzy rules and integrated into FISs.

**Objective 3:** Evaluate the developed approach in terms of the inference accuracy, computational time and comprehensibility, and compare it with the notable baseline gene network inference algorithms.

## 3. Desired Skills from the Student

The student needs to have experience in computing, statistics, and preferably bioinformatics.

## 4. Supervisory Team

**Primary:** Dr Jun Chen, Senior Lecturer in Engineering Science.

**Secondary:** Prof Rob Krams, Professor in Molecular Bioengineering.