**Statistical learning of epistasis, genetic maps and microbial networks via graphical models**

Prof. Ernst C. Wit

Professor of Data Science

Universita della Svizzera italiana

Institute of Computational Science



**Abstract**

High-throughput genomic data in the field of genomics has revolutionized the research field. It is able to penetrate deeply into the functional, evolutionary and even social environment of the species under study. Often, however, existing data analytic methods are unable to capture the intricate nature of these processes.

In this presentation we present a single methodological tool, called Copula Graphical Models, that are able to capture the functional, evolutionary and social interactions of organisms through high-throughput data. We present three examples:

1. Recombinant Inbred Lines (RILs) derived from divergent parental lines can display extensive segregation distortion and long-range linkage disequilibrium (LD) between distant loci on same or different chromosomes. These genomic signatures are consistent with epistatic selection having acted on entire networks of interacting parental alleles during inbreeding. The reconstruction of these interaction networks from observations of pair-wise marker-marker correlations or pair-wise genotype frequency distortions is challenging as multiple testing approaches are under-powered and true long-range LD is difficult to distinguish from drift, particularly in small RIL panels.
2. A side product of the previous study showed that the same method also clearly detected a genetic map signature. Especially for challenging polyploid organisms we have used this method to detect the genetic map.
3. Microorganisms often live in symbiotic relationship with their environment and they play a central role in many biological processes. They form a complex system of interacting species. Until now correlation-based network analysis and graphical modelling have been used to identify the putative interaction networks formed by the species of microorganisms, but these methods do not take into account all features of microbiota data.

The methods have been implemented in two R packages (netgwas and rMAGMA), available to the practitioner. This is collaborative work with Pariya Behrouzi and Arnaud Cougoul.