

Data: Martedì 8 luglio 2025, ore 14,00-15,00 – su Teams

Speaker: Thomas Bartlett, University College London

Title: A stochastic networks approach to modelling genomic count data

Abstract:

Important tasks in the study of genomic data include the identification of groups of similar cells (for example by clustering), and visualisation of data summaries (for example by dimensional reduction). In this talk, I will present a novel view of these tasks in the context of single-cell genomic data. To do so, I propose modelling the observed count-matrices of genomic data by representing these measurements as a bipartite network with multi-edges. Starting with this first-principles network model of the raw data, I will show improvements in clustering single cells via a suitably-identified d-dimensional Laplacian Eigenspace (LE) using a Gaussian mixture model (GMM-LE), and apply UMAP to non-linearly project the LE to two dimensions for visualisation (UMAP-LE). From this first-principles viewpoint, the LE representation of the data-points estimates transformed latent positions (of genes and cells), under a latent position statistical model of nodes in a bipartite stochastic network. By applying this proposed methodology to data from three recent genomics studies in different biological contexts, I will show how clusters of cells independently learned by this proposed methodology are found to correspond to cells expressing specific marker genes that were independently defined by domain experts, with an accuracy that is competitive with the industry-standard for these data. I will then show how this novel view of these data can provide unique insights, leading to the identification of a LE breast-cancer biomarker that significantly predicts long-term patient survival outcome in two independent validation cohorts with data from 1904 and 1091 individuals.

Si può seguire online al link

<https://u.garr.it/j00OT>





Scansiona per l'elenco completo
dei Convegni e Seminari