AI-Powered Bayesian Methods for Analyzing Spatial Omics Data

Qiwei Li[1]

Department of Mathematical Sciences, The University of Texas at Dallas

ABSTRACT

Recent technology breakthroughs in gene expression profiling have enabled the comprehensive molecular characterization of single cells while preserving their spatial and morphological contexts. This new bioinformatics scenario advances our understanding of molecular and cellular spatial organizations in tissues, fuelling the next generation of scientific discovery. Bayesian statistics relies more on human analyses with computer aids, while AI relies more on computer algorithms with aids from humans. This talk will outline methodologies for bridging AI capabilities with Bayesian frameworks, aiming to resolve key issues in spatially resolved transcriptomics (SRT) data analysis. Particularly, I will focus on two major problems: spatial domain identification and gene expression reconstruction.

Current clustering analysis of SRT data primarily relies on molecular information and fails to fully exploit the morphological features present in histopathology images, leading to compromised accuracy and interpretability. To overcome these limitations, we have developed a multi-stage statistical method called iIMPACT. It identifies and defines histology-based spatial domains based on AI-reconstructed histopathology images and spatial context of gene expression measurements, and detects domain-specific differentially expressed genes. Through multiple case studies, we demonstrate iIMPACT outperforms existing methods in accuracy and interpretability and provides insights into the cellular spatial organization and landscape of functional genes within spatial transcriptomics data.

Most next-generation sequencing-based SRT techniques are limited to measuring gene expression in a confined array of spots, capturing only a fraction of the spatial domain. Typically, these spots encompass gene expression from a few to hundreds of cells, underscoring a critical need for more detailed, single-cell resolution SRT data to enhance our understanding of biological functions within the tissue context. Addressing this challenge, we introduce BayesDeep, a novel Bayesian hierarchical model that leverages cellular morphological data from histology images, commonly paired with SRT data, to reconstruct SRT data at the single-cell resolution. BayesDeep effectively models count data from SRT studies via a negative binomial regression model. This model incorporates explanatory variables such as cell types and nuclei-shape information for each cell extracted from the paired histology image. A feature selection scheme is integrated to examine the association between the morphological and

molecular profiles, thereby improving the model robustness. We applied BayesDeep to two real SRT datasets, successfully demonstrating its capability to reconstruct SRT data at the single-cell resolution. This advancement not only yields new biological insights but also significantly enhances various downstream analyses, such as pseudotime and cell-cell communication.

Keywords: Spatial transcriptomics; Spatial clustering; Shape analysis; Deep learning

Back to Sessions List

Covariate-Assisted Graph Matching

Trisha Dawn, **Jesús Arroyo**

Department of Statistics, Texas A&M University

ABSTRACT

Data integration is essential across diverse domains, from historical records to biomedical research, facilitating joint statistical inference. A crucial initial step in this process involves merging multiple data sources based on matching individual records, often in the absence of unique identifiers. When the datasets are network data -- a flexible and increasingly prevalent structure representing entities as vertices and their relationships as edges -- this problem is typically addressed through graph matching methodologies. For such cases, auxiliary features or covariates associated with nodes or edges can be instrumental in achieving improved accuracy. However, most existing graph matching techniques do not incorporate this information, limiting their performance against non-identifiable and erroneous matches. To overcome these limitations, we propose two novel covariate-assisted seeded graph matching methods, where a partial alignment for a set of nodes, called seeds, is known. The first one utilizes the quadratic assignment problem (QAP), while the second one leverages the local neighborhood structure of non-seed nodes to guide the matching process. Both methods are grounded in a conditional modeling framework, where elements of one graph's adjacency matrix are modeled using a generalized linear model (GLM), given the other graph and the available covariates. We establish theoretical guarantees for model estimation error and exact recovery of the solution of the QAP, demonstrating perfect alignment accuracy with high probability under sufficient signal strength. The effectiveness of our methods is demonstrated through numerical experiments to accommodate varied parameter settings and number of seeds. Finally, we apply our proposed approach to match two real-world network data. Our work highlights the power of integrating covariate information in the classical graph matching setup, offering a practical and improved framework for combining network data with wide-ranging applications.

Keywords: Network data; node and edge covariates; optimization problem; generalized linear model

Abhishek Roy

Back to Sessions List

Seeing Is Believing: Challenges and Opportunities for Super-Resolution Microscopy Image Data Analysis for Quantitative Molecular Biology

Chongzhi Zang
University of Virginia

ABSTRACT

Advanced technologies have been the driving force behind modern biomedical research, with rigorous data analysis serving as the critical link between technological applications and scientific discoveries. While high-throughput sequencing-based technologies have advanced to the single-cell level with spatial information, they still lack the ability to directly measure the dynamic structures and activities of protein molecules in the cell nucleus. Super-resolution microscopy technologies, such as three-dimensional structured illumination microscopy (3D-SIM) and 3D stochastic optical reconstruction microscopy (STORM), have been increasingly applied in molecular biology research, as they enable the direct detection of 3D structures of chromatin elements at the single-molecule resolution. However, data analysis remains a major challenge in this field due to the noisy, sparse, and highly variable nature of microscopy image data across samples, whether in replicates or under different biological conditions. In this talk, I will introduce super-resolution microscopy technologies, discuss the unique characteristics of the big data and challenges of the analysis, and emphasize the need for innovative and rigorous statistical methods to better understand these image data to elucidate molecular biology mechanisms.

Keywords: Image analysis, super-resolution microscopy, bioinformatics