

A LOCALLY ADAPTIVE ALGORITHM FOR MULTIPLE TESTING WITH NETWORK STRUCTURE

Ziyi Liang, T. Tony Cai, Wenguang Sun, and Yin Xia*

*University of California, University of Pennsylvania,
Zhejiang University and Fudan University*

Abstract: Incorporating auxiliary information alongside primary data can significantly enhance the accuracy of simultaneous inference. However, existing multiple testing methods face challenges in efficiently incorporating complex side information, especially when it differs in dimension or structure from the primary data, such as network side information. This paper introduces a locally adaptive structure learning algorithm (LASLA), a flexible framework designed to integrate a broad range of auxiliary information into the inference process. Although LASLA is specifically motivated by the challenges posed by network-structured data, it also proves highly effective with other types of side information, such as spatial locations and multiple auxiliary sequences. LASLA employs a p -value weighting approach, leveraging structural insights to derive data-driven weights that prioritize the importance of different hypotheses. Our theoretical analysis demonstrates that LASLA asymptotically controls the false discovery rate (FDR) under independent or weakly dependent p -values, and achieves enhanced power in scenarios where the auxiliary data provides valuable side information. Simulation studies are conducted to evaluate LASLA's numerical performance, and its efficacy is further illustrated through two real-world applications.

Key words and phrases: Covariate-assisted inference, distance matrix, false discovery rate, p -value weighting, structure learning.

1. Introduction

1.1. Motivating application for network data

Statistical analysis of network-structured data is an important topic with a wide range of applications. Our study is motivated by genome-wide association studies (GWAS), where a primary objective is to identify disease-associated single-nucleotide polymorphisms (SNPs) across diverse populations. Previous studies have indicated that linkage analysis can provide insights into the genetic basis of complex diseases. Particularly, SNPs in linkage disequilibrium (LD) can jointly contribute to the representation of the disease phenotype (Schaub et al., 2012; Joiret et al., 2019). However, existing research in GWAS has often overlooked or underutilized the LD network information, representing

*Corresponding author. E-mail: xiayin@fudan.edu.cn