Statistica Sinica **35** (2025), 1-20
doi:https://doi.org/10.5705/ss.202023.0298

INFERRING HUB NODES ON DIFFERENTIAL GAUSSIAN GRAPHICAL MODELS

Xin Zhou, Kean Ming Tan and Junwei Lu*

University of California, Berkeley, University of Michigan and Harvard Chan School of Public Health

Abstract: Identifying changes between two networks, also referred to as differential network analysis, has brought new insights to many biological applications. A lot of progress has been made in the development of statistical inference tools for detecting changes between two networks, with most work focused on testing whether two networks are exactly the same, or whether there is an edge that is missing in one network but present in another. However, in many scientific settings, it is often more interesting to identify nodes that have different conditional dependency structures between two networks, which we refer to as differential hub nodes. In this paper, we propose an inferential framework to test whether there is at least one differential hub node in a differential Gaussian graphical model. As a by-product, our proposed test statistic can also be used to test the hypothesis on whether there is a differential edge and construct a confidence interval for the corresponding differential edge. Theoretically, we show that the proposed method yields an asymptotic valid test and that the type II error decreases to zero asymptotically. The proposed method is applied to both simulated data and the Genotype-Tissue Expression (GTEx) data to evaluate whether gene regulatory networks between males and females for different tissues are different.

Key words and phrases: Differential network, Gaussian multiplier bootstrap, hypothesis testing, maximum degree.

1. Introduction

Undirected graphical models have been used extensively for modeling conditional dependence relationships among a set of random variables in many scientific domains (Markowetz and Spang, 2007; Rubinov and Sporns, 2010). An undirected graph consists of d nodes and a set of edges: each node represents a random variable, and an edge between two nodes indicates that the corresponding two random variables are conditionally dependent, conditioned on all the other variables. Given a number of independent and identically distributed random samples, many methods were proposed to estimate a sparse undirected graphical model under various assumptions on the random variables (Meinshausen and Bühlmann, 2006; Friedman, Hastie and Tibshirani, 2008; Cai, Liu and Luo, 2011; Lee and Hastie, 2015; Tan et al., 2016; Yang, Ning and Liu, 2018). We refer the

*Corresponding author. E-mail: junweilu@hsph.harvard.edu