

# Identifying the Influential Variable using Centrality Measure: A Case of Multivariate Time Series

Nur Syahidah Yusoff<sup>1</sup>, Shamshuritawati Sharif<sup>2</sup>

<sup>1</sup> Faculty of Industrial Sciences & Technology  
Universiti Malaysia Pahang, 26300, Gambang, Pahang, Malaysia

<sup>2</sup> School of Quantitative Sciences  
UUM -College Arts and Science  
Universiti Utara Malaysia, Sintok, Kedah, Malaysia

Analysing the structure of multivariate system has been an important part in reliability analysis especially in identifying the influential variables. The complexity of the analysis increases when high dimensional data involved. To simplify the information in multivariate system, a network topology which is based on an Escoufier's RV-coefficient is constructed and centrality measure will be used to interpret the network. Statistically, RV-coefficient is a multivariate generalization of the squared Pearson correlation coefficient. An example in finance industry will be discussed to illustrate the structure of network topology and a recommendation will be presented.

Keywords: RV-coefficient, Network Topology, Minimum Spanning Tree.

## 1. INTRODUCTION

Financial industry incorporates a wide range of money management organizations such as banks, credit card companies, insurance companies, stocks market, real estate funds, currency exchange rates, and others. In recent years, it has become a common feature in financial industry to have more than one high-dimensional data set.

In this paper, we focus on the currency exchange rate data set. This data are the collection of data attained from large number of currencies where every currency will have four different prices per day, i.e, closing, opening, highest and lowest prices. Consequently, currency exchange rate data set is now considered to be a multivariate time series of those four prices but not a univariate time series of one price only<sup>1,2</sup>. The multivariate

other. Furthermore, a research work by Liangyue<sup>3</sup> found multivariate time series gives good prediction. This consideration should be substantial advantages in modeling and prediction when using multivariate time series.

In the literature, in the case of multivariate time series, there are two common strategies to measure the association between set of variables which are Canonical Correlation Analysis (CCA) and RV-coefficient<sup>4</sup>. The first approach is proposed by Hotelling<sup>5</sup> in year 1936. It seeks a linear combination of one set of variables that maximally correlated with linear function of the other set of variables. The second approach is RV-coefficient. This approach is put forward by Escoufier<sup>6</sup> in year 1973 to measure of similarity between positive semidefinite matrices<sup>7</sup>. The coefficients are similar to the