

# Complex Systems and Networks HW 1

Rachael Judy, Connor Klein, Josh Smith

February 15, 2024

## 1 Q1: Annotated Bibliography

### 1.1 A network perspective of the stock market

C. K. Tse, J. Liu, and F. C. M. Lau, "A network perspective of the stock market," *Journal of Empirical Finance*, vol. 17, no. 4, pp. 659–667, Sep. 2010, doi: 10.1016/j.jempfin.2010.04.008.

This paper presents network constructions of US stocks and proposes a new methodology for computing stock indices. The nodes of the networks consist of stocks, and the edges are established with a winner-take-all approach based on a threshold for cross correlation of daily stock prices, price returns, and trading volumes. The networks display a power-law distribution and suggest that the majority of stocks' time indices are strongly influenced by a small subset of stocks, particularly in the financial sector. They propose a new approach of using the networks for selection of stocks to represent the market indices. The paper highlights some flaws that may be present in the current methodology of determining stock indices and expands the importance of interconnectivity in examining the stock market. However, their methods are not examined over enough data to generalize the approach; the two windows examined by the authors are each two years in length and centered around the financial crisis of 2007-2008 which may not represent the overall market over time. Further, it does not examine the impact of adjusting several somewhat arbitrary parameters such as the threshold for creating an edge.

### 1.2 Pearson Correlation Coefficient-Based Performance Enhancement of Broad Learning System for Stock Price Prediction

G. Li, A. Zhang, Q. Zhang, D. Wu, and C. Zhan, "Pearson Correlation Coefficient-Based Performance Enhancement of Broad Learning System for Stock Price Prediction," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 5, pp. 2413–2417, May 2022, doi: 10.1109/TCSII.2022.3160266.

In this paper, the Pearson Correlation Coefficient (PCC), a measure of strength and direction of the relationship between two variables, is used with a Broad Learning System (BLS) for feature selection to perform time series prediction of stocks selected from the

Shanghai Stock Exchange. The authors compared the results of their method with ten machine learning methods including Adaboost, Gradient Boosting Decision Tree, Convolutional Neural Networks, and others. The BLS system is based on a random vector functional link neural network that transforms input features to feature nodes that connect to the output layer. Weights are initially randomly generated and then improved through pseudoinversion. This provides a less time intensive alternative to a deep learning framework. The paper provides a detailed explanation of the features and equations used and provides metrics showing its superior performance over other prediction methods. The main deficiencies evident in the paper are the assumptions that the small sample size of selected stocks can demonstrate the system's performance overall; these selections are regional and do not necessarily represent the overall market. The paper does state that future work will involve creating a strategy and evaluating the model through returns.

### **1.3 System abnormality detection in stock market complex trading systems using machine learning techniques**

P. A. Samarakoon and D. A. S. Athukorala, "System abnormality detection in stock market complex trading systems using machine learning techniques," in 2017 National Information Technology Conference (NITC), Sep. 2017, pp. 125–130. doi: 10.1109/NITC.2017.8285660.

Using several different supervised learning approaches, this paper explores detecting faults and anomalous behaviors in stock market systems. These methods can reduce the human domain knowledge necessary to detect and correct issues in a trading system. The system developed predicts the overall system state based on the individual states of the sequencing component, distribution component, and matching component. Key features were selected based on filter selection methods using statistics from the data. For the classification, the group used the C4.5 tree algorithm, the naive Bayesian classifier, and the Random Forest algorithm. From metrics such as accuracy, precision, recall, and ROC, the authors conclude the Random Forest approach, which builds decision trees during training, can best represent system state. While the paper provides a good basis for developing a warning system for anomalous behavior and suggests future work in specific areas such as false positive reductions, it does not clearly interpret its results or provide a control group of human anomaly detectors. While it discusses the challenge filtering the features, it is somewhat biased by the assumption that the classic statistical relationships and components selected by domain expertise will best represent the market.

**1.4 Article 4**

**1.5 Article 5**

**1.6 Article 6**

**1.7 Article 7**

**1.8 Article 8**

**1.9 Article 9**

**2 Q2: ChatGPT Topic Description**