# Development of stock correlation networks

Lixin Huang, Princeton International School of Mathematics and Science, USA, henry.huang@prismsus.org

***Abstract***

How to depict the relationships between stocks has always been a focus for scholars. Knowing the relationship between two stocks means that we can adjust the investment plans based on the correlation between the stocks. We are able to lower the risk of the portfolio while maintaining its expected return if we know the correlations between the stocks in the portfolio, assuming that information carries on through time In this study, we establish a method to depict the relationship between two stocks in a more generalized way, as to provide a new approach to find the relationship between two stocks other than correlation. The following four categories are taken into account: the correlation between the stocks, how close the stocks are in case of the category of the companies that issue the stocks, how frequently that the two stocks are mentioned together, and possible transaction in the business between the two companies that issue the stocks. To determine the relationships between stocks, an algorithm is initiated to generate a score between 0 and 1 for all four categories described above. Typically, a higher score indicates a more significant relationship between the two stocks. The data of the stocks are imported from the wind database, including the price and category of the stocks. The business transactions between the companies have been taken from D&B Hoovers. Primary and secondary sources about the stocks will be considered as textual evidence. On the basis of the algorithm, the following 4-step analyses have been conducted. First, the correlation between the two stocks is calculated using the covariance matrix from the DCC-GARCH model. We assume the score of the correlation section equals the correlation between the two stocks. Second, if the fields of the two companies that issued the stocks are closer, the score for this section will be higher. Third, the score for the business transaction between the companies is determined by the proportion of transactions between the two companies. Last but not least, the score for textual evidence will be calculated using the equation below.

$$s_t = 1 - \frac{1}{\ln(n)}$$

Where $\mathbf{s_t}$ is the score for textual evidence and **n** is the number of articles that mentioned both stocks. The final score between the two stocks is calculated by the weighted average of the scores for the four categories. After the score between each pair of stocks in the market is determined, an app is developed to display the top ten correlated stocks with the user's search to facilitate and optimize their selections in the stock exchange market. In the future, this research could be conducted in the following three aspects. To begin with, the weight for the score for each part can be adjusted with more stock examples in order to depict the relationship between two stocks more accurately. Furthermore, the graph can be more user-friendly to increase the engagement of the users. Last but not least, a more sophisticated, multi-level categorization can be developed to optimize the categorization of the stocks given.

978-1-7281-7520-1/20/$31.00 ©2020 IEEE