ISSN 2616-5775 Vol. 6, Issue 10: 85-92, DOI: 10.25236/AJCIS.2023.061013

Portfolio Optimization Based on Complex Networks and Genetic Algorithms

Bo Liu*

School of Mathematics, Jilin University, Changchun, 130012, China
**Corresponding author: paulliu2004@163.com*

Abstract: *Portfolio optimization is a crucial endeavor in finance which aims to effectively manage investment risks and maximize returns. This paper explores the application of complex networks and genetic algorithms as a solution to the challenges associated with portfolio optimization. This paper strives to optimize the composition of portfolios, mitigate risks, and enhances potential returns by analyzing the interdependencies and correlations among financial assets using complex networks and utilizing genetic algorithms as an optimization technique. The results demonstrate that the portfolio resulting from the optimization of genetic algorithms applied to complex networks exhibits remarkable risk control capabilities. This integrated approach effectively minimizes risks associated with investments, contributing to the creation of a more stable and resilient portfolio, particularly in volatile financial markets.*

Keywords: *Portfolio Optimization, Complex Network, Genetic Algorithm, Risk Control*

1. Introduction

Portfolio optimization is a fundamental aspect of finance, aiming to achieve an optimal allocation of assets to maximize returns while minimizing risks. The integration of complex networks and genetic algorithms have emerged as a promising approach to address the challenges faced in portfolio optimization. This study aims to leverage the power of complex networks and genetic algorithms to optimize portfolio composition, enhance risk control capabilities, and contribute to the advancement of portfolio management strategies.

Portfolio optimization has witnessed significant advancements throughout its history. It originated with Markowitz's seminal work in 1952, where he introduced mean-variance analysis and laid the foundation for modern portfolio theory[1]. This framework was further developed by Sharpe (1964) and Lintner (1965), who incorporated concepts such as the Capital Asset Pricing Model (CAPM) and the efficient market hypothesis to enhance risk and return analysis in portfolio construction[2,3]. The emergence of complex networks provided a fresh perspective on understanding the interconnections among financial assets[4-8]. Bonanno et al. (2003) applied network theory to explore the correlation structure of stocks and revealed complex patterns within financial markets[9]. This work laid the groundwork for further investigations in this field. Li, Y., et al. (2019) and Clemente, G. P., et al. (2021) demonstrated the possibility of using a network approach to solve the portfolio optimization problem [10,11]. The application of genetic algorithms in portfolio optimization was conducted by Zhang, P. M. (2022)., demonstrating their ability to handle large-scale investment portfolios and enhance the efficiency of the optimization process[12].

However, challenges persist in effectively integrating complex networks and genetic algorithms for portfolio optimization while considering risk control. This paper aims to address this gap by investigating the performance of genetic algorithms applied to complex networks. We focus on enhancing risk control capabilities in portfolio optimization to achieve more robust and resilient investment portfolios.

We employ complex networks to capture the intricate interdependencies and correlations among financial assets. We tend to identify an optimal portfolio composition that strikes a balance between risk and returns by leveraging genetic algorithms as an optimization technique. The research aims to demonstrate the superior risk control capabilities of the portfolio obtained through the application of genetic algorithms to complex networks.

ISSN 2616-5775 Vol. 6, Issue 10: 85-92, DOI: 10.25236/AJCIS.2023.061013

2. Preliminary

In this section, we present the foundational knowledge and concepts that underpin our research. We introduce the relevant techniques and establish a unified notation for the entire paper.

2.1 Portfolio Optimization

Portfolio optimization is a vital component of modern finance, aiming to construct an optimal mix of financial assets to achieve the best risk-return trade-off. Mathematically, given a set of N financial assets, a portfolio is represented as a column vector $x = [x_1, x_2, ..., x_N]$ ', where each x_i denotes the proportion of wealth invested in asset i, and $\Sigma x_i = 1$.

2.2 Genetic Algorithms

Genetic algorithms (GAs) are optimization techniques inspired by the process of natural selection. They are particularly well-suited for solving complex and non-linear problems. In our study, GAs are employed to find the optimal portfolio composition by mimicking the process of genetic evolution. The algorithm iteratively generates and evolves a population of potential portfolios through selection, crossover, and mutation operations, converging toward an optimal solution.

2.3 Betweenness structural entropy

Developed by Qi Zhang and Meizhu Li[13], the betweenness structural entropy is a measure used in this paper to quantify the centrality and influence of each stock in connecting different parts of the network. It assesses the extent to which a stock serves as a bridge or intermediary between other stocks in the portfolio. A lower betweenness structural entropy indicates that a stock has less influence in connecting different parts of the network, suggesting a more balanced distribution of influence among the stocks.

2.4 Sharpe ratio

The Sharpe ratio is a risk-adjusted performance measure that helps investors assess the return of an investment relative to its risk. It provides a metric to compare different investment opportunities by considering both the potential returns and the associated volatility or risk. A higher Sharpe ratio indicates a more favorable risk-return trade-off, as it implies a higher return achieved for each unit of risk assumed. It suggests that the investment or portfolio is generating better risk-adjusted returns compared to alternatives with lower Sharpe ratios.

2.5 Data Set Description

The data set utilized in this paper is sourced from the Wind data terminal, a comprehensive financial database. We selected a combination of five leading stocks, commonly referred to as "blue-chip" stocks, and forty-five randomly chosen stocks from the A-share market. This selection ensures a diverse representation of the market. The data set consists of daily returns spanning a ten-year period from 2010 to 2020, with 2 trading days. As a result, we obtained a matrix of size 2520 ×50, representing the returns of the selected stocks over the ten-year period.

2.6 Notation

The notation of the whole paper is shown in the Table 1 below.

ISSN 2616-5775 Vol. 6, Issue 10: 85-92, DOI: 10.25236/AJCIS.2023.061013

Table 1: Notation of the whole paper

Symbol	Description
N	Number of financial assets
x_i	The proportion of wealth invested in asset i
A	Adjacency matrix
E	The betweenness structural entropy of the network
p_{ij}	Connection probabilities
r	A random number between 0 to 1
σ_{st}	The number of shortest paths from node s to node t.
$\sigma_{st}(i)$	The number of shortest paths from node s to node t that pass through node i
$B(i)$	The betweenness of node i
$P(i)$	The distribution of the structural components of node i in the whole network

3. Model Construction

3.1 Problem Analysis

The primary objective of this study is to optimize the composition of a portfolio to achieve the best risk-return trade-off using complex networks and genetic algorithms. We can solve the problem by following steps:

Network Construction. The research utilizes the correlation coefficients between the stocks as a measure of their interdependencies. These correlation coefficients serve as the basis for constructing a complex network representation of the portfolio. The network represents each stock as a node, and the connectivity between stocks is determined based on the corresponding correlation coefficients. Higher correlation coefficients imply a greater likelihood of a connection between stocks in the network.

Optimization Objective. The optimization objective is to minimize the betweenness structural entropy of the network. The betweenness structural entropy measures the centrality and influence of each stock in connecting different parts of the network. By minimizing the betweenness structural entropy, the research aims to identify an optimized network configuration that achieves a well-balanced allocation of investment among the stocks.

Genetic Algorithm Optimization. The research employs a genetic algorithm approach to achieve the optimization objective. The genetic algorithm iteratively generates and evolves a population of networks, representing different portfolio allocations. The fitness of each network is evaluated based on its betweenness structural entropy. Through successive generations, the genetic algorithm identifies and refines networks with lower entropy, leading to an optimized portfolio allocation.

Portfolio Allocation and Risk Control. The optimized network obtained from the genetic algorithm represents an allocation of investment among the stocks. The allocation is based on the network's node degrees, which reflect the importance of each stock in the network. Stocks with higher degrees are assigned greater investment proportions.

3.2 Model building

3.2.1 Network Construction

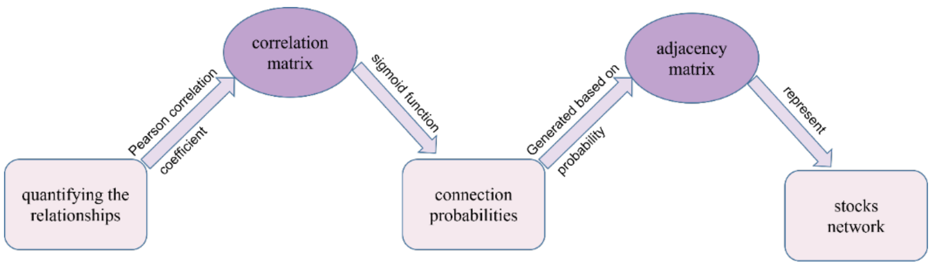

Figure 1: Network construction flow chart

As Figure 1 shows, the process of constructing the network from the data set starts with quantifying the relationships between the stocks which captures the correlation matrix that is computed based on the

ISSN 2616-5775 Vol. 6, Issue 10: 85-92, DOI: 10.25236/AJCIS.2023.061013

returns data. This matrix measures the statistical correlation between each pair of stocks in the portfolio, reflecting the degree of similarity or dissimilarity in their return patterns. A higher correlation coefficient suggests a stronger positive relationship, while a lower coefficient indicates a weaker or negative relationship. Having the correlation matrix serves as the basis, we are able to construct the network by representing each stock as a node in the network, and the connections between stocks are determined based on the corresponding correlation coefficients. The network captures the interdependencies and linkages among the stocks, enabling a comprehensive analysis of their collective behavior.

In the practice of constructing the network, the correlation coefficients are transformed into connection probabilities(denoted as p_{ij}) using a sigmoid function:

$$f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

This function maps the correlation coefficients to probabilities in the range of 0 to 1. Stocks with higher correlation coefficients have higher probabilities of being connected in the network, indicating a stronger relationship between them. This probabilistic approach allows for capturing the varying strengths of relationships within the portfolio.

Based on the connection probabilities, an adjacency matrix is constructed to represent the network. The adjacency matrix is a binary matrix where the presence of a connection between two stocks is denoted by a value of 1, while the absence of a connection is represented by a value of 0. This matrix serves as a fundamental representation of the network structure, facilitating further analysis and optimization.

3.2.2 Optimization Algorithm

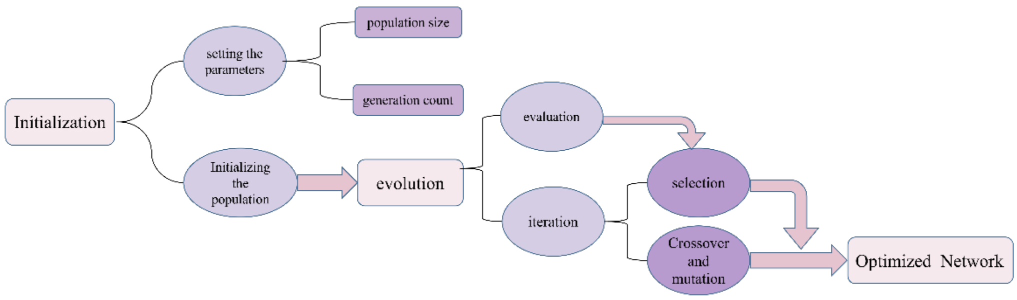

Figure 2: Optimization algorithm flow chart

In this research, a genetic algorithm is employed as the optimization algorithm to refine the network structure and achieve an optimized portfolio composition. The genetic algorithm iteratively evolves a population of networks to minimize the betweenness structural entropy and identify an optimal portfolio allocation. Figure 2 shows how the genetic algorithm operates in this paper.

The optimization process starts with setting the parameters for the genetic algorithm. The population size is determined to be 50, representing the number of networks or potential portfolio allocations within each generation. The generation count is set to 200, indicating the number of iterations or generations for which the genetic algorithm evolves the population.

Initialization of the population with networks is the vital job of the optimization process. Each network is represented as an adjacency matrix, capturing the connections between stocks. The adjacency matrix is generated by a function that constructs a random network based on the given correlation coefficients. The probability that two nodes are connected is denoted as p_{ij}, and r is a random number. This step ensures diversity in the initial population.

$$A(a_{ij}) = \begin{cases} 1, & p_{ij} > r \\ 0, & p_{ij} \le r \end{cases} \tag{2}$$

The key to a genetic algorithm is evolution, which is based on evaluation and selection. The fitness of each network in the population is evaluated based on its ability to minimize the betweenness structural entropy. The betweenness structural entropy measures the centrality and influence of each stock in connecting different parts of the network. A lower betweenness structural entropy indicates that a stock has less influence in connecting different parts of the network, suggesting a more balanced distribution of influence among the stocks, while a higher value suggests a more concentrated or unequal distribution. A function calculates the betweenness structural entropy by considering the network represented by the adjacency matrix.

According to Qi Zhang and Meizhu Li[13]. The definition of the betweenness centrality of node i is:

$$B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{3}$$

where the numerator $\sigma_{st}(i)$denotes the number of shortest paths from node *s* to node *t* that pass through node *i,* while the denominator σ_{st} represents the total number of shortest paths from node s to node t. The next equation calculates the probability *P(i)* of node *i*, which represents the distribution of the structural components of node *i* in the entire network.

$$P(i) = \frac{B(i)}{\sum_{i=1}^{N} B(i)} \tag{4}$$

We can define the betweenness structural entropy *(E)* of the complex networks as follows:

$$E = -\sum_{i=1}^{N}\left(P(i) * ln\left(P(i)\right)\right) \tag{5}$$

The selection process aims to choose the fittest networks as parents for the next generation. Networks with lower betweenness structural entropy, indicating better performance, are more likely to be selected. The selection process is performed by comparing the fitness of each network and probabilistically selecting them based on their fitness values. The chosen networks undergo crossover, which involves exchanging genetic information which is represented by the adjacency matrix, or the connections between stocks, to generate offspring networks with potentially improved performance.

Another way to introduce diversity and explore new solutions is mutation. Random changes are made to the connections within the network with a certain probability that is 10%. This random alteration enables the exploration of different network configurations beyond the existing population, enhancing the chances of finding an optimal solution. The process of the network's crossover and mutation is illustrated in Figure 3.

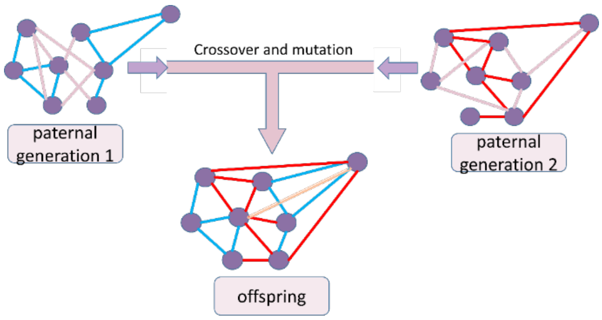

Figure 3: The process of iteration

After crossover and mutation, the offspring networks replace the previous generation's population. This population update ensures the evolution of the population towards networks with lower betweenness structural entropy. The updated population is then subjected to subsequent iterations of selection, crossover, and mutation, leading to the continuous refinement of the network structure.

The best network with the lowest betweenness structural entropy is continuously updated and stored as the "best network." This network represents an optimal portfolio allocation with minimized risk and maximized returns based on the network structure and stock connections.

3.2.3 Portfolio Optimization and Comparison

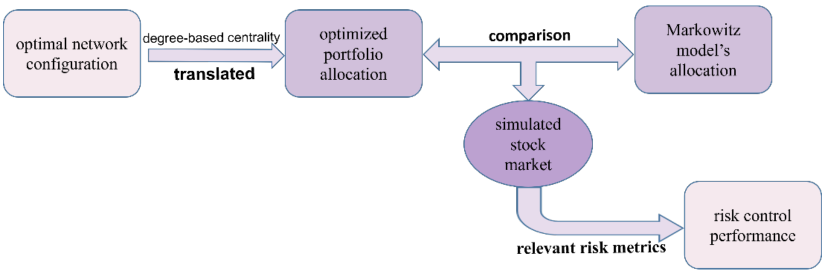

Figure 4: Portfolio optimization and comparison flow chart

In this study, the optimization process extends beyond network refinement to portfolio optimization. The optimal network configuration obtained through the genetic algorithm is translated into an optimized

ISSN 2616-5775 Vol. 6, Issue 10: 85-92, DOI: 10.25236/AJCIS.2023.061013

portfolio allocation. Furthermore, a comparison is made between the optimized portfolio based on the genetic algorithm and the traditional Markowitz model. This comparison is performed using a simulation of the stock market to observe their respective performances in risk control. The process of portfolio optimization and comparison is shown in Figure 4.

The optimal network configuration obtained through the genetic algorithm provides valuable insights into the relative importance and relationships between stocks. Each stock's degree centrality is determined, reflecting its significance within the network. This degree-based centrality is then used to allocate investment weights to each stock in the portfolio. There is a direct proportionality between the degree centrality and the weight assigned to the corresponding stock. This process allows for an optimized portfolio allocation based on the network configuration.

The traditional Markowitz model is employed as a benchmark for comparison. In this model, the optimization is based on the expected returns and covariance matrix of the stocks. The Markowitz model aims to find the portfolio allocation that maximizes returns for a given level of risk.

A simulated stock market environment is created to assess the performance of the optimized portfolio derived from the genetic algorithm and the Markowitz model. The historical returns data used to construct the network and calculate the correlation matrix are utilized in the simulation. By simulating the market conditions and the performance of the optimized portfolios over a specified time horizon, the risk control capabilities of the two approaches can be observed and compared.

The risk control performance of the optimized portfolios is evaluated using relevant risk metrics, such as the Sharpe ratio. The Sharpe ratio measures the risk-adjusted returns of a portfolio, considering both the returns and the volatility or risk level. Insights into their respective risk control abilities can be obtained by comparing the Sharpe ratios of the optimized portfolios derived from the genetic algorithm and the Markowitz model.

4. Experiment and analysis

4.1 Design

The experiment involves implementing the optimization algorithm and comparing the performance of the optimized portfolio derived from the genetic algorithm with that of the traditional Markowitz model. The experimental procedure is shown below:

Firstly, the stock price data for the selected stocks are collected over the specified time period from the Wind data terminal. The daily stock returns are then calculated from the price data to construct the returns matrix. This step ensures the availability of accurate and reliable data for analysis.

Next, the correlation matrix is computed from the returns matrix, capturing the interdependencies and relationships between stocks. Based on the correlation coefficients, the network connections are established by assigning probabilities to connect the stocks. These probabilities are determined by mapping the correlation coefficients to the sigmoid function, providing a probabilistic framework for network construction.

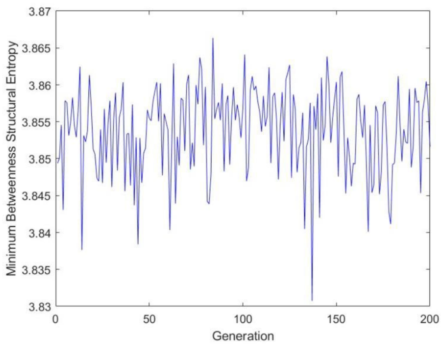

Figure 5: Evolution of minimizing betweenness structural entropy

The genetic algorithm is implemented to optimize the network structure and derive the optimal portfolio allocation. The algorithm iteratively evolves the population of networks, guided by the objective of minimizing the betweenness structural entropy. The evolution of minimizing betweenness structural

ISSN 2616-5775 Vol. 6, Issue 10: 85-92, DOI: 10.25236/AJCIS.2023.061013

entropy is shown in Figure 5. And the optimal network obtained from the minimum betweenness structural entropy is displayed in Figure 6. The optimal network configuration is then translated into an optimized portfolio allocation. The degree-based centrality of each stock in the network serves as a basis for determining the investment weights in constructing the portfolio.

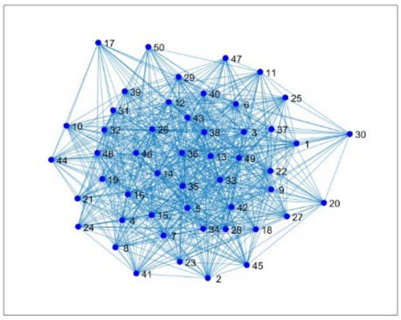

Figure 6: The optimal network

A simulation of the stock market is conducted using the historical returns data. The optimized portfolios derived from the genetic algorithm and the Markowitz model are evaluated based on risk metrics such as the Sharpe ratio. The performance of the portfolios is analyzed and compared to assess their risk control capabilities. The results are obtained by conducting this experiment by programming based on MATLAB.

4.2 Results

Table 2: Performance of two portfolios in stimulated stock market

Method	Total returns	Sharpe ratio
Degree-Based	0.2667	0.0685
Markowitz	0.1058	0.0299

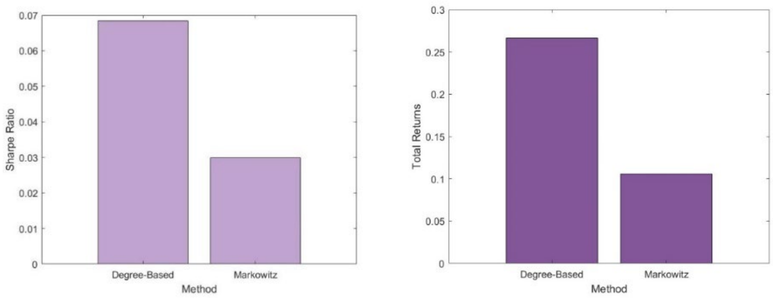

Figure 7: Comparison of Sharpe Ratios and Total Returns

The experimental results revealed intriguing insights into the risk control capabilities of the optimized portfolio derived from the network-based approach compared to the traditional Markowitz portfolio. The performance of the portfolios was evaluated using two key metrics: total returns and the Sharpe ratio.

The performance of two portfolios in the stimulated stock market is displayed in Table.2. The degree-based portfolio achieved a total return of 0.2667, indicating a higher overall gain compared to the Markowitz portfolio, which recorded a total return of 0.1058. However, focusing solely on total returns does not provide a comprehensive assessment of a portfolio's performance, as it neglects the level of risk associated with achieving those returns.

The Sharpe ratio was utilized as a valuable risk-adjusted performance measure to account for risk. The degree-based portfolio exhibited a Sharpe ratio of 0.0685, surpassing the Markowitz portfolio's Sharpe ratio of 0.0299. The higher Sharpe ratio of the degree-based portfolio indicates a superior risk-adjusted return per unit of risk taken. This implies that the degree-based portfolio achieved a more favorable trade-off between risk and returns compared to the Markowitz portfolio.

The comparison of Sharpe ratios and total returns shown in Figure 7 suggests that the degree-based

ISSN 2616-5775 Vol. 6, Issue 10: 85-92, DOI: 10.25236/AJCIS.2023.061013

portfolio not only generates higher total returns than the Markowitz portfolio but also exhibits a stronger ability to control risk. This is a significant advantage, particularly in volatile market conditions, as the degree-based portfolio demonstrated remarkable resilience in minimizing losses during market downturns.

The results highlight the value of the network-based approach in portfolio optimization. By incorporating the complex network structure and leveraging genetic algorithms, the approach effectively identifies an optimized portfolio allocation that prioritizes risk control. The degree-based portfolio's emphasis on the connectivity and importance of each stock in the network enables the construction of a well-diversified and robust portfolio.

However, it is essential to acknowledge the limitations of the experimental findings. The study utilized a specific dataset and focused on a defined set of stocks, which may limit the generalizability of the results. Additionally, the evaluation was conducted using historical data, and the performance in simulated market conditions may not directly translate to real-world scenarios.

5. Conclusion

The paper aims to optimize portfolio composition using complex networks and genetic algorithms and evaluate their risk control capabilities. This optimization process achieved a well-balanced allocation of investments among the stocks, considering their interdependencies and risk factors. The experimental findings demonstrated that the optimized portfolio derived from our network-based approach outperformed the traditional Markwotiz approach. It indicates that our approach effectively mitigates risks and provides a more stable and resilient investment portfolio. The significance and value of this research lie in its ability to enhance risk control in portfolio optimization. We provide a novel perspective for constructing and optimizing portfolios that consider interdependencies among stocks. This may benefit investors in making more informed decisions and achieving better risk-adjusted returns. There are several avenues for future research, such as: exploring more advanced optimization techniques or alternative network-based approaches, investigating the impact of incorporating additional factors, conducting real-world validation, and exploring the scalability of the approach to larger portfolios and more diverse markets.

References

[1] Markowitz, H. (1952). Portfolio Selection. Journal of Finance 7(1): 77-91.
[2] Sharpe, W. F. (1964). Capital-Asset Prices - A Theory of Market Equilibrium under Conditions of Risk. Journal of Finance 19(3): 425-442.
[3] Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios And Capital Budgets. Review of Economics and Statistics 47(1): 13-37.
[4] Pontes, L. S. D. and L. C. Rego (2022). Impact of macroeconomic variables on the topological structure of the Brazilian stock market: A complex network approach. Physica a-Statistical Mechanics and Its Applications 604: 14.
[5] Zhou, Y., et al. (2023). Analysing and forecasting co-movement between innovative and traditional financial assets based on complex network and machine learning. Research in International Business and Finance 64.
[6] Hu, S. Y., et al. (2019). An analysis of the clustering effect of a jump risk complex network in the Chinese stock market. Physica a-Statistical Mechanics and Its Applications 523: 622-630.
[7] Liu, C. and N. Arunkumar (2019). Risk prediction and evaluation of transnational transmission of financial crisis based on complex network. Cluster Computing-the Journal of Networks Software Tools and Applications 22(2): S4307-S4313.
[8] Samal, A., et al. (2021). Network geometry and market instability. Royal Society Open Science 8(2).
[9] Bonanno, G., et al. (2003). Topology of correlation-based minimal spanning trees in real and model markets. Physical Review E 68(4).
[10] Li, Y., et al. (2019). Portfolio optimization based on network topology. Physica a-Statistical Mechanics and Its Applications 515: 671-681.
[11] Clemente, G. P., et al. (2021). Asset allocation: new evidence through network approaches. Annals of Operations Research 299(1-2): 61-80.
[12] Zhang, P. M. (2022). Research on Securities Portfolio Model Based on Genetic Optimization Neural Network. Security and Communication Networks 2022: 10.
[13] Zhang, Q. and M. Z. Li (2022). A betweenness structural entropy of complex networks. Chaos Solitons & Fractals 161.