

Asset Subsets Identification via Investment Universe Complex Network

Mert Arda Asar^{1,†}, Günce Keziban Orman^{1,†}

¹Galatasaray University, Ortaköy, Çırağan St, 34349 Beşiktaş, İstanbul, Turkey

Abstract

This study seeks to identify an optimal asset subset by ensuring diversity through *company description* and *price behavior over time*. We propose the *Investment Universe Complex Network (InUCoN)* framework, which models the investment universe as a complex network. InUCoN consists of three components: dynamic network generation, snapshot aggregation, and universe filtering. Experiments on S&P stocks demonstrate that InUCoN reduced risk by selecting a more independent stock set, with our proposed portfolio yielding a 32% higher return compared to the unfiltered universe.

Keywords

Complex networks, Financial Networks, Dynamic Network Modeling, Portfolio Allocation

1. Introduction

In the complex and ever-evolving landscape of financial markets, investors face the significant challenge of selecting an optimal subset of assets from a vast investment universe. Filtering this universe into stocks from different sectors and behaviors is essential for distributing risk and diversifying the portfolio. The primary difficulty lies in filtering assets by not only considering recent performance and correlations but also preserving their inherent, often subtle, relationships. These relationships can be intricate and dynamic, reflecting the deeply interconnected nature of financial markets. Overlooking these complexities can lead to suboptimal portfolio construction, exposing investors to unforeseen risks. Traditional methods often depend heavily on time-series correlations or simply cluster stocks by the sectors in which they operate. This approach neglects the dynamic, emergent relationships among assets because it fails to consider the investment universe as a complex, interdependent system. While there have been studies modeling financial markets with complex networks [1, 2, 3] and efforts to identify the most appropriate metrics for constructing financial market networks [4], none have yet applied this modeling approach specifically to the challenge of stock allocation.

This study aims to identify an optimal subset of assets by ensuring diversity through *company description* and *price behavior over time*. While recent price behaviors reflect short-term market trends and sentiment, enduring characteristics provide a foundation for long-term decisions. We introduce the *Investment Universe Complex Network (InUCoN)* framework, integrating these aspects to address previous limitations. Our major contributions are: (i) developing a framework using complex networks for dynamic and static modeling to diversify assets; (ii) generating a hybrid similarity metric combining both aspects; (iii) optimizing this metric using the network's structural consistency; and (iv) experimentally validating the effectiveness of the filtered asset subsets.

2. Method

The main strategy of InUCoN is to treat the investment universe as a complex system, model it with a complex network, and analyze it using this model. The framework has three components: (i)

Discovery Science - Late Breaking Contributions 2024

*Corresponding author.

✉ maasar@gsu.edu.tr (M. A. Asar); korman@gsu.edu.tr (G. K. Orman)

ORCID 0009-0007-6357-8204 (M. A. Asar); 0000-0003-0402-8417 (G. K. Orman)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

dynamic network generation; (ii) snapshot aggregation; and (iii) investment universe filtering. In dynamic network generation, assets are modeled through a series of snapshots, each representing the system’s state at a specific time interval. This approach captures the evolving nature of asset relationships over time. While the assets are represented as fixed nodes, their relationships (links) change dynamically, reflecting the fluid nature of financial markets where correlations can shift rapidly due to economic factors. Constructing these relationships carefully is essential, as they directly influence the accuracy and effectiveness of the analysis. We define a hybrid asset similarity $sim_{hib}(a_i, a_j) = w_1 \cdot sim_{tx}(a_i, a_j) + w_2 \cdot sim_{ts}(a_i, a_j)$ considering both stock description similarity sim_{tx} , which is calculated using the cosine similarity of embedding vectors generated by the DistilBert [5] model using asset descriptions, and time-series similarity sim_{ts} . For sim_{ts} , we propose four metrics: dynamic time warping, distance correlation, Euclidean distance, and Pearson correlation. During the time-series similarity analysis, we apply an overlapping window approach. This ensures that the relationship between two consecutive snapshots is not entirely lost, providing a smoother transition between time intervals and capturing ongoing trends. The weights w_1 and w_2 are experimentally optimized to achieve the highest average network consistency for all snapshots, based on a “structural consistency” index from first-order matrix perturbation [6]. Pairs of nodes with sim_{hib} values above a certain *threshold* are linked in the snapshots. *threshold* is obtained during experiments.

In the snapshot aggregation phase, a single market network G_{final} was formed to represent all snapshots. Nodes in G_{final} correspond to the same nodes of snapshots. Links were established between nodes that shared a community in at least one snapshot, with weights indicating the frequency of such occurrences. The infomap algorithm identified communities in all snapshots. This community-based aggregation approach allows us to capture both stable and evolving relationships between assets over time, providing a more comprehensive view of market dynamics. In the investment universe filtering phase, we aimed to select the most diversified stock subset using network centrality measures. We employed a two-step process based on closeness centrality and PageRank scores, designed to identify stocks that are both independent and influential within the network. First, we selected nodes with lower than average closeness centrality, identifying stocks relatively disconnected from overall market trends. From this subset, we then chose stocks with PageRank scores above a calculated threshold. A node with a high PageRank score and low closeness centrality suggests that the stock is influential within the network but not necessarily closely connected to other nodes. For determining the PageRank threshold, we used the following equation:

$$threshold_{pr} = \mu_{PageRank} + 0.5 \cdot \sigma_{PageRank} \quad (1)$$

Where $\mu_{PageRank}$ is the mean PageRank score of the selected nodes (those with below-average closeness centrality), and $\sigma_{PageRank}$ is the standard deviation of these scores. This approach aims to construct a portfolio that maximizes diversification while ensuring each selected asset contributes significantly to overall network dynamics.

3. Experiments and Results

We selected the 200 highest average trading volume stocks from the S&P 500 for our experiments, covering 01/06/2021 to 01/06/2024. To evaluate InUCoN’s effectiveness, we applied the Markowitz mean-variance optimization algorithm [7] for portfolio allocation and calculated the average portfolio return based on the allocated weights. We constructed a dynamic network using a six-month period, with the remaining time used to measure performance, following a buy-and-hold strategy with no rebalancing. We compared InUCoN with the baseline (the unfiltered universe) and with portfolios using single similarity metrics instead of our proposed hybrid approach. As shown in Figure 1, the historical cumulative returns demonstrate that InUCoN effectively reduced risk and improved portfolio performance by selecting a more independent set of stocks.

Notably, the text-only model showed strong performance, outperforming several time-series-based approaches. This suggests that company descriptions contain valuable information for portfolio diversification that may not be fully captured by price movements alone. By incorporating static textual

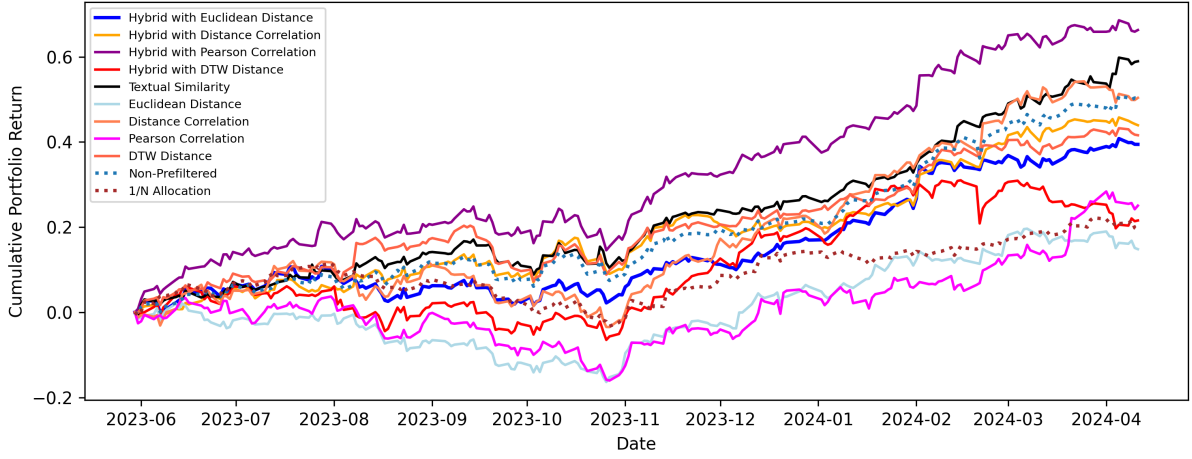


Figure 1: Average Allocation Return for US Stocks After 5 Runs

information, we may have been able to avoid some of the noise inherent in price data during the filtering process. By leveraging both quantitative price data and qualitative company information, InUCoN appears to capture a more holistic view of each asset's role in the investment universe. This indicates that while both types of data provide valuable insights, integrating them allows for a more comprehensive understanding of a stock's behavior and its relationships within the market network. This comprehensive approach leads to more effective portfolio diversification and improved risk-adjusted returns, as evidenced by the superior performance of the hybrid method in our results.

4. Conclusion

In this study, we introduced the Investment Universe Complex Network (InUCoN) framework as a novel method to refine asset selection in dynamic financial markets. By leveraging textual and time series similarities within a complex network model, we aimed to enhance portfolio allocation algorithms. We demonstrated the utility of the structural consistency metric in optimizing network generation. InUCoN significantly improved portfolio outcomes compared to both the unfiltered stock universe and previously highlighted metrics. We outperformed the unfiltered investment universe by 32% increased return over the time period by modeling financial markets based on Pearson correlation.

Acknowledgments

This work is supported by the Galatasaray University Research Fund (BAP) within the scope of project number FBA-2024-1240, titled *Bağlantı Tahmini Yöntemleri ve Sınıf Dengesizliği Sorununa Çözüm Yaklaşımları*.

References

- [1] W. Huang, H. Wang, Y. Wei, J. Chevallier, Complex network analysis of global stock market co-movement during the covid-19 pandemic based on intraday open-high-low-close data, *Financial Innovation* 10 (2024) 1–16. doi:10.1186/s40854-023-00548-5.
- [2] D. Jothimani, C. Kavaklioğlu, A. Başar, Financial networks: A study of the toronto stock exchange (2018) 4684–4691. doi:10.1109/BigData.2018.8621969.
- [3] N. Khan, S. Bano, M. Durrani, A network analysis of the pakistan stock exchange, *Webology* 18 (2021) 5744–5763.

- [4] S. C. Ugwu, P. Miasnikof, Y. Lawryshyn, Distance correlation market graph: The case of s&p500 stocks, *Mathematics* 11 (2023) 3832. doi:10.3390/math11183832.
- [5] V. Sanh, L. Debut, J. Chaumond, T. Wolf, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, *arXiv.org* (2019).
- [6] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, H. E. Stanley, Toward link predictability of complex networks, *Proceedings of the National Academy of Sciences of the United States of America* 112 (2015) 2325–2330. doi:10.1073/pnas.1424644112.
- [7] H. Markowitz, Portfolio selection, *The Journal of Finance* 7 (1952) 77–91.