

그래프 필터링 및 중심성 측도에 기반한 포트폴리오 구성 전략

Graph-Based Portfolio Selection via Filtering and Centrality Measures

이우식¹

Woosik Lee¹

요 약

본 연구는 그래프 이론의 필터링 방법(MST, TMFG)과 세 가지 중심성 측도(차수, 고유벡터, 부분 그래프)의 조합을 통해 포트폴리오의 종목 선택 능력을 분석하였다. 다우존스 산업평균지수 내 종목 변경이 없는 29개 종목을 대상으로 금융 네트워크에서 주변부에 위치한 자산 수($K=5, 10, 15, 20$)를 선택하였고, 이를 균등 배분하여 포트폴리오 성과를 분석하였다. 실험 결과, TMFG와 고유벡터 중심성 조합이 자산 수가 10개에서 샤프 비율 2.65로 가장 우수한 위험 조정 성과를 보였으며, 자산 수가 5와 15에서도 일관되게 높은 성과를 나타냈다. 차수 중심성은 MST와 TMFG 모두에서 자산 수가 증가할수록 성과가 개선되는 경향을 보였다. MST와 차수 중심성의 경우, 자산 수가 5에서 20까지 증가함에 따라 샤프 비율도 증가하였다. 이러한 결과는 필터링 방법과 중심성 측도의 선택이 포트폴리오 성과에 미치는 영향을 규명함으로써, 주변부 종목으로 구성된 포트폴리오가 우수한 성과를 보임을 확인하였다.

핵심어 : 비즈니스 애널리틱스, 계량 금융, 금융수학, 네트워크 분석, 위험 관리

Abstract

This study investigates the impact of graph filtering methods and centrality measures on portfolio stock selection. Two graph filtering methods - Minimum Spanning Tree (MST) and Triangulated Maximally Filtered Graph (TMFG)-are combined with three centrality measures-Degree, Eigenvector, and Subgraph centrality-to identify peripheral assets in financial networks. The analysis is conducted on 29 stocks from the Dow Jones Industrial Average with no constituent changes. Portfolios are constructed by selecting the least central assets at varying portfolio sizes(5,10,15, and 20 assets) and applying equal weight allocation. The empirical results indicate that the combination of TMFG and Eigenvector centrality shows the highest Sharpe ratio of 2.65 with 10 selected assets, and maintains high performance with 5 and 15 assets. Degree centrality exhibits improving performance as portfolio size increases across both filtering methods ; for MST with Degree centrality, the Sharpe ratio rises from 1.24 with 5 assets to 2.55 with 20 assets. These findings clarify the impact of the choice of filtering method and centrality measure on portfolio performance, demonstrating that portfolios composed of peripheral assets achieve higher risk-adjusted returns.

Keyword : Business Analytics, Quantitative Finance, Financial Mathematics, Network Analysis, Risk Management

¹ Gyeongsang National University, Jinju, Korea [Professor]
e-mail: woosiklee@gnu.ac.kr

Received(April 18, 2026), Review Result(1st: May 6, 2026), Accepted(June 13, 2026), Published(June 30, 2026)



© 2026 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

1. 서론

재무 이론에서 투자 위험을 완화하기 위한 기본 원리 중 하나는 상호 연관성이 낮은 다양한 금융자산을 결합하여 자금을 운용하는 분산투자이다 [1]. 이를 체계적으로 정립한 Markowitz의 평균-분산 최적화 모형은 자산 배분 이론의 기초를 제공하였으나, 공분산 행렬의 추정 불안정성과 역행렬 연산에 대한 의존이라는 구조적 한계를 지닌다 [2][3].

이러한 한계를 극복하기 위해 자산 간 수익률 상관계수 행렬을 거리 측도로 변환하고, 이를 기반으로 네트워크를 구성하여 활용하는 접근법이 제시되었다. Mantegna의 연구 [4]는 상관계수 기반 최소신장트리(Minimum Spanning Tree, MST)를 통해 금융시장의 계층적 구조를 처음 제시하였으며, Tumminello 외 3인의 연구 [5]는 평면 최대 필터 그래프(Planar Maximally Filtered Graph, PMFG)를 통해 MST보다 많은 엣지를 보존하여 자산 간 다자간 의존 관계를 표현할 수 있는 방법을 제안하였다. 이후 Massara 외 2인의 연구 [6]는 PMFG의 계산 효율성을 개선한 삼각 최대 필터 그래프(Triangulated Maximally Filtered Graph, TMFG)를 제안하였다. 이들 그래프 필터링(MST, PMFG, TMFG)은 이 완전 그래프에서 특정 위상적 조건을 충족하는 엣지만을 보존하여 핵심적인 의존 구조를 추출하는 과정이다 [5][7].

그래프 이론 기반 포트폴리오의 핵심 가설은 Pozzi 외 2인의 연구 [8]에 의해 제시되었다. 이 연구에서는 금융 네트워크에서 주변부에 위치한 자산, 즉 중심성이 낮은 자산으로 구성된 포트폴리오가 중심부 자산 포트폴리오보다 우수한 위험 조정 성과를 보인다는 것을 실증적으로 확인하였다. 이후 Peralta와 Zareci의 연구 [9]에 의해 최적 포트폴리오 전략이 고유벡터 중심성의 역에 비례하여 자산을 배분해야 한다고 제시하였다.

그러나 기존 연구들은 주로 단일 그래프 필터링 방법인 MST, PMFG 또는 TMFG 만을 사용하거나, 여러 중심성 측도를 결합한 혼합 측도를 활용함으로써, 개별 중심성 측도의 효과를 분리하여 평가하지 못하였다. 또한, 각 중심성 측도에 의해 선택되는 종목 수가 상이하여 비교가 어려운 문제가 존재하였다.

이에 본 연구는 두 가지 그래프 필터링 방법(MST, TMFG)과 세 가지 중심성 측도(차수 중심성, 고유벡터 중심성, 부분 그래프 중심성)의 조합을 비교하여, 총 6가지 전략의 종목 선택 능력을 분석한다. 종목 수를 5, 10, 15, 20개로 고정하고 균등 배분을 적용하였다. 이를 통해 필터링 방법의 선택과 중심성 측도의 선택이 각각 포트폴리오 성과에 미치는 영향을 비교하고, 특정 필터링-중심성 조합에서 나타나는 성과 특성을 분석하여 선행 연구와의 차별점을 확립하고자 하였다.

본 논문은 다음과 같이 구성된다. 제1장에서는 본 연구의 배경과 목적을 기술하며, 제2장에서는 그래프 필터링 방법과 중심성 측도에 대한 이론적 배경을 고찰한다. 제3장에서는 자료의 구성과

실증 분석 방법을 제시하고, 분석 결과를 제시한다. 제4장에서는 주요 결과를 바탕으로 결론과 시사점을 논의한다.

2. 이론적 배경

2.1 그래프 필터링

자산간 수익률 상관 계수를 그래프로 변환하기 위해, López de Prado[3]가 식(1)의 상관 기반 거리 측도를 활용하였다.

$$d_{i,j} = \sqrt{\frac{1}{2}(1 - \rho_{i,j})} \quad (1)$$

여기서 $\rho_{i,j}$ 는 자산 i 와 j 의 수익률 간 Pearson 상관관계수이다. 식 (1)에 의해 n 개 자산의 모든 쌍에 대한 거리를 산출하여 $n \times n$ 대칭 거리 행렬 D 를 구성한다. 이 거리 행렬을 기반으로 두 가지 그래프 필터링 방법을 적용한다. n 개 자산의 거리 행렬은 $n(n-1)/2$ 개의 엣지를 가진 완전 그래프로 표현되나, 이 중 상당수는 표본의 유한성으로 인해 우연히 발생한 통계적 노이즈에 해당한다 [7]. 그래프 필터링은 이 완전 그래프에서 특정 위상적 조건을 충족하는 엣지만을 보존하여 핵심적인 의존 구조를 추출하는 과정이다 [5][6][7].

MST는 모든 노드를 연결하면서 엣지 가중치의 합이 최소인 트리를 구성한다. n 개 자산에 대해 $n-1$ 개의 엣지만을 보존하므로 가장 희소한 필터링이다. MST는 트리 구조이므로 자산 간 관계가 단일 경로로만 연결되며, 주변부 노드는 차수(Degree)가 1인 잎 노드로 명확히 식별된다. TMFG는 평면 그래프 조건을 유지하면서 정보 보존량을 최대화하는 삼각화를 구성한다. n 개 자산에 대해 $3(n-2)$ 개의 엣지를 보존하며, 노드를 삼각형 면에 삽입하는 방식으로 구성되어 Chordal 구조가 보장된다 [7]. TMFG의 모든 면이 삼각형이므로 최소 차수가 3이며, 이로 인해 차수 기반의 주변부 식별에서 MST와 상이한 특성을 보인다.

2.2 중심성 측도

그래프가 구성되면, 각 노드의 네트워크 내 위치적 중요도를 정량화하기 위해 중심성 측도를 계산한다. 차수 중심성(Degree Centrality)은 노드에 직접 연결된 엣지의 수를 측정하는 가장 단순한 중심성 측도이다 [10].

$$DC_n = A \cdot \mathbf{1}_n \quad (2)$$

여기서 DC_n 은 모든 노드의 차수를 포함하는 $n \times 1$ 크기의 벡터이다. 1_n 은 모든 원소가 1인 $n \times 1$ 크기의 열벡터를 의미하며, 이 벡터와 A 와 곱하여 각 노드의 연결 수를 계산한다. 차수 중심성은 길이 1인 walk만을 고려하므로, 직접 연결의 수만으로 노드의 중요도를 판단한다.

반면에 고유벡터 중심성(Eigenvector Centrality)은 노드의 직접 연결뿐 아니라 연결된 이웃 노드의 중요도까지 재귀적으로 고려한다 [10].

$$EC_n = \frac{1}{\lambda_{\max}} A \cdot q_{\max} \quad (3)$$

여기서 EC_n 도 모든 노드의 고유벡터 중심성 값을 담은 $n \times 1$ 차원의 벡터이다. $\lambda_{\max}$ 는 벡터의 크기를 조정하여 전체 시스템 내에서 각 노드의 중심성 값을 비교 가능한 수준으로 유지하고, $q_{\max}$ 는 연결된 노드들의 영향력을 나타낸다. 이로써 길이 1보다 긴 walk를 고려하여 노드가 얼마나 중요한 이웃과 연결되어 있는지를 계산한다.

부분 그래프 중심성(Subgraph Centrality)은 노드가 참여하는 모든 닫힌 walk의 수를 가중 합산하여, 네트워크 내 소그룹 참여도를 측정한다 [11].

$$SC_n = \text{diag}(e^A) \quad (4)$$

여기서 SC_n 도 모든 노드 전체 중심성 값을 담은 $n \times 1$ 차원의 벡터이다. e^A 는 닫힌 walk를 통합하여 계산한 행렬 지수함수를 의미하며, diag 는 행렬의 대각성분만을 추출하여 각 원소의 값이 해당 노드의 부분 그래프 중심성 점수가 된다. 짧은 닫힌 경로에 더 높은 가중치를 부여하므로, 노드의 국소적 네트워크 구조 참여도를 포착할 수 있다 [11].

3. 실증분석

3.1 자료의 구성

분석대상은 다우존스 산업평균지수의 구성 종목이다. 분석기간 중 구성 종목의 변동이 없는 안정적인 유니버스를 확보하기 위해 2020년 9월 1일부터 2024년 2월 23일까지를 표본 기간으로 설정하였으며, 해당 기간 동안 편입과 편출이 발생하지 않은 29개 종목을 대상으로 하였다. 일별 수정 종가를 바탕으로 수익률을 산출하였다.

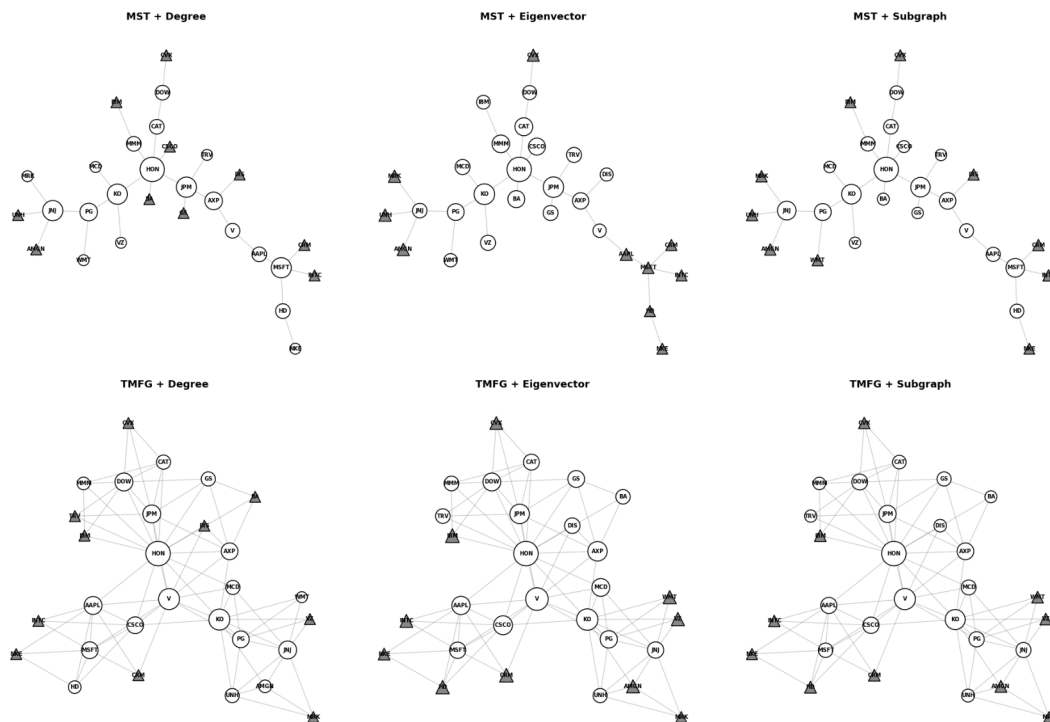
포트폴리오 성과를 평가하기 위해 롤링 백테스팅을 수행하였다. 전체 표본의 80%를 학습 구간, 나머지 20%를 테스트 구간으로 설정하였다. 학습 구간의 데이터를 기반으로 가중치를 산출하고, 20일 거래일 간격으로 리밸런싱을 수행하였다. 공매도를 허용하지 않았으며, 거래 수수료 및 세금

등의 제반 비용은 고려하지 않았다. 선택 자산 수(이하 K)는 Pozzi 외 2인의 연구 [8]를 참고하되, 본 연구의 유니버스를 초과하지 않는 범위에서 $K = 5, 10, 15, 20$ 으로 설정하였다.

3.2 모형의 추정 및 분석

자산 간 수익률 상관계수 행렬을 식(1)의 거리 척도로 변환하고, 이를 통해 거리 행렬을 산출하였다. 거리 행렬을 기반으로 MST와 TMFG를 각각 구성하였다. MST는 29개 노드에 28개 엣지($n - 1$) 그리고 TMFG는 29개 노드에 81개 엣지($3(n - 2)$)를 가진다.

각 그래프에 대해 세 가지 중심성 벡터(식(2)~(4))를 계산한다. 차수 중심성은 직접 연결 수가 적은 자산을, 고유벡터 중심성은 영향력 있는 허브로부터 멀리 떨어진 자산을, 부분 그래프 중심성은 닫힌 경로에 참여도가 낮은 자산을 주변부로 식별한다. 이 중심성이 가장 낮은 하위 K 개($K = 5, 10, 15, 20$) 자산을 선택하여 균등 배분으로 포트폴리오를 구성하였다.



[그림 1] 네트워크에서 중심성 척도별 노드 구조 비교

[Fig. 1] Node Structures by Centrality Measures in Networks

[그림 1]에서 먼저 MST에서는 세 중심성 척도 간 선택 종목의 차이가 관찰된다. MST와 차수 중심성의 경우 허브(HON, JPM, AXP) 근처에도 선택 종목(삼각형)이 나타난 반면, MST와 고유벡

터 중심성 및 MST와 부분그래프 중심성에서는 선택 종목이 대부분 주변부에 위치하였다. 다만 고유벡터 중심성과 부분그래프 중심성 사이에도 차이가 있다. MST와 고유벡터 중심성에서는 MSFT를 중심으로 AAPL, CRM, INTC, HD, NKE가 연결되어 있으며 이들 모두가 주변부로 선택되었으나, MST와 부분그래프 중심성에서는 동일한 구조에서 CRM, INTC, NKE만 선택되었다. 이는 고유벡터 중심성이 연결 수뿐 아니라 연결된 이웃이 누구인가까지 고려하기 때문이다. MSFT의 이웃인 AAPL과 HD은 다른 노드와의 연결이 적어 비영향력적이며, 고유벡터 중심성은 이러한 이웃과 연결된 MSFT 자체도 주변부로 판정한다. 반면 부분그래프 중심성은 닫힌 경로의 참여도를 측정하나, MST에는 측정할 닫힌 경로가 제한적이므로, MSFT처럼 다수의 노드와 연결된 자산은 선택에서 제외된다.

TMFG에서는 세 중심성 측도 간 선택 종목의 차이가 MST만큼 크지 않으며, 대부분의 선택 종목이 네트워크 가장자리에 집중되어 있다. 다만 TMFG와 차수 중심성 조합에서는 허브 근처에 위치한 BA와 DIS가 선택에 포함되었다. TMFG는 모든 노드가 최소 3개의 엣지를 가지므로 차수가 3인 자산이 다수 존재하며, 이들 중 허브에 인접한 BA와 DIS까지 주변부로 판정된 것으로 보인다. TMFG와 부분그래프 중심성 조합은 TMFG와 고유벡터 중심성 조합과 유사한 종목 배치를 보였다.

[표 1] 그래프 이론 기반 포트폴리오 전략의 성과

[Table 1] Performance of Graph-Based Portfolio Strategies

Number of Assets (K)	Strategy	Annual Return (%)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
5	MST+Deg	16.60	13.08	1.24	-10.83
5	MST+Eig	32.29	13.22	2.18	-6.40
5	MST+Sub	27.12	12.87	1.93	-9.24
5	TMFG+Deg	18.00	13.08	1.33	-9.24
5	TMFG+Eig	32.60	13.68	2.13	-7.29
5	TMFG+Sub	25.79	13.58	1.76	-5.85
10	MST+Deg	27.38	11.42	2.18	-7.13
10	MST+Eig	25.17	10.63	2.17	-5.70
10	MST+Sub	31.40	11.24	2.49	-7.39
10	TMFG+Deg	22.36	10.74	1.93	-6.06
10	TMFG+Eig	32.06	10.73	2.65	-5.49
10	TMFG+Sub	30.66	11.35	2.41	-6.60
15	MST+Deg	25.05	9.78	2.33	-6.10
15	MST+Eig	18.64	10.46	1.69	-8.37
15	MST+Sub	23.68	10.04	2.17	-6.47
15	TMFG+Deg	28.39	10.80	2.37	-8.00
15	TMFG+Eig	26.34	10.09	2.37	-6.39
15	TMFG+Sub	27.75	10.44	2.40	-5.86

20	MST+Deg	27.15	9.62	2.55	-6.22
20	MST+Eig	24.47	9.99	2.24	-7.58
20	MST+Sub	28.22	9.98	2.54	-6.85
20	TMFG+Deg	25.61	10.16	2.30	-7.43
20	TMFG+Eig	25.82	9.96	2.36	-6.27
20	TMFG+Sub	25.31	9.95	2.32	-6.88

[표 1]의 K=10 기준, TMFG와 고유벡터 중심성 조합이 샤프 비율 2.65로 가장 우수하였다. 이는 [그림 1]에서 확인할 수 있듯이 TMFG와 고유벡터 중심성 조합에서 선택된 종목은 네트워크 가장 자리에 집중되어 있으며, 이는 고유벡터 중심성의 재귀적 특성이 허브로부터 멀리 떨어진 자산을 식별한 결과로 판단된다. MST와 부분그래프 중심성 조합, TMFG와 부분그래프 중심성 조합이 뒤를 이었다. 반면 TMFG와 차수 중심성 조합은 샤프 비율이 1.93으로 가장 낮았다. [그림 1]에서 확인할 수 있듯이, 허브 근처에 위치한 BA와 DIS가 선택에 포함되었기 때문이다.

이 패턴은 K=5에서도 유지되어, 고유벡터 중심성이 MST와 TMFG 모두에서 최상위를 기록하였다. 하지만 K=20에서는 모든 전략 간 샤프비율의 범위가 2.24~2.55로 축소되었으며, MST와 차수 중심성 조합이 가장 우수하였다.

[그림 2]는 K=10 기준, 테스트 기간의 누적 수익률을 나타낸다. TMFG와 고유벡터 중심성 조합과 MST와 부분 그래프 중심성 조합이 높은 수준의 누적 수익률을 보였으며, TMFG와 부분그래프 중심성 조합이 그 뒤를 이었다. 반면 TMFG와 차수 중심성 조합은 최종적으로 가장 낮은 누적 수익률을 기록하였다. 2023년 9~10월 하락기에서 MST와 고유벡터 중심성 조합이 가장 깊은 낙폭을 보인 반면, TMFG와 고유벡터 중심성 조합은 하락을 억제하여, 동일한 고유벡터 중심성이라도 그 래프 구조에 따라 방어력이 달라짐을 확인할 수 있다.



[그림 2] 누적 수익률

[Fig. 2] Cumulative Return

4. 결론

본 연구는 그래프 이론에서 두 가지 필터링 방법(MST, TMFG)과 세 가지 중심성 측도(차수, 고유벡터, 부분 그래프)의 조합을 통한 금융 포트폴리오 성과를 분석하였다. 포트폴리오 내 자산 수를 통제하고 균등 배분을 적용함으로써 주변부 정의의 차이가 성과에 미치는 영향을 비교하였으며, 자산 수 변화에 따른 결과의 일관성을 검증하였다.

분석 결과는 다음과 같다. 첫째, TMFG와 고유벡터 중심성 조합이 자산 수가 10개에서 샤프 비율 2.65로 가장 우수한 위험 조정 성과를 보였으며, 자산 수가 5와 15에서도 일관되게 높은 성과를 나타냈다. 이는 TMFG의 엣지 구조에서 고유벡터 중심성의 재귀적 특성이 영향력 있는 허브로부터 상대적으로 독립적인 자산을 식별한 결과로 판단된다. 이러한 특성은 네트워크 그래프([그림 1])에서도 확인된다. Zhao 외 5인의 연구 [12]에서도 주변부 종목 포트폴리오가 우수한 성과를 보임을 확인하였으며, 본 연구에서 고유벡터 중심성 기반 주변부 전략이 효과적이라는 결과는 주변부 투자 전략의 유효성이라는 맥락에서 이와 부합한다. 둘째, 차수 중심성은 MST와 TMFG 모두에서 자산 수가 증가할수록 성과가 개선되는 경향을 보였다. MST와 차수 중심성의 경우, 자산 수가 5에서 20까지 증가함에 따라 샤프 비율도 증가하였다. 이는 차수 중심성이 대규모 포트폴리오에 적합하다는 Wang 외 4인의 연구 [10]의 결과와 일관된다.

본 연구의 한계점과 향후 연구 방향은 다음과 같다. 첫째, 분석 기간이 특정 시장 국면에 해당할 가능성이 있으므로, 향후 연구에서는 다양한 시장 환경을 포함한 장기 분석이 필요하다. 둘째, 본 연구는 소규모 유니버스를 대상으로 수행되었으므로, 대규모 유니버스에서의 검증이 요구된다. 셋째, 본 연구는 종목 선택 효과를 균등 배분을 적용하였으나, 가중치 배분 방식에 따른 성과 차이에 대한 추가적인 분석이 필요하다.

References

- [1] E. J. Elton, M. J. Gruber, S. J. Brown, and W. N. Goetzmann, *Modern Portfolio Theory and Investment Analysis*, 9th ed., John Wiley & Sons, 2014.
- [2] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77-91, Mar. 1952, doi: 10.2307/2975974.
- [3] M. López de Prado, "Building diversified portfolios that outperform out of sample," *The Journal of Portfolio Management*, vol. 42, no. 4, pp. 59-69, 2016, doi: 10.3905/jpm.2016.42.4.059.
- [4] R. N. Mantegna, "Hierarchical structure in financial markets," *European Physical Journal B*, vol. 11, no. 1, pp. 193-197, Sep. 1999, doi: 10.1007/s100510050929.

- [5] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna, "A tool for filtering information in complex systems," *Proceedings of the National Academy of Sciences*, vol. 102, no. 30, pp. 10421-10426, Jul. 2005, doi: 10.1073/pnas.0500298102.
- [6] G. P. Massara, T. Di Matteo, and T. Aste, "Network filtering for big data: Triangulated maximally filtered graph," *Journal of Complex Networks*, vol. 5, no. 2, pp. 161-178, 2016, doi: 10.1093/comnet/cnw015.
- [7] T. Aste, "Information filtering networks: Theoretical foundations, generative methodologies, and real-world applications," Apr. 2025, doi: 10.48550/ARXIV.2505.03812.
- [8] F. Pozzi, T. Di Matteo, and T. Aste, "Spread of risk across financial markets: Better to invest in the peripheries," *Scientific Reports*, vol. 3, no. 1665, pp. 1-7, Apr. 2013, doi: 10.1038/srep01665.
- [9] G. Peralta and A. Zareei, "A network approach to portfolio selection," *Journal of Empirical Finance*, vol. 38, pp. 157-180, Sep. 2016, doi: 10.1016/j.jempfin.2016.06.003.
- [10] G.-J. Wang, H. Huai, Y. Zhu, C. Xie, and G. S. Uddin, "Portfolio optimization based on network centralities: Which centrality is better for asset selection during global crises?," *Journal of Management Science and Engineering*, vol. 9, no. 3, pp. 348-375, Sep. 2024, doi: 10.1016/j.jmse.2024.04.001.
- [11] E. Estrada and J. A. Rodríguez-Velázquez, "Subgraph centrality in complex networks," *Physical Review E*, vol. 71, no. 5, Art. no. 056103, May 2005, doi: 10.1103/PhysRevE.71.056103.
- [12] L. Zhao, G.-J. Wang, M. Wang, W. Bao, W. Li, and H. E. Stanley, "Stock market as temporal network," *Physica A*, vol. 506, pp. 1104-1112, Sep. 2018, doi: 10.1016/j.physa.2018.05.039.