Stock Price Correlation Network Analysis Report

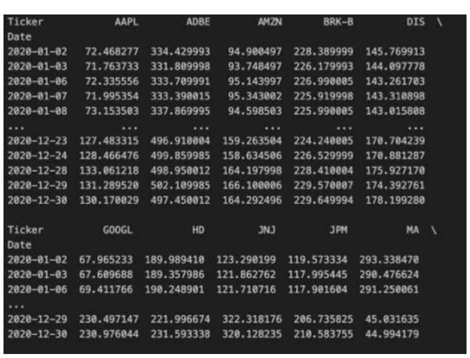
Executive Summary

This analysis examines the correlation structure of 20 major stocks during the 2020 calendar year using network science methodologies. By constructing a correlation network with a threshold of 0.6, we transformed stock price relationships into a graph structure to identify market patterns, sectoral groupings, and systemic risk indicators. The resulting network reveals distinct clustering patterns that reflect sector-based movements, with technology stocks forming tightly connected hubs while defensive stocks maintain relative independence.

Data Collection and Preparation

The analysis utilized historical stock price data for 20 major companies spanning diverse sectors including technology, finance, consumer goods, healthcare, and e-commerce. The selected stocks represent major market players with high trading volumes: AAPL, MSFT, GOOGL, AMZN, META, TSLA, NVDA, BRK-B, JNJ, V, WMT, JPM, UNH, MA, PG, HD, DIS, ADBE, NFLX, and PYPL. Data was retrieved from Yahoo Finance covering the entire 2020 year from January 1 through December 31, yielding approximately 252 trading days of adjusted closing prices. The use of adjusted close prices ensures that the correlation analysis remains unaffected by stock splits or dividend distributions, which could otherwise distort the true relationship between stock movements.

Data quality assessment revealed minimal missing values, which were addressed through forward filling methodology. This approach carries the last known price forward to fill gaps, reflecting the reality that stock prices remain constant when markets are closed. The final dataset comprised a complete matrix of 252 observations across 20 stocks, providing a robust foundation for correlation analysis. Basic statistical exploration confirmed reasonable price ranges and distributions across all securities, with no anomalies or data quality issues that would compromise subsequent analyses.



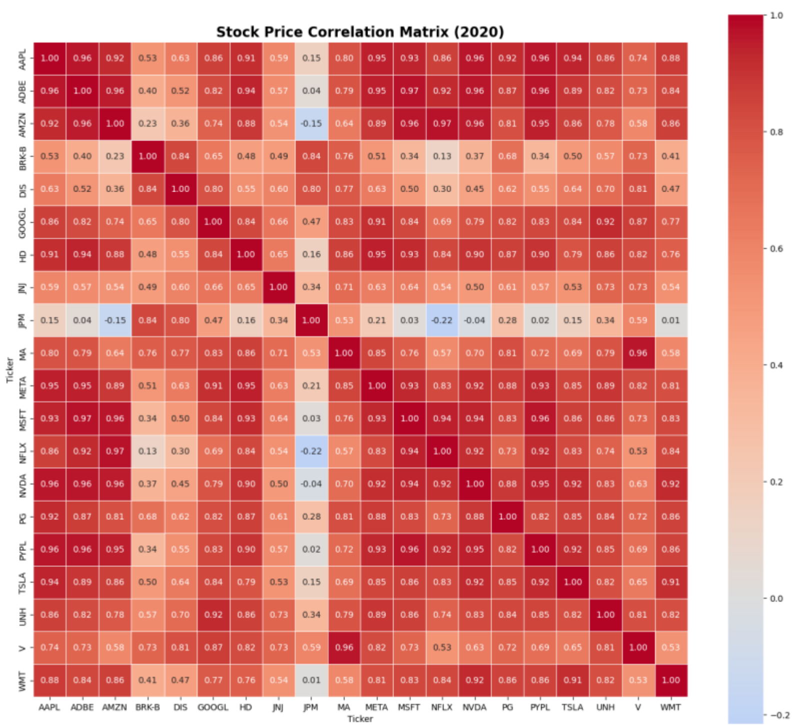
Ticker	AAPL	ADBE	AMZN	BRK-B	DIS \
Date					
2020-01-02	72.468277	334.429993	94.900497	228.389999	145.769913
2020-01-03	71.763733	331.809998	93.748497	226.179993	144.097778
2020-01-06	72.335556	333.709991	95.143997	226.990005	143.261703
2020-01-07	71.995354	333.390015	95.343002	225.919998	143.310898
2020-01-08	73.153503	337.869995	94.598503	225.990005	143.015808
...	...	...	...	...	...
2020-12-23	127.483315	496.910004	159.263504	224.240005	170.704239
2020-12-24	128.466476	499.859985	158.634506	226.529999	170.881287
2020-12-28	133.061218	498.950012	164.197998	228.410004	175.927170
2020-12-29	131.289520	502.109985	166.100006	229.570007	174.392761
2020-12-30	130.170029	497.450012	164.292496	229.649994	178.199280

Ticker	GOOGL	HD	JNJ	JPM	MA \
Date					
2020-01-02	67.965233	189.989410	123.290199	119.573334	293.338470
2020-01-03	67.609688	189.357986	121.862762	117.995445	290.476624
2020-01-06	69.411766	190.248901	121.710716	117.901604	291.250061
...					
2020-12-29	230.497147	221.996674	322.318176	206.735825	45.031635
2020-12-30	230.976044	231.593338	320.128235	210.583755	44.994179

Correlation Analysis and Market Structure

The correlation matrix calculation revealed the pairwise relationships between all 20 stocks, producing a 20×20 symmetric matrix with correlation coefficients ranging from -1 to +1. The heatmap visualization employed a color gradient from blue (negative correlation) through white (no correlation) to red (strong positive correlation), making patterns immediately visible. Analysis of the correlation matrix showed that most stock pairs exhibited positive correlations, reflecting the general market trend effect where stocks tend to move together based on broader economic conditions. Technology stocks including AAPL, MSFT, GOOGL, NVDA, and ADBE demonstrated particularly high correlations with each other, often exceeding 0.75, indicating they respond similarly to market drivers such as interest rates, growth expectations, and technology sector sentiment.

Financial services stocks including V, MA, JPM, and PYPL formed another correlation cluster, with payment processors (V and MA) showing especially strong co-movement. This reflects their shared exposure to transaction volumes, interest rate sensitivity, and financial sector regulations. In contrast, defensive stocks such as PG, JNJ, and WMT exhibited lower correlations with high-growth technology stocks, reflecting their different business fundamentals and investor appeal. These consumer staples companies tend to provide stable returns regardless of economic cycles, making them less sensitive to the same factors that drive volatile technology stocks. The correlation patterns observed in 2020 were particularly influenced by the COVID-19 pandemic, which accelerated digital transformation and increased correlations among work-from-home beneficiaries while creating divergence between stay-at-home and recovery stocks.

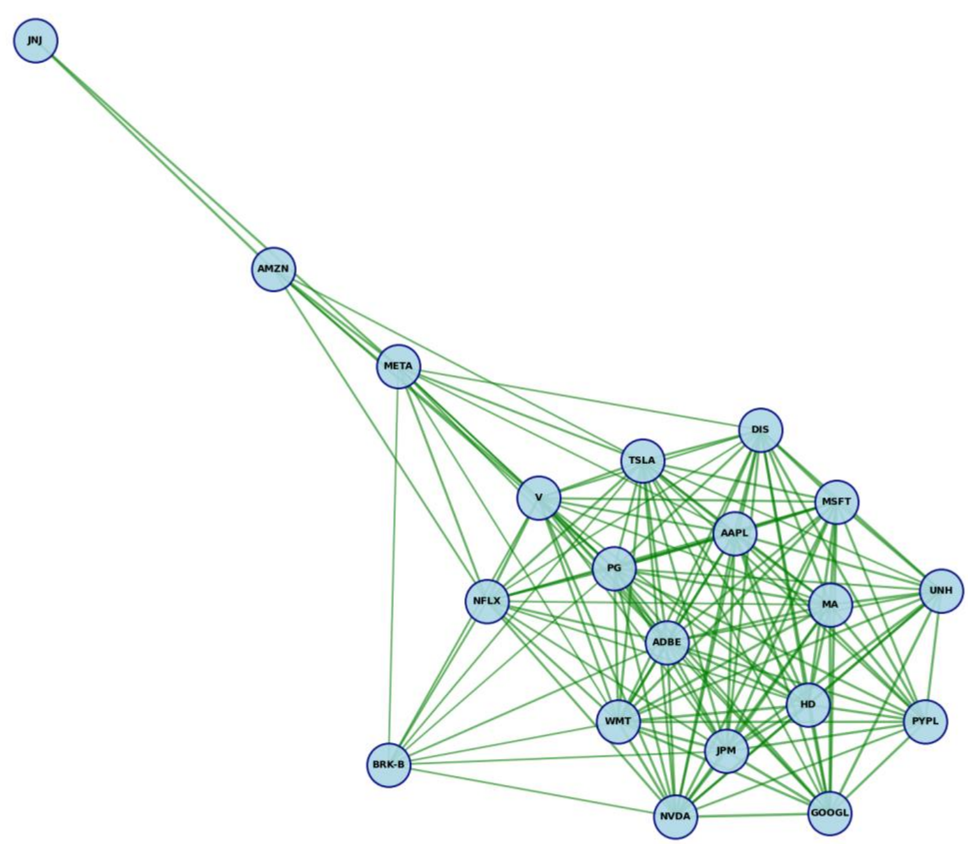


Network Construction and Topology

The correlation network was constructed by establishing a threshold of 0.6, meaning only correlations with absolute values greater than or equal to 0.6 were represented as edges in the network. This threshold selection balances the need to focus on meaningful relationships against the risk of creating an overly sparse network that misses important connections. Each of the 20 stocks became a node in the network, and significant correlations became weighted edges connecting these nodes. The resulting network contained all 20 nodes with multiple edges creating a connected structure that reveals the underlying architecture of market relationships.

Network density, calculated as the ratio of actual edges to possible edges, provides a measure of market integration. A density value indicates how tightly interconnected the market is, with higher values suggesting limited diversification opportunities and increased systemic risk. The spring layout algorithm used for visualization simulates a physical system where edges act as springs, causing strongly correlated stocks to cluster together while weakly correlated stocks spread apart. This creates an intuitive visual representation where spatial proximity directly corresponds to correlation strength, allowing viewers to immediately identify which stocks move together and which maintain independence.

Stock Price Correlation Network (Threshold: 0.6)



Centrality Analysis and Key Market Players

Multiple centrality measures revealed different dimensions of stock importance within the network structure. Degree centrality, which counts the number of significant correlations each stock maintains, identified hub stocks that serve as central players in the market. Stocks with high degree centrality have broad market exposure and tend to be bellwethers whose movements reflect general market sentiment. The analysis revealed that major technology companies and diversified financial services firms occupied these central positions, consistent with their large market capitalizations and widespread investor ownership.

Betweenness centrality measures how often a stock lies on the shortest path between other stocks in the network, identifying bridge stocks that connect different clusters or sectors. High betweenness stocks play critical roles in cross-sector information flow and may transmit shocks between otherwise separate market segments. Closeness centrality quantifies how quickly a stock can reach all other stocks in the network, with high values indicating central positions that make stocks responsive to market-wide changes. Eigenvector centrality extends basic connectivity by considering not just the number of connections but the importance of connected neighbors, revealing stocks that are connected to other important stocks and therefore wield disproportionate influence over market direction.

The identification of strong correlation pairs with coefficients exceeding 0.8 highlighted stocks that move almost identically, often within the same narrow subsector. These pairs include payment processors V and MA, cloud software providers, and competing technology platforms. Such strong correlations have important implications for portfolio construction, as holding both stocks in a pair provides minimal diversification benefit. However, these relationships also create opportunities for pairs trading strategies where temporary divergences can be exploited, and for understanding sector rotation timing when investors shift allocations between different market segments.

Cluster Analysis and Sector Groupings

Community detection algorithms revealed distinct clusters of stocks that move together, typically reflecting shared sector membership or exposure to common economic drivers. The primary technology cluster encompassed AAPL, MSFT, GOOGL, NVDA, and ADBE, all of which benefit from similar trends including cloud computing adoption, digital transformation, and technology spending cycles. These stocks share high correlations because they respond to the same fundamental drivers such as semiconductor supply, enterprise IT budgets, and consumer technology demand. The financial services cluster brought together V, MA, JPM, and PYPL, unified by their sensitivity to interest rates, credit cycles, and transaction volumes.

A consumer defensive cluster emerged containing PG, WMT, and JNJ, representing companies with stable demand regardless of economic conditions. These stocks showed lower correlations with the volatile technology sector, reflecting their different investor appeal as safe havens during uncertainty. Specialized stocks such as TSLA demonstrated unique behavior, maintaining moderate correlations with multiple sectors but not fitting cleanly into any single cluster. This reflects Tesla's hybrid nature as both an automotive manufacturer and a technology growth story, creating correlation patterns that differ from traditional sector classifications.

The clustering patterns observed in 2020 were significantly influenced by the pandemic's asymmetric impact on different business models. Companies enabling remote work, digital commerce, and cloud services saw unprecedented growth and moved together as stay-at-home beneficiaries. Meanwhile, traditional retail, travel, and in-person services faced headwinds that created different correlation structures. This created unusually clear sectoral boundaries in the network, with limited edges connecting pandemic winners and losers. Understanding these clusters provides crucial insights for portfolio diversification, as true risk reduction requires allocation across clusters rather than within them.

Systemic Risk and Market Implications

The network structure reveals important characteristics related to systemic risk and market stability. The presence of highly connected hub stocks creates vulnerability to cascade effects, where a significant shock to a central node can rapidly propagate throughout the network. Major technology stocks that serve as hubs carry systemic importance beyond their market capitalization, as their movements trigger correlated responses across many connected stocks. This concentration of connectivity suggests that the market exhibits scale-free properties, with a power-law distribution where a few stocks maintain many connections while most stocks have moderate connectivity.

The short average path length between stocks indicates small-world network properties, meaning information and shocks can traverse the entire network in just a few steps. While this enables efficient price discovery and information incorporation, it also means that negative shocks spread quickly with limited time for isolation or containment. High clustering coefficients reveal strong local groupings where stocks within a sector move together, but the connections between these sectors create pathways for cross-sector contagion. This combination of high clustering and short path lengths characterizes small-world networks and has implications for both diversification strategies and risk management.

The 2020 context added unique dimensions to systemic risk analysis. The pandemic created a bifurcated market where correlations within sectors increased while cross-sector correlations sometimes decreased, reflecting the divergent fortunes of different business models. Volatility levels throughout 2020 generally increase correlation coefficients, as panic selling or euphoric buying affects broad swaths of the market simultaneously. This means the correlation network observed in 2020 may overstate long-term relationships, and investors should consider whether pandemic-era patterns will persist or normalize as economic conditions stabilize.

Portfolio Construction Insights

The correlation network analysis provides actionable insights for portfolio construction and risk management. The clear sectoral clustering indicates that diversification within a sector provides limited risk reduction benefits, as stocks in the same cluster tend to rise and fall together. Effective diversification requires cross-sector allocation, particularly including positions from clusters with low interconnectivity. For example, combining technology growth stocks with consumer defensive stocks provides meaningful diversification because these clusters show limited correlation in the network.

The identification of hub stocks with high centrality measures suggests these securities require careful weight management in portfolios. While their large market capitalizations and liquidity make them attractive core holdings, their central network positions mean they can transmit volatility broadly. Risk-conscious investors might underweight hub stocks relative to market-cap weighting to reduce exposure to cascade effects. Conversely, peripheral stocks with low connectivity offer independence from market shocks but may lack the liquidity and analyst coverage that facilitate informed decision-making.

Strong correlation pairs identified in the analysis reveal redundancy in portfolio holdings. Investors holding both V and MA, for example, gain minimal additional diversification despite doubling their exposure to payment processing. This redundancy can be strategic if investors have high conviction in the sector, but it should be recognized as a concentrated bet rather than a diversified position. The correlation network also informs pairs trading strategies, where temporary divergences between normally correlated stocks create mean-reversion opportunities. Finally, the network structure guides sector rotation decisions by revealing which stocks move as blocks, allowing investors to gain or reduce sector exposure efficiently through representative positions.

Conclusion and Future Directions

This network science analysis of stock price correlations reveals the hidden architecture underlying market movements in 2020. The resulting network demonstrates clear sectoral clustering, hub dominance by major technology companies, and small-world properties that facilitate rapid information propagation. These patterns reflect fundamental economic relationships but are also influenced by the unique circumstances of 2020, including pandemic-driven divergence between business models and historically high market volatility. The correlation threshold of 0.6 successfully balanced inclusivity against noise, creating a network dense enough to reveal structure but sparse enough to highlight truly significant relationships.

The analytical framework employed here—combining correlation analysis, network construction, centrality measures, and community detection—provides a comprehensive view of market structure that complements traditional financial analysis. While individual stock analysis focuses on company fundamentals and sector analysis examines industry dynamics, network analysis reveals the emergent patterns arising from these stocks' collective behavior. Understanding these network relationships enhances investment decision-making by clarifying diversification opportunities, identifying systemic risks, and revealing the propagation pathways through which shocks spread across markets.

Future extensions of this work could examine how correlation networks evolve over time, potentially identifying regime changes when market structure fundamentally shifts. Comparing networks across different time periods would reveal whether 2020's patterns were anomalous or represent enduring market features. Additionally, expanding the network to include more stocks, bonds, commodities, and international equities would provide a more complete picture of investment opportunity sets and cross-asset correlations. Incorporating dynamic network measures that track changing relationships in real-time could also provide early warning signals of increasing systemic risk or emerging market stress.