NORTHERN TRUST
ASSET MANAGEMENT

RESEARCH INSIGHTS

ALPHA YOUR MANAGER MIGHT BE MISSING

Understand how AI-based models can unlock alpha signals found in company networks and access a detailed framework for evaluating your equity manager’s approach.

Uncovering Alpha In The Networked Economy

SHINING A LIGHT ON THE HIDDEN CONNECTIONS BETWEEN COMPANIES

Gijsbert De Lange
Investment Solutions Expert

Jan Rohof, CAIA
Director of Quantitative Strategies, APAC

Guido Baltussen, PhD
Head of Quantitative Strategies, International

Sri Kancharla, CFA
Head of Quantitative Portfolio Management

In the ever-evolving landscape of quantitative investing alternative data and advanced modelling techniques offer novel and compelling alpha opportunities. Resulting from the combination of low cost computing power and increased data accessibility, quants can now apply the most advanced algorithms to outperform the market and their peers. As a result, the frontier of quant investing has shifted. Today, the most compelling opportunities lie in the nuanced, the uncorrelated, and the hidden. Among these, network-based signals, that shine light on hidden connections among and between companies, stand out as a particularly rich and underexploited source of alpha.

EXECUTIVE SUMMARY

This paper explores the transformative potential of networks in quantitative asset management, emphasizing how they create opportunities for alpha. By leveraging advances in Artificial Intelligence (AI), particularly in Natural Language Processing (NLP) and Large Language Models (LLMs), new value-additive insights are discussed. This paper provides insightful observations on how networks inform investment decisions. Specifically, it demonstrates how networks have the potential to improve factor-based strategies by uncovering hidden opportunities for idiosyncratic returns in networks of companies.

THE DATA AND COMPUTING POWER EXPLOSION AND THE RISE OF AI

The last decade has seen an explosion in data availability and computational power. Corresponding data costs through storage and computational power have plummeted. While the storage cost per gigabyte was $10,000 in 1990, today it is only $ 0.01. Cloud computing per hour was $0.10 in 2011 vs $0.005 in 2025[1] .

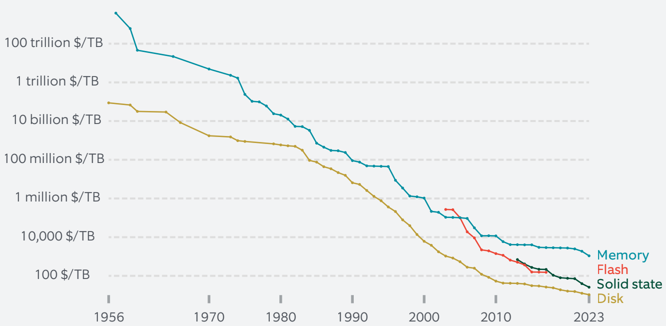
[1] **Sources:** Backblaze, Statista.

EXHIBIT 1:

Trends In Computing Cost And Power

Historical price of computer memory and storage

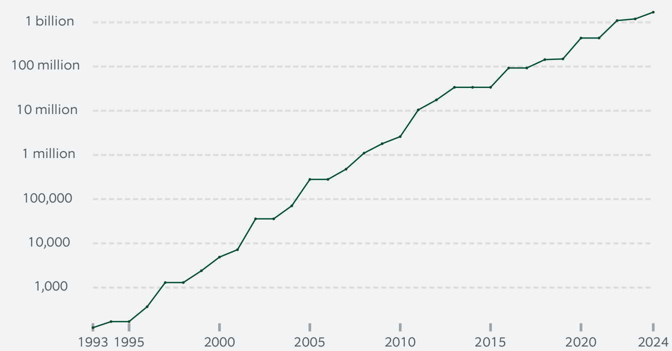
This data is expressed in US dollars per terabyte (TB), adjusted for inflation. "Memory" refers to random access memory (RAM), "disk" to magnetic storage, "flash" to special memory used for rapid data access and rewriting, and "solid state" to solid-state drives (SSDs).



Note: For each year, the time series shows the cheapest historical price recorded until that year. This data is expressed in constant 2020 US$.

Computational capacity of the fastest supercomputers

The number of floating-point operations carried out per second by the fastest supercomputer in any given year. This is expressed in gigaFLOPS, equivalent to 10^9 floating-point operations per second.



Source: Northern Trust Asset Management, LHG: John C. McCallum (2023); U.S. Bureau of Labor Statistics (2024) RHG: Dongarra et al. (2024).

Breakthroughs in generative AI, particularly in LLMs have transformed our ability to extract meaning from unstructured text. In 2019 GPT2 was launched, with its human-like text generation, followed in 2022 by ChatGPT, that allowed for reinforcement learning from human feedback. Estimates state that approximately 50-60% of global data is text-based. This figure increases further to 70–80% when we consider data relevant to enterprises, including annual reports, earnings calls, corporate filings, news articles[2] . As an example, the average annual report comprises two-thirds of text based information. NLP models, which enable computers to understand, interpret and generate human language, allow us to turn this unstructured text data into structured data, that can be translated into stock selection signals for our quantitative investment strategies.

COMPANY NETWORKS: ANALYZING A COMPANY'S COMPETITIVE LANDSCAPE AND SUPPLY CHAIN

One area in particular where we make good use of the explosion in data availability and computing power is the construction of groups, or networks, of companies and their use in evaluating investment information. Instead of basing an investment decision solely on a company itself, we now also assess information about other companies that the company in question is connected to. The model we use for these connections is that of a network. The networks we build are collections of companies, where the nodes are companies and the connections between them are the edges. An example is a supply chain network: each node represents a company and the edges are the business relationships between them – like supplying parts or distributing products. When constructing a supply chain network we use data from primary sources only. These can be disclosures of companies through annual filings, company announcements, or company websites. The data is used to recognize relationships between suppliers and customers and direct the edges for the flow of goods or services. We can assign

Mechanism of constructing networks

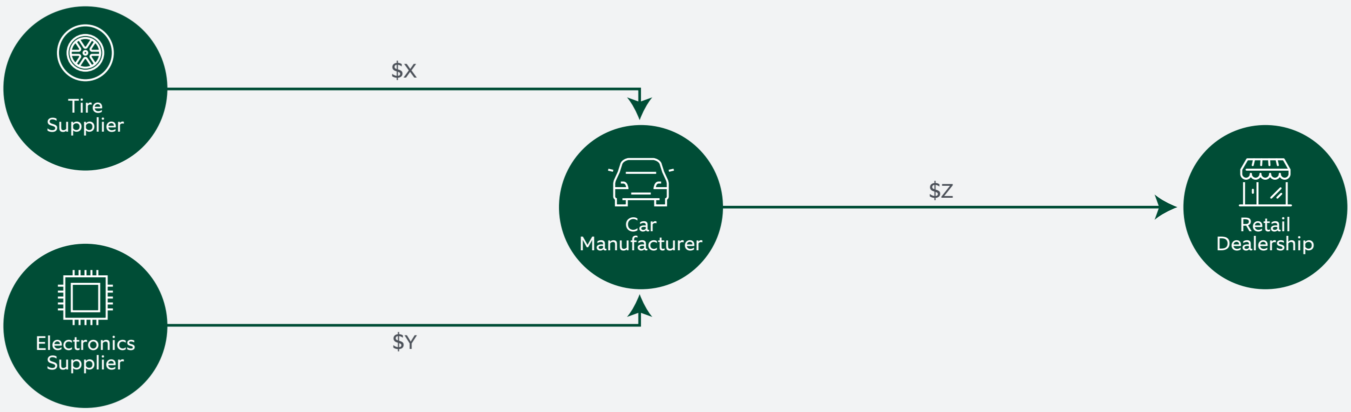
In a network, the basic building blocks used to model relationships and structures are nodes and edges. ***Nodes*** *are the individual entities in the network, and the edges are the connections or relationships between nodes.* ***Edges*** *can be directed (with a direction) or undirected, and can be weighted (carrying a value) or unweighted. Some of the concepts we use when analysing networks are degree (number of edges connected to a node), path (sequence of edges connected to a node), and clustering (tendency of nodes to form tightly-knit groups).*

[2] **Sources:** IBM, Gartner.

weights to the edges for the total transaction value between the two companies or a percentage expressing the relative importance of the relation, from either the customer's or the supplier's perspective. The node of suppliers with diversified sales over many customers have higher degrees than those with more concentrated customers. If we are interested not only in 1st order relationships but also 2nd or higher orders, then we will analyze paths in the network. Exhibit 2 below shows such a network based on the supply-chain of a car manufacturer.

EXHIBIT 2:

Supply Chain Network: Nodes and Edges



Source: Northern Trust Asset Management.

Another type of network is that of technology-linked companies. To build this we use patent ownership data and the Cooperative Patent Classification (CPC)[3] system . We classify the patents into different technology classes using the CPC. For each company with patents, we construct a technology vector with the firm's proportional share of each technology class. Finally, we calculate a proximity score for every two companies. As with the supply chain network, we now have a network that identifies not only if two companies are linked, but also with weights to express the importance of the link to each company. Other than the supply chain network, that has directed edges – from supplier to customer, the technology linked network is undirected and the weights of the edges are two-fold: the relative importance of the link can differ between the two companies.

In general, any type of connection between companies can be used to construct a network: for example similarity of fundamentals, correlations of stock returns, shared ownership, geographical sales, exposure to climate risks, parent-subsidiary. There are many academic studies[4] on the use of company networks. Some of the more interesting networks are constructed using text data as we shall discuss in the next section.

[3] The Cooperative Patent Classification (CPC) is a comprehensive system for classifying patents, developed jointly by the European Patent Office (EPO) and the United States Patent and Trademark Office (USPTO).

[4] We refer interested readers to, for example: "Economic links and predictable returns", L Cohen, A Frazzini, The Journal of Finance, 2008; "Text-based network industries and endogenous product differentiation", G Hoberg, G Phillips, Journal of political economy, 2016; "Supply chain network structure and firm returns", J Wu, J Birge, Working Paper, 2014; "Competition links and stock returns", A Eisdorfer, e.a., The Review of Financial Studies, 2022; "Technological links and predictable returns", C Lee, e.a., Journal of Financial Economics, 2019.

TEXT-BASED NETWORKS: A NEW LENS ON COMPANY NETWORKS

One of the most exciting applications of NLP is the construction of text-based networks. Over the last 10 years the ability to accurately extract information from text has evolved enormously. First the release of Word2Vec by Google enabled semantic relationships to be captured (for example king – man + woman = queen) and gave a big impetus to text-based network building. More recently the arrival of LLMs trained on huge volumes of text-data enables contextual interpretation and automated processing of corporate filings with less need for human error checking. Of course LLMs have well-known pitfalls too, for example when LLMs produce entirely made-up responses (hallucinations) or when their outputs are biased due to the inherent biases in their training data, and these need to be managed carefully.

A prominent example of a text-based network is a competition network, which shows how companies are linked to their competitors. Parsing corporate filings, we identify the competitor relationships. For example, Pepsi and Coca Cola's filings (see Exhibit 3 below) reveal that they view each other as their main competitor. Both are beverage giants, but their business models differ significantly. Pepsi is more diversified, while Coke is more focused. This is reflected in their filings, that reveal both shared competitors in the beverage business and distinct competitors.

EXHIBIT 3:

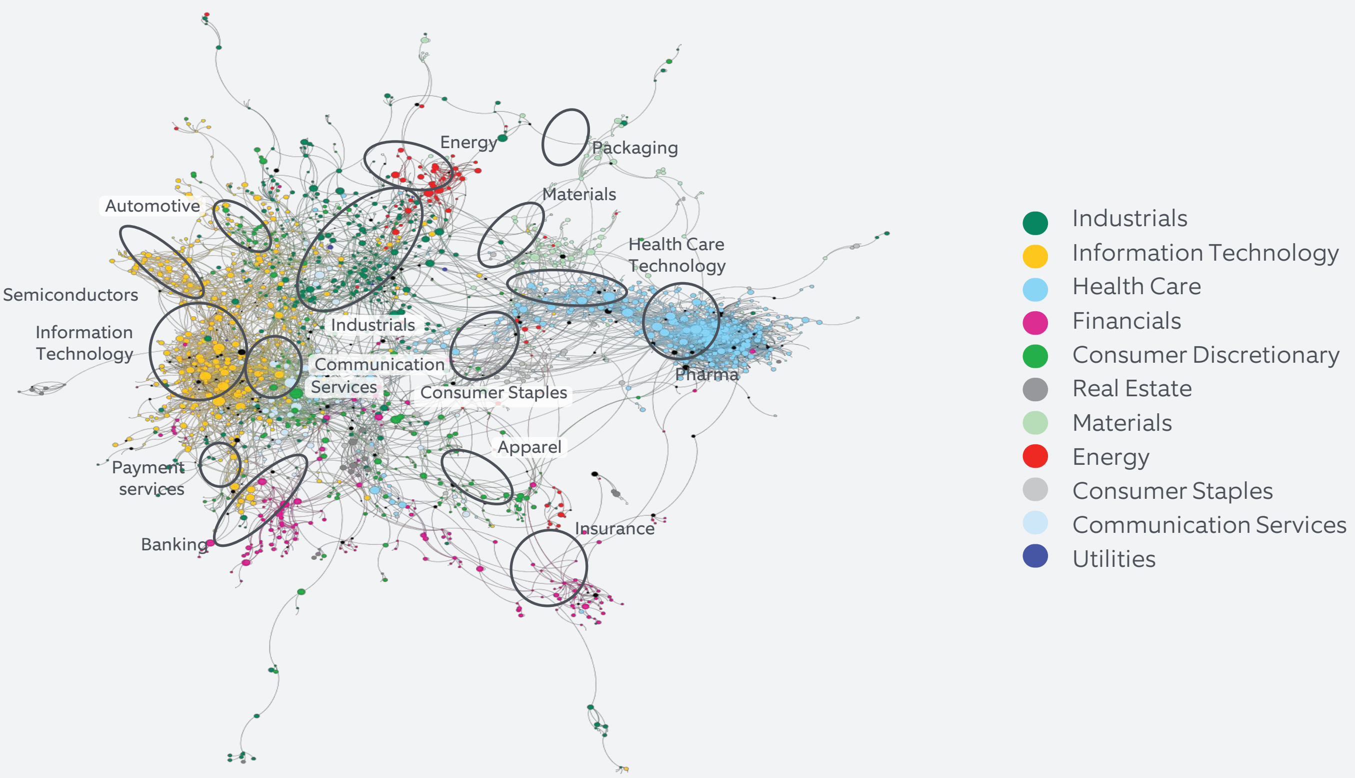
Extracts from Corporate Filings of Major Beverage Companies

Source: Northern Trust Asset Management, PepsiCo, The Coca Cola Company.

We use filings to construct a competitor network with all companies. This network is depicted in Exhibit 4 below. Nodes again are companies and edges denote a mention in its filing by one company of the other company as competitor. The edges are directed, and if both companies mention each other as competitor the edge directs both ways. Here we assign weights to the edges, instead to the nodes. The more a company is mentioned as competitor the higher the weight. In the below chart the size of the nodes depicts the weights. Observe how some nodes are clustered in groups of mutual competitors. The colors identify the 11 GICS sectors. Interestingly, competition links exist between companies from different industries. For example, Information Technology (IT) companies are often identified as competitors by companies outside of the IT sector. The reverse holds true as well, for example Amazon is classified as Consumer Discretionary, but through its subsidiary AWS competes with IT companies. Clearly competition doesn't stop at the GICS boundaries.

EXHIBIT 4:

Example of a Competition Network in the MSCI ACWI Universe



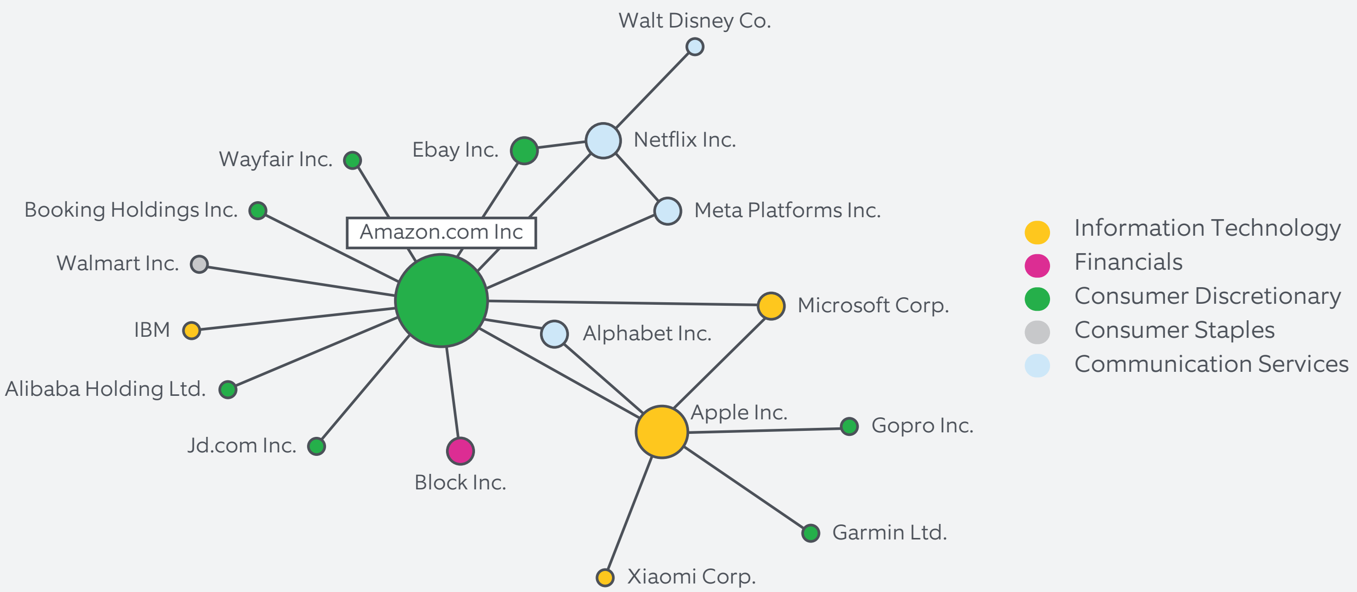
Source: Northern Trust Asset Management

Another example of a text-based network is that of peers. Here we look at the description companies' business models by screening annual reports of thousands of companies in our investable universe. For each company's business description a text vector is created, then a similarity score is calculated with the text vectors of all other companies in the investable universe. The result is an undirected network of business peers. The weights of the edges represent a similarity score measuring how closely linked two companies are.

This network also has clusters of similar companies. In the example of Exhibit 5 below we again observe Amazon having peers beyond Consumer Discretionary.

EXHIBIT 5:

Example of a More Detailed Network Of Business Peers



Source: Northern Trust Asset Management

DATA COVERAGE OF TEXT-BASED NETWORKS

While corporate filings are standardized in the US, for example they have clear sections on the business' competitive landscape, we can construct networks using corporate filings and large language models in all markets, even where reporting is less standardized such as in emerging markets. Knowledge of global reporting customs and language translation models allows for network creation across the globe. There are almost 8,300 public companies in MSCI ACWI IMI, that in total have approximately. 230,000 past and current filings in twenty-one languages. This is where proven and thoroughly vetted LLMs become useful for the translation of filings. The table (Exhibit 6) below shows the coverage of the developed and emerging markets universes we achieve in three of our networks we discussed above. The coverage is near complete for the peer network and ranges from fair to very good in the other networks. Over time we see coverage improving.

EXHIBIT 6:
ACWI IMI Coverage

	# Names	Peer Network	Competitor Network	Supply Chain Network	Technology-Link Network
MSCI World	1,325	100%	89%	81%	73%
MSCI World small	3,853	99%	61%	56%	51%
MSCI EM	1,203	100%	51%	59%	53%
MSCI EM small	1,896	97%	31%	50%	28%

Source: Northern Trust Asset Management. As of June 30, 2025.

ADVANTAGES OF NETWORKS OVER TRADITIONAL (INDUSTRY) CLASSIFICATIONS

We showed in the discussions of the competitor network and the business peer network that when using NLP to extract information from corporate filings, we identify connections between companies from different industries. The same holds true for other networks, companies often have connections that one would not suspect when using traditional industry classification systems only. This is a key advantage of using text-based networks, we get a more granular picture of the corporate landscape. Other advantages over standard industry classifications are the weights we can assign to the edges, directions and paths and clusters we can observe. Finally, in the networks we observe a unique set of peers for each company, if companies A and B are linked to company C this does not necessarily imply that B and C are linked to each other. With all these advantages, text-based networks offer us a rich source of information about the connections between companies in the corporate landscape - this is where investment value lies.

NETWORK ALPHA IN PRACTICE

Traditionally, factors have been captured using financial statement information of every company individually. This information is still important, but it overlooks the networked nature of the economy and of financial markets. In order to stay ahead of the curve, quants developed enhanced factor definitions and created signals that offer fundamentally new insights. Signals that incorporate the information from networks of companies can offer improved returns by refining the implementation of traditional styes or by accessing information that had previously been unattainable for quantitative investment approaches. Network-based impulse signals help us understand a company's position within its ecosystem—how it is perceived by peers, how it interacts with competitors, and how information about it propagates through the market. The use of impulse signals as alpha signals is fundamentally inspired and systematically applied. It blends deep economic insight with rigorous data science.

Signals through the network

*How do we capture the investment value of networks? For that we need to construct one or more **impulse signals.** Technically an impulse signal on a network is a signal that is zero at all nodes except one, where it takes a nonzero, usually unit, value. It's like "poking" one node to see how the signal or effect spreads through the network over time. Impulse signals are useful to probe or test the network's response, similar to how we would tap a surface to understand its structure.*

COMPETITIVENESS AS ALPHA SIGNAL

One network-based alpha signal comes from using our competitor networks to measure the competitiveness of companies. We proxy competitiveness by the number and quality of competition mentions by other companies. These can be thought of as the impulse signal in this network. The size, or weight of nodes represents the competitiveness of the company, bigger nodes for more direct competitors. Cross-referencing, by being mentioned as a competitor by companies that are themselves formidable competitors, adds more weight. We adjust the weight of each company by removing unintended biases such as the region they are from or the size of the company. The final weight captures idiosyncratic competitiveness—information about a company not explained by other alpha drivers nor fundamentals. This makes this alpha signal uncorrelated and complementary to traditional factors.

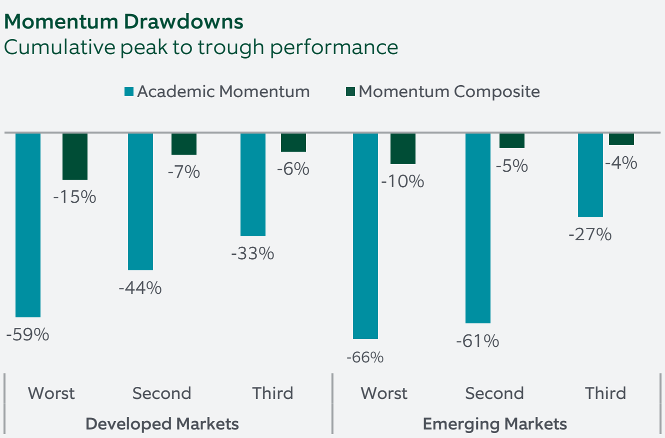
NETWORK MOMENTUM: INFORMATION IN MOTION

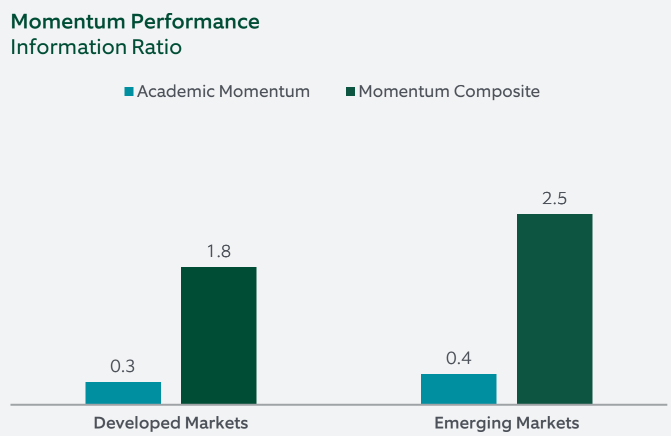
Momentum is one of the most pervasive anomalies in finance, but it is also risky and prone to crashes. Network momentum offers a solution. By identifying price trends as an impulse signal and tracking how they propagate through networks of connected companies, we create a more stable and predictive alpha signal. We apply this methodology across various networks: business peers, competitors, customers, and technology-linked companies. Each captures a different dimension of inter-company relationships, behavior of the impulse signal, which we then aggregate into one alpha signal.

The alpha signals are risk-managed by residualizing for beta exposures on market and stock level. Our empirical validation shows that such a risk-efficient approach leads to significantly reduced 'crash risk' by way of smaller and shorter drawdowns. In doing so, the pain ratio of the Momentum factor becomes less painful. Performance metrics in Exhibit 7 below show superior risk-adjusted returns compared to academic price momentum. This is momentum reimagined for the networked age.

EXHIBIT 7:

Drawdowns and Information Ratio of Academic Versus Network Momentum





Source: Northern Trust Asset Management. Based on long-short backtests, weighted by alpha times the square root of market cap for the MSCI All Country World Investible Market Index universe. Period: January 1995 - December 2024. For illustrative purposes only.

VALUATION IN CONTEXT: NETWORK-RELATIVE ASSET-BASED VALUE

Classic book-to-price metrics often favor old economy companies and penalize innovative companies. Quantitative strategies often apply thorough risk management frameworks like sector neutralization to address some of these shortcomings. What's more, by adjusting for off-balance-sheet items and accounting for intangibles we can further enhance our valuation framework by adding human and knowledge capital to the balance sheet. Using networks enables us to go a step further by comparing each company's valuation to that of its peers and competitors. As the relative cheapness (expensiveness) of the focal company is (partially) a mispricing that will correct over time, we arrive at a more accurate, forward-looking valuation measure. It captures unique information and improves portfolio balance. It also reflects a deeper truth: value is not absolute—it is contextual. The table below (Exhibit 8) compares the Information Ratios of classic, Academic Value with that of our Network-based Value signal.

EXHIBIT 8:

Information Ratio of Academic Versus Network Value

	Academic Value	Network Value
MSCI World IMI	0.14	1.34
MSCI Emerging Markets IMI	0.43	1.02

Source: Northern Trust Asset Management. Back tested long-short risk-adjusted returns (Information Ratio) over the period January 1995 – May 2025. Academic Value is defined as book-to-price. Network Value is based on NTAM's proprietary network-relative asset-based value signal. For illustrative purposes only..

CONCLUSION: THE FUTURE IS NETWORKED

As quantitative investors, we are always seeking the next edge. In a world awash with data, the challenge is not scarcity but discernment. Network-based signals offer a powerful new lens—one that is grounded in economic theory, enabled by modern technology, and validated by empirical results. We have shown the opportunities in networks, the technology and investor knowledge needed to extract the information efficiently, and how to turn the opportunity into systematic alpha based on building a competitor network. Text-based networks offer rich and nuanced insights into how companies are connected. These insights allow us to see companies not as isolated entities, but as nodes in a complex, dynamic system. They help us understand how information flows, how sentiment spreads, and how value is perceived. In doing so, they open up new frontiers for alpha generation.

The future of quant investing is not just about better models or faster compute. It's about better context. And in that future, networks allow us to make even better informed and risk-efficient investment decisions.

Artificial Intelligence (AI): AI refers to computational systems designed to perform tasks that typically require human intelligence, such as pattern recognition, decision-making, and prediction. In investment management, AI may be used to support portfolio construction, risk assessment, and trading strategies.

Machine Learning (ML): ML is a subset of AI that enables systems to identify patterns based on data inputs without being explicitly programmed. ML models may be used in stock selection to identify investment opportunities based on historical and real-time data.

Natural Language Processing (NLP): NLP is a field of AI focused on the interpretation and generation of human language by machines. In financial contexts, NLP may be applied to analyze textual data such as earnings reports to inform investment decisions.

Large Language Models (LLMs): LLMs are advanced NLP systems trained on extensive datasets to understand and generate human-like text. In investment management, LLMs may assist in synthesizing qualitative information or generating insights, but do not independently make investment decisions.

091025-4799572-101026