Phoenix Peers: Using Double-Entry Bookkeeping Graphs for Comparable Firm Identification

Eyup O. Gun
Tepper School of Business
Carnegie Mellon University
orhun@cmu.edu

Pierre Jinghong Liang
Tepper School of Business
Carnegie Mellon University
liangj@andrew.cmu.edu

Burton Hollifield
Tepper School of Business
Carnegie Mellon University
burtonh@cmu.edu

Draft: December 1, 2025

Abstract

We introduce "Phoenix Peers," a peer selection methodology grounded in the topological structure of double-entry bookkeeping. Unlike traditional approaches that rely on categorical industry codes or observable outcomes, we measure firm similarity using the Jensen-Shannon distance between directed graphs representing the flow of economic value through financial statements. Using Compustat data from 1987–2024, we document that structural accounting similarity captures a dimension of operational comparability that is largely orthogonal to industry and size groupings, exhibiting less than 1% overlap with traditional benchmarks. Empirically, Phoenix Peers triple the explanatory power of comparable-firm valuation regressions, increasing the adjusted R^2 for enterprise-value-to-sales ratios from 13% to 39%. These findings demonstrate that the mathematical laws governing accounting systems encode fundamental information about business models that is invisible to standard classification schemes.

Keywords: peer selection; graph entropy; Jensen–Shannon divergence; kNN forecasts; comparable-firm valuation.
JEL codes: G12, M41, C55

1 Introduction

Consider the transformation of a phoenix: destruction followed by rebirth into something fundamentally different yet intrinsically connected to its origins. This ancient metaphor captures our approach to peer firm identification. We deconstruct financial statements to their atomic elements—the individual account relationships recorded through double-entry bookkeeping—then reconstruct them as network graphs that reveal operational structure invisible in traditional tabular displays. Like burning down the familiar "table of numbers" to rebuild it as an interconnected topology, we introduce Phoenix Peers: a methodology that identifies comparable firms through the structural similarity of their accounting-graph architectures.

To understand why this transformation matters, consider two firms operating in the same retail sector with comparable market capitalizations and overlapping customer segments. Traditional peer selection methods—whether based on SIC industry codes, size deciles, or product description based—would likely classify these firms as comparable. Yet beneath the surface, their accounting structures reveal fundamentally different economic architectures. The first firm operates a capital-intensive model with substantial inventory holdings, complex supplier financing arrangements, and debt-heavy capital structure reflected in dense interconnections between asset, liability, and equity accounts. The second operates an asset-light platform business model connecting third-party sellers, characterized by minimal inventory, transaction-based revenue recognition, and equity-heavy financing reflected in a sparser, more centralized account structure. Despite appearing similar through conventional metrics, these firms may face entirely different economic risks, capital allocation constraints, and operational leverage—differences that manifest systematically in the structural relationships embedded within their financial statements.

This example illustrates a fundamental challenge in financial statement analysis: existing peer selection methods often fail to capture the deeper operational similarities that drive economic outcomes. While industry codes provide broad sectoral groupings and factor models identify systematic risk exposures, neither approach directly examines the structural relationships within firms' accounting systems that reflect their underlying business models and operational characteristics.

While professional analysts possess the discretion to qualitatively adjust for these structural factors, relying on their subjective assignments introduces host of challenges including unreplicability, agency conflicts, and cognitive biases (e.g., Michaely and Womack (1999); Hirshleifer and Teoh (2003); De Franco et al. (2013)).

Selecting appropriate peer firms is central to valuation, performance benchmarking, and risk modeling, yet existing approaches remain fragmented across practice and academia. In practice, peer definition is often application-specific: valuation groups, compensation consultants, and risk managers all rely on different peer-construction logics tailored to their task. Industry practice gravitates toward factor-based systems such as MSCI's BARRA fundamental factor model, whereas academic work typically defaults to SIC/GICS industry groupings, sometimes augmented by size, momentum, or profitability screens (Alford, 1992; Hou et al., 2020; Bender and Nielsen, 2010).

BARRA models are the workhorse of portfolio risk management because they capture exposures to characteristics—industry membership, leverage, growth, momentum—that explain systematic return variation. Their purpose, however, is risk control: to decompose returns into systematic and idiosyncratic components, monitor unintended factor bets, and attribute performance. Importantly, this objective differs from identifying economically comparable firms for tasks such as valuation multiples, forecasting, or peer-based benchmarking. A firm that is close in factor space may share similar return sensitivities but still operate under a fundamentally different business model or accounting structure—exactly the type of mismatch documented in financial-statement benchmarking research (e.g., Hoitash et al. (2023)).

The consequences of peer choice extend well beyond methodological details. Early empirical evidence from Alford (1992) demonstrates that the accuracy of price-earnings valuation varies systematically with comparable firm selection: industry membership yields substantially greater accuracy than size or growth screens alone, with further improvements from finer industry partitions. More recent evidence underscores the pervasive reliance on peer-based methods: Bastianello et al. (2025) document that roughly 65% of analyst reports employ at least one multiples-based approach, with peer selection directly anchoring valuation multiples and investment recommendations. In academic settings, peer-based benchmarking underlies performance evaluation, abnormal return calculations, and cross-sectional tests in many areas of accounting and finance research.

1.1 Limitations of Existing Approaches

Despite this widespread reliance, existing peer selection methods share fundamental limitations that constrain their effectiveness for financial statement analysis.

Categorical approaches such as SIC/GICS industry classifications provide broad groupings but often mask substantial within-industry heterogeneity in business models and operational characteristics. Fairfield et al. (2009) demonstrate this limitation empirically: while industry-specific models improve out-of-sample forecasts of sales growth relative to economy-wide benchmarks, they provide no gains for profitability prediction, suggesting that broad industry codes are too coarse to capture the firm-level drivers of earnings performance.

Factor-based approaches such as BARRA fundamental factor models excel at systematic risk decomposition but are designed primarily for portfolio risk management and performance attribution rather than identifying operational comparability. These models capture exposures to macroeconomic factors through firm-level characteristics, yet their construction optimizes for explaining return variation around expected levels rather than for defining economic similarity relevant to valuation or forecasting applications.

Outcome-based approaches define similarity through observable results—product descriptions (Hoberg and Phillips, 2010), analyst co-coverage (Kaustia and Rantala, 2021), earnings correlation patterns (Franco et al., 2011), or financial statement line-item overlap (Hoitash et al., 2023). While these methods capture important aspects of comparability, they are inherently backward-looking and may conflate correlation with causation. More fundamentally, they define similarity through outcomes or disclosed characteristics rather than the underlying structural factors that generate those outcomes.

A recent frontier leverages large language models to embed firm disclosures into continuous similarity spaces. Vamvourellis et al. (2023) use 10-K business descriptions to generate company embeddings that reproduce GICS classifications with high accuracy and yield peers with stronger return comovement than traditional industry buckets. While representing a significant advance over categorical approaches, these text-based methods remain dependent on disclosed language and product descriptions rather than the mathematical structure underlying all financial reporting.

1.2 The Structural Alternative: Accounting-Graph Distances

We propose that structural similarity in firms' double-entry bookkeeping systems provides a more fundamental basis for peer identification than existing categorical, factor-based, or outcome-based approaches. This approach builds on the insight that every firm's economic activities are recorded through the same universal system of double-entry bookkeeping, governed by mathematical laws of balance, conservation, and linearity. The resulting financial statements encode not merely the outcomes of business activities but the structural relationships between stocks and flows that reflect underlying operational architectures.

Building on the bookkeeping-graph[1] framework of Liang (2025), we represent each firm's financial statements as a directed, weighted graph $G = (V, E, W)$ where vertices V correspond to T-accounts, edges E represent journal entry flows, and weights W capture the magnitude of debits and credits between accounts. This graph-theoretic representation transforms the traditional view of financial statements as static tabular displays into dynamic network structures that capture the flow of economic value through firms' operations.

For peer selection, we measure structural similarity through information-theoretic distances between firms' accounting-graph entropy distributions. Specifically, we compute the Jensen-Shannon distance.

1.3 Research Questions and Empirical Strategy

Our study addresses two fundamental research questions. First, do accounting-graph distances capture a distinct dimension of firm similarity relative to existing peer selection methods? Second,

[1] Throughout, we use "bookkeeping-graph distance" and "accounting-graph distance" interchangeably.

does structural similarity in bookkeeping graphs translate to superior performance in the economic applications where peer quality matters such as valuation accuracy, forecasting precision, and return predictability?

We evaluate Phoenix Peers through an empirical design adapted from Hoitash et al. (2023) that emphasizes economic validity in valuation applications. Our primary test examines whether accounting-graph distance peers outperform traditional benchmarks in predicting enterprise-value-to-sales ratios using Fama–MacBeth regressions with strict out-of-sample protocols. We benchmark Phoenix Peers against industry peers (SIC2) and industry-size peers (four closest-sized firms within SIC2), following the validation framework of Bhojraj and Lee (2002) and Hoitash et al. (2023). To assess whether Phoenix Peers capture orthogonal information, we compute Jaccard similarity coefficients measuring peer-set overlap: Phoenix Peers exhibit less than 1% overlap with industry-size peers, indicating that structural similarity captures (operational) comparability invisible to traditional categorical groupings.

1.4 Contribution to Literature

Our study makes several contributions to the accounting and finance literature on peer selection and financial statement comparability. **Methodologically**, we introduce the first systematic application of information-theoretic graph distances to peer identification, providing a structural foundation for comparability grounded in the universal mathematical laws of double-entry bookkeeping rather than industry conventions or observable outcomes.

Empirically, we establish whether structural similarity contains incremental information beyond existing approaches through comprehensive benchmark testing.[2]

Theoretically, we advance understanding of what constitutes meaningful economic comparability by demonstrating that the structural relationships within firms' accounting systems capture operational similarities that transcend traditional industry boundaries and factor exposures. This extends the growing literature on alternative data applications in accounting and finance by showing how fundamental accounting structures can be leveraged for peer identification.

The study design evaluates whether accounting-graph distances contain incremental information beyond prevailing peer-selection methods, establishing their potential role in valuation, forecasting, and asset-pricing applications. By grounding peer selection in the mathematical structure of double-entry bookkeeping rather than categorical assignments or observable outcomes, we provide a transparent, replicable method that leverages the universal laws governing all financial reporting.

The remainder of this paper proceeds as follows. Section 2 reviews related literature and positions our contribution within existing peer selection approaches. Section 3 describes our data sources and methodology for constructing accounting-graph distances. Section 4 outlines our empirical design for validating peer selection methods. Section 5 presents comparative performance results across peer identification approaches. Section 6 explores robustness through alternative specifications and sensitivity analyses. Section 7 concludes with implications for research and practice.

2 Related Literature and Positioning

Our study builds on three distinct streams of research: traditional peer selection methods, modern approaches using textual and network data, and the emerging literature on accounting-graph representations.

[2] Unlike prior studies that typically compare new methods against one or two existing approaches, we will be evaluating AG-D peers against the full spectrum of established alternatives using identical out-of-sample protocols to ensure fair comparison.

2.1 Practitioner Approaches

In practice, peer identification follows several distinct paradigms that differ in their inputs, objectives, and operational definitions of comparability (Table 1). Barra-style factor models remain the industry standard for portfolio risk management and performance attribution. Originating with Rosenberg and Marathe (1976), these models capture the impact of macroeconomic events through firm-level microeconomic characteristics. In practice, Barra's *fundamental factor models* link each stock to exposures such as industry membership, size, value, growth, leverage, yield, momentum, currency sensitivity, and volatility, estimated from standardized descriptors of financial statements, analyst forecasts, and market data (Bender and Nielsen, 2010). At the firm level, factor exposures resemble the building blocks used by fundamental analysts, but their purpose differs: the analyst seeks to forecast returns, while the factor model forecasts variation around expected returns. At the portfolio level, factor models provide risk decomposition—separating systematic from firm-specific drivers—and help monitor unintended bets, attribute performance, and manage exposures over time. In empirical terms, Barra-style models show that common factors (style and industry) explain the majority of portfolio risk, far outweighing idiosyncratic components. Their primary function, however, is risk control and attribution, not the identification of economically comparable peers for valuation or forecasting. Our study positions Phoenix Peers as complementary to factor-based approaches: while BARRA captures systematic risk exposures, Phoenix Peers capture operational structure through the double-entry system. Future work will directly compare Phoenix Peers to BARRA factor neighbors (see Online Appendix). The second are comparables chosen by third parties, often data providers such as Bloomberg, investment banks, or analyst reports. Analyst-selected peers are especially important: they are widely used in industry reports, valuation exercises, and target-price setting, and thus influence both market practice and academic research (Bastianello et al., 2025). These practitioner methods highlight the prevalence of both statistical factor-based peers and judgment-based comparables, but neither is designed with the explicit goal of identifying structurally comparable firms.

A recent frontier in practice leverages large language models (LLMs) to embed firm disclosures into continuous similarity spaces. In a BlackRock Investment Institute working paper, Vamvourellis et al. (2023) use Item 1 of 10-K business descriptions for the Russell 3000 (2,590 filings in 2022) to generate company embeddings via state-of-the-art language models (BERT, SBERT, Longformer, GPT, PaLM). These embeddings reproduce GICS classifications with high accuracy ($\approx 90\%$ at the sector level; F1 ≈ 0.79 at the industry level) and yield peers that exhibit higher return comovement than GICS (average pairwise correlation 47% vs. 41%). Clusters formed from embeddings also explain more cross-sectional monthly returns (11.9% R^2) than sector (5.2%) or industry (10.6%) groupings. This practitioner evidence highlights a shift from discrete taxonomies toward embedding-based, continuous peer definitions. Our study complements this trend by deriving embeddings not from textual disclosures but from the structural laws of double-entry bookkeeping, thereby capturing operational similarity beyond language or product descriptions.

These practitioner methods highlight the prevalence of both statistical factor-based peers and judgment-based comparables, but neither is designed with the explicit goal of identifying structurally comparable firms. Table 1 contrasts these approaches with our bookkeeping-graph entropy methodology, which derives continuous similarity from the structural laws of double-entry accounting.

2.2 Academic Approaches

2.2.1 Traditional Academic Approaches

Academic work has long mirrored these practitioner methods. Early evidence from Alford (1992) shows that the accuracy of the price–earnings valuation method depends critically on the set of comparable firms: industry membership (at the two- or three-digit SIC level) yields the most accurate valuations, with further improvements when combined with return-on-equity or size screens, whereas growth or leverage alone perform poorly. These results established industry-based peers as the default academic benchmark.

Industry-level analysis and forecasting limits. A complementary line of evidence evaluates whether industry-specific models improve out-of-sample forecasts relative to economy-wide models. Fairfield et al. (2009) estimate rolling mean-reversion models for firm performance (sales growth, growth in book value and net operating assets, ROE, and RNOA) and compare industry-specific (IS) to economywide (EW) predictions. IS models materially improve forecasts of sales growth at both one- and five-year horizons and offer some gains for long-horizon investment growth, but they do not improve forecasts of profitability (ROE, RNOA) at either horizon. Analyst forecasts and market-based tests align with these patterns. These results suggest that broad industry peers are too coarse for earnings/profitability forecasting—precisely where our structural peers, defined via accounting-graph distances (AG-D/JSD), are designed to add information beyond industry membership.

Life-cycle partitions. A complementary approach replaces industry with economically motivated states. Vorst and Yohn (2018) classify firms by life cycle (cash flow signs; Dickinson, 2011) and show that life-cycle models deliver substantially more accurate out-of-sample forecasts of both growth and profitability than economy-wide or industry-specific models, with improvements persisting at one- to three-year horizons and holding under alternative (Anthony–Ramesh; Hribar–Yehuda) classifications. Gains are broad across stages and larger when uncertainty is high (e.g., idiosyncratic risk, volatile operating performance, R&D and intangible intensity, special items). Analysts partially impound life-cycle information, as forecast errors co-vary with the model's incremental improvements. These results reinforce that coarse industry peers are insufficient for profitability/earnings forecasting and motivate richer, economically grounded groupings. Our structural peers push this logic further by replacing state labels with distances computed from the double-entry system's internal structure, allowing peer sets to flexibly capture within-stage heterogeneity by capturing the actual topology of firms' transaction flows.

Valuation–Based Peers. Guided by valuation theory, Bhojraj and Lee (2002) develop a firm specific "warranted multiple" and define peers as firms with the closest warranted enterprise value to sales (WEVS) or warranted price to book (WPB) ratio. This approach addresses the long recognized subjectivity of practitioner comp-set selection by systematizing peer choice with cross-sectional drivers of multiples (profitability, growth, and cost of capital). In extensive out-of-sample tests, warranted-multiple peers markedly outperform industry and industry–size matches in predicting one to three year ahead EV/S and P/B, nearly tripling adjusted R^2 for EV/S and more than doubling it for P/B—with especially pronounced gains for "new economy" firms (Bhojraj and Lee, 2002). We treat BL02's warranted-multiple peers as a valuation-centric, outcome-based benchmark in our tests; by contrast, our method is price agnostic and defines proximity from the internal structure of the double-entry system and asks whether such *structural* peers add information beyond outcome-based comparable.

More recent contributions expand the comparability construct further. Text-based product-market similarity (TNIC; Hoberg and Phillips, 2010) exploits firm disclosures; correlation-based networks (Kelly et al., 2019) use stock-price co-movement; and other work emphasizes product-market overlap (Bhojraj and Lee, 2002) or analyst-reported comparables (Franco et al., 2011). These approaches focus on observable outcomes such as industries, products, returns, or disclosure content—rather than the internal accounting system.

Economic content of text-based product-market peers. Hoberg and Phillips (2010) demonstrate that text-based product-market similarity has first-order economic consequences, using 10-K product descriptions to show that acquirer–target proximity predicts deal incidence and stronger ex post outcomes—including higher announcement returns, cash flows, sales growth, and new product introductions—especially when targets are similar to acquirers yet differentiated from rivals. Their approach reveals that merger pairs are far more similar in text space than SIC/NAICS codes suggest, highlighting the limits of coarse industry classifications. Hoberg and Phillips (2016) extend this framework by demonstrating that text-space proximity interacts with rival network density to shape firms' product differentiation, investment, and innovation outcomes. Together, these findings establish that continuous, disclosure-based distances for peer construction are economically meaningful and capture endogenous competitive dynamics. This evidence validates the use of continuous, outcome-driven peer measures and motivates our complementary approach that derives proximity from the internal accounting system rather than reported product language. We therefore treat TNIC-style product-market peers as a leading benchmark for evaluating whether our structural, bookkeeping-graph approach provides incremental information beyond disclosure-driven distances[3].

Graph representations from financial filings. Recent advances have also explored automated graph construction from financial filings. Zhang et al. (2023) develop a comprehensive NLP pipeline to extract competition relationships directly from 10-K business descriptions, focusing on the "Competition" section. Their system identifies firm mentions using finance-tuned named-entity recognition and maps them to Compustat GVKEYs, producing large-scale competition graphs. In their S&P 500 sample (2000–2015), the resulting network covers roughly 700 unique firms and 1,200 rival edges, recovering about 80–85% of the rivalries disclosed in hand-collected benchmarks. The graphs reveal intuitive hubs (e.g., Microsoft, Apple, ExxonMobil) and dense rivalry clusters by sector, with network statistics (degree, clustering, k-core) aligning closely with known industry structure. Relative to TNIC's continuous product-market proximities (Hoberg and Phillips, 2010, 2016), these competition graphs encode discrete, disclosed rivalries that highlight network position and strategic market structure. While such rivalry networks capture inter-firm competitive positioning, our bookkeeping-graph approach constructs networks from the mathematical structure of double-entry accounting itself, potentially capturing deeper operational similarities that may not be explicitly disclosed. Together, these complementary graph-based approaches highlight the potential for mining different aspects of financial reports to construct peer networks.

LLM-based entity extraction. Recent work applies large language models directly to the problem of peer identification. Covas (2023) use GPT-3.5 to perform named entity recognition (NER) on company descriptions, extracting products and services to form peer groups based on entity overlap.

[3]Further discussion of our planned integration of disclosure-based text similarity into the broader peer-selection framework is provided in the Online Appendix

In tests on 3,890 firms with Wikipedia summaries, GPT substantially outperformed standard spaCy NER models (zero-shot F-score $\approx$ 56%, rising to 85–88% with few-shot prompts), and the resulting peers align intuitively by industry (e.g., energy firms for Total SE, satellite firms for Iridium). Their approach highlights the potential of LLMs to automate comparable-firm construction, especially in private equity settings where disclosure is limited. Relative to these disclosure-driven and entity-extraction peers, our study offers a complementary structural foundation: we construct comparables from the double-entry bookkeeping system itself, thereby capturing operational similarity even when product-market text is sparse or unavailable.

Outcome-Based Peers and Forecasting Applications A central outcome-based benchmark is Franco et al. (2011), who construct a measure of financial-statement comparability that captures how similarly two firms' accounting maps economic events into earnings. In an earnings-correlation variant (CompAcct-R^2), comparability is the adjusted R^2 from firm i's earnings on firm j's earnings over 16 quarters, averaged over the top four matches within 2-digit SIC; they also partial out economics using analogous cash flow and return correlations. Comparability is positively associated with analyst following and forecast accuracy and negatively with forecast dispersion, consistent with lower information-acquisition costs and richer information sets for analysts. For construct validity, they hand-collect 1,000 Investext reports (2005) and show that higher comparability raises the probability that an analyst lists j as a peer for i by roughly 3–5 percentage points after industry and firm controls. These findings establish that earnings-based comparability captures economically meaningful peer relationships, motivating our investigation of whether structural similarity in accounting graphs provides incremental information beyond outcome-based measures. Our AG-D methodology evaluates this question directly through valuation tests that compare the predictive power of graph-based peers against traditional industry-size benchmarks.

Peers also matter directly for forward-looking information production. Bastianello et al. (2025) show that analysts' "mental models" hinge on their implicit peer choices, which systematically shape both the accuracy and optimism of forecasts. When analysts' mental-model peers align with economic fundamentals, forecasts improve; when misaligned, errors increase. This highlights that peer definitions condition not only valuation multiples but also the information sets analysts use to generate forecasts. Our empirical framework addresses this directly by testing whether AG-D peers improve valuation accuracy in enterprise-value-to-sales predictions, where our baseline results show accounting-graph peers ($K = 4$) achieve approximately 39% adjusted R^2 compared to 13% for industry-size benchmarks—a near tripling of predictive power that suggests structural similarity captures fundamental peer relationships beyond traditional categorical groupings.

Financial analysts play a central role as information intermediaries in capital markets. Bradshaw et al. (2016) review how analysts contribute to price formation through information discovery, interpretation, and dissemination, and highlight that comparables are embedded in their research process. Analysts' forecast behavior is shaped by peer choice, as well as institutional incentives such as career concerns, investment-banking ties, and herding. Coverage itself matters: initiation and withdrawal of coverage alter liquidity, institutional investment, and the degree of firm-specific information in prices. These findings underscore that peer definitions are not merely methodological constructs for researchers but have direct implications for capital-market functioning[4].

Beyond valuation and compensation applications, comparables can also be used directly for forecasting. Easton et al. (2023) develop a nonparametric k-nearest neighbors (kNN) approach that predicts earnings by matching each firm-year to those with similar recent earnings histories. Specif-

[4] Further discussion of our planned integration of analyst chosen peers to our research is provided in the Online Appendix.

ically, they forecast one-year-ahead earnings before special items (EBSI) scaled by market equity, and use the median of the k nearest neighbors' future outcomes as the prediction. This method avoids the linearity assumptions of traditional autoregressive or panel regressions and naturally produces an ex ante indicator of forecast accuracy. In particular, they define

$$\mathrm{MAD}_{i,t} = \frac{1}{k}\sum_{j=1}^{k}\left|\frac{\mathrm{EBSI}_{j,t+1}}{\mathrm{ME}_{j,t}} - \frac{\widehat{\mathrm{EBSI}}_{i,t+1}}{\mathrm{ME}_{i,t}}\right|, \tag{2.2.1}$$

which provides a firm-year specific ex ante measure of forecast precision based on the dispersion among neighbors' outcomes. They show that kNN forecasts outperform analyst consensus at longer horizons, are incrementally value relevant, and support profitable trading strategies. Importantly, the peer sets implied by kNN diverge sharply from industry or SIC groupings, underscoring that the choice of comparables is a first-order research question[5].

Our study follows in their footsteps: we use the kNN framework both as a methodological template and as a benchmark for our own peer selection. Whereas Easton et al. (2023) define similarity using past earnings histories, we replace this outcome-based distance with accounting-graph entropy measures of structural similarity. In doing so, we generalize the nonparametric peer-based forecast paradigm to a foundation grounded in the laws of double-entry bookkeeping. Our valuation accuracy tests directly assess whether this structural approach provides superior peer identification. Using Fama-MacBeth cross-sectional regressions over the period 1987–2024, we find that AG-D peers substantially outperform both industry-only ($R^2 = 0.124$) and industry-size ($R^2 = 0.131$) benchmarks, achieving $R^2 = 0.388$ when $K = 4$ peers are used. This represents approximately a 28 percentage point improvement in one-year-ahead valuation accuracy, demonstrating that structural similarity captures economically meaningful peer relationships beyond what is captured by industry classification or outcome-based earnings histories.

Finally, Hoitash et al. (2023) introduce the FSB measure, based on board- and analyst-disclosed peers in large samples. Their work demonstrates that overlap in disclosed peer sets is predictive of actual peer choice, making FSB the current leading outcome-based benchmark against which new measures should be evaluated.

Financial-Statement Benchmarking (FSB) peers. Hoitash et al. (2023) define financial-statement benchmarkability for a firm pair using the Jaccard coefficient over reported line items—i.e., the size of the intersection divided by the size of the union of non-missing items—constructed first from Compustat's standardized statements and, in robustness, from "as reported" XBRL tags. Compustat provides broad coverage (776 items, of which a subset are non-missing in practice), whereas XBRL offers much finer granularity (over 7,000 monetary tags post-2011); the authors validate that presence/overlap (rather than amounts) avoids imputation bias and undue influence from a few large values. In revealed-preference tests, a one–standard-deviation increase in pairwise FSB raises the probability that analysts or boards select a peer by about 10%, and *low* FSB is associated with strategic peer choices—higher analyst forecast optimism and greater CEO overpay (roughly 6%)—consistent with mis-benchmarking costs. They further show applications to multiples valuation and other benchmarking tasks. We therefore treat FSB as the leading outcome-based comparator and ask whether our accounting-graph distance (AG-D/JSD) peers add incremental information for forecasting and returns beyond FSB.

[5]Please see the Online Appendix, where we outline an expansion that adapts Easton et al. (2023)'s temporal-matching framework to our accounting-graph distance methodology by computing peer similarity over trailing multi-year windows rather than relying solely on contemporaneous annual data.

2.2.2 Accounting Graph Approach

Foundations and computational underpinnings. Liang (2023) formalizes the representation of double-entry bookkeeping as a directed, weighted graph and lays out the computational toolkit for analyzing these graphs in practice. The monograph motivates "journal entries as graphs" and develops an information-theoretic/MDL perspective that supports graph-mining tasks such as anomaly detection and graph summarization, while also detailing how to convert real-world general-ledger data—especially compound entries—into analyzable graph objects. This framework provides the conceptual and data-engineering basis for structural similarity measures on accounting graphs that we use for peer selection.

A recent advance departs from outcome-based measures by grounding similarity in the laws of double-entry bookkeeping. Liang (2025) formalizes three universal "bookkeeping laws"—balance, conservation, and linearity—and recasts double-entry bookkeeping as a directed, weighted graph. Within this framework, Liang defines three entropy-based summaries (and associated divergences): (i) a *Balance-sheet Node-entropy* that treats the two-way classified balance sheet as a joint probability distribution; (ii) a *Transaction Edge-entropy* based on normalized debit–credit flows; and (iii) a *Bookkeeping Graph-entropy* computed from the Laplacian eigenvalue spectrum. These measures are internal to the accounting system and quantify structure in stocks, flows, and topology.

Our paper is the first to systematically apply this accounting-graph entropy framework to peer selection. Specifically, we measure peer proximity as the divergence between two firms' double-entry structures. For symmetry and boundedness, we replace Liang's canonical Kullback–Leibler divergence with the Jensen–Shannon distance, while retaining the same bookkeeping-graph foundation. This structural approach allows us to test whether peer similarity defined by accounting laws contains incremental information beyond traditional outcome-based definitions such as FSB, TNIC, or industry-size buckets.

Building on the bookkeeping-graph framework of Liang (2025), Pyo et al. (2025) construct quarterly GAAP graphs for U.S. firms (1995–2024) and quantify *structural change* as the Kullback–Leibler divergence (KLD) between consecutive periods' normalized edge-weight distributions. They document that large structural changes entail short-run profitability declines but stronger long-run sales growth, yet elicit negative market reactions—especially when concurrent earnings news is positive—consistent with elevated uncertainty.[6] While their focus is intertemporal change, their pairwise distance illustrations demonstrate that bookkeeping-graph distances produce sensible cross-sectional clusters, validating the construct for mapping firm-to-firm proximity. Our paper is the first to *systematically* deploy these distances for peer selection—replacing asymmetric KLD with a symmetric, bounded Jensen–Shannon distance (JSD)—and to evaluate incremental predictive content relative to outcome-based peers such as FSB, TNIC, industry–size buckets, and factor-space neighbors.

In sum, the literature has evolved from categorical industry codes toward continuous similarity measures, and from outcome-based metrics toward structural approaches grounded in accounting fundamentals. Figure 1 maps this evolution and highlights how AG-Distance represents a natural progression from FSB's binary account-presence approach to full transaction-flow analysis.

3 Data Construction and Methodology

Our approach departs from traditional outcome-based peer selection by grounding comparability in the structural properties of the double-entry bookkeeping system. In this section, we outline

[6] They also show that option-implied volatility rises after high-KLD filings.

the construction of accounting graphs, the metric used to quantify structural similarity, and the formation of the Phoenix Peer variables used in our empirical tests.

3.1 Accounting Graph Representation

Building on the framework of Liang (2025), we represent each firm's financial statements not as a static list of account balances, but as a directed, weighted graph $G = (V, E, W)$. In this representation:

- **Vertices (V):** Correspond to specific financial statement accounts (e.g., Cash, Inventory, Retained Earnings).
- **Edges (E):** Represent the flow of economic value between accounts, capturing the articulation of the balance sheet and income statement via journal entries.
- **Weights (W):** Capture the magnitude of debits and credits flowing between accounts during the reporting period.

Figure 2 provides a visual representation of this topology for NVIDIA Corporation (NVDA) in Q4 2024. As illustrated, the graph structure explicitly maps the articulation between flow accounts (e.g., Revenue, Expense, CFO) and stock accounts (e.g., Net Noncash Assets), characterizing the firm's operational architecture for that period.

To ensure comparability across firms of vastly different sizes, we normalize these flows into probability distributions, P_i, representing the relative allocation of economic activity within firm i's accounting system. This transformation allows us to treat financial statement analysis as a problem of comparing probability distributions over a fixed accounting schema.

3.2 Measuring Structural Similarity (Jensen-Shannon Distance)

To quantify the pairwise similarity between firms, we employ information-theoretic divergence measures. While Kullback-Leibler (KL) divergence is a standard measure of distributional difference, it is asymmetric and undefined for zero-probability events—a frequent occurrence in sparse accounting data.

Therefore, we measure structural similarity using the **Jensen-Shannon Distance (JSD)**, which is the square root of Jensen-Shannon Divergence. JSD is symmetric, bounded to the interval $[0, 1]$, and robust to sparse support and metric. For two firms i and j with accounting graph distributions P_i and P_j, the distance is defined as:

$$JSD(P_i, P_j) = \sqrt{\frac{1}{2} D_{KL}(P_i || M) + \frac{1}{2} D_{KL}(P_j || M)} \tag{3.2.1}$$

where $M = \frac{1}{2}(P_i + P_j)$ is the mixture distribution, and $D_{KL}(\cdot || \cdot)$ denotes the Kullback-Leibler divergence. A JSD of 0 indicates identical accounting flow structures, while a JSD of 1 indicates completely disjoint structures.

Figure 3 illustrates the distributional impact of using the distance metric versus the raw divergence. As shown in the right panel, the Divergence distribution exhibits a pronounced right tail, representing pairs with high structural dissimilarity—such as comparisons between capital-intensive manufacturing firms and asset-heavy financial institutions. The dense mid-section reflects the universal constraints of GAAP, which impose a baseline structural similarity across the economy.

For our primary application of selecting nearest neighbors, either metric suffices as the ordinal ranking remains invariant. However, as shown in the left panel, the square-root transformation inherent in the Jensen-Shannon Distance compresses this right tail. This resulting scale is more distinct and better behaved, which is critical for subsequent analyses where structural distance is employed as a continuous explanatory variable.

3.3 Constructing Phoenix Peers

We construct our peer sets at the end of each fiscal year t. For every focal firm i, we compute the pairwise JSD against all other firms in the sample universe. We then rank potential peers by ascending distance (descending similarity).

To generate the regressor used in our valuation models, we select the top K structurally similar peers (Phoenix Peers). The peer-implied valuation multiple is calculated as the equal-weighted mean of the peers' observed multiples:

$$EVS_{Phoenix,i,t}^{(K)} = \frac{1}{K} \sum_{j \in \mathcal{N}_i^K} EVS_{j,t} \tag{3.3.1}$$

where $\mathcal{N}_i^K$ is the set of the K nearest neighbors to firm i in the accounting-graph space. In our baseline analysis, we set $K = 4$ to maintain consistency with prior literature (Bhojraj and Lee, 2002; Hoitash et al., 2023), though we test sensitivity to $K \in \{3, 5, 10, \ldots, 200\}$.

Figure 4 illustrates the resulting peer network for a cross-section of major firms. The spatial arrangement, derived directly from pairwise Jensen-Shannon distances, effectively recovers economic clusters without relying on categorical industry labels. For instance, retailers (WMT, TGT) and banks (BAC, JPM) form distinct, tight neighborhoods, highlighting that structural accounting similarity captures the underlying operational homogeneity of these business models.

3.4 Sample Selection and Filters

Our primary dataset is constructed from quarterly Compustat data covering the period 1987 through 2024. We apply standard filters following Bhojraj and Lee (2002): restricting to the industrial format (INDL), standardized data format (STD), consolidated accounts (C), and U.S. dollar reporting (USD).

We exclude financial firms (SIC 6000–6999) and utilities (SIC 4900–4999) to ensure benchmarkability with existing research, though our results are robust to their inclusion. We further exclude firms with non-positive sales or shares outstanding, non-U.S. incorporated firms (ADRs), and firms with total assets below \$100 million.[7]

After applying these filters and requiring valid future enterprise-value-to-sales ($EVS_{t+\tau}$) data for our dependent variables, our final sample for the main regression analysis consists of approximately 94,000 firm-year observations.

3.5 Benchmark Datasets

To evaluate the incremental value of Phoenix Peers, we compare them against both traditional and novel peer identification schemes:

[7] We use a \$100 million total assets threshold in this analysis due to data availability constraints in our accounting graph construction. Firms below this threshold have insufficient coverage in our bookkeeping relationship data. In subsequent analyses with updated graph data, we will lower this threshold to the standard \$1 million cutoff commonly used in the literature.

Industry and Size Peers: We construct two standard baselines. **Industry Peers** are defined as the leave-one-out mean of all firms in the same SIC 2-digit industry. **Industry-Size Peers** are defined by selecting the four firms closest in log market capitalization within the same SIC 2-digit industry-year, following Bhojraj and Lee (2002).

Financial Statement Benchmarking (FSB): We use the FSB data (1990–2023) introduced by Hoitash et al. (2023). FSB defines similarity via the Jaccard coefficient of overlapping Compustat line items:

$$FSB_{i,j,t} = \frac{|Items_{i,t} \cap Items_{j,t}|}{|Items_{i,t} \cup Items_{j,t}|}$$

This measure captures similarity in accounting choices (reporting behavior) rather than structural flows. We utilize the firm-level dataset aggregating the top-4 FSB peers.

Yahoo Finance Comparables (Web-Collected): We collected "Also Watched" peers from Yahoo finance on August 5, 2025, via a polite Python script, reporting progress in blocks of 50 tickers and saving checkpoints every 100. Requests sleep uniformly between 0.3–0.7 seconds after each successful response. We obtained 11,986 focal firm‘s and their peer lists. Nearly all retrieved focal firms have exactly five peers (11,907); fewer than 80 have fewer than five. Yahoo provides a continuous similarity score for each focal–peer pair; we will treat this as a regression covariate and, when needed, convert it to a ranked top-k list.

4 Empirical Design

To evaluate the economic content of accounting-graph similarity, we employ a relative valuation framework. Following the methodology of Bhojraj and Lee (2002) and Hoitash et al. (2023), we test the ability of different peer selection methods to predict future valuation multiples.

4.1 Valuation Model

We estimate cross-sectional Fama-MacBeth regressions to predict future Enterprise-Value-to-Sales (EVS) ratios. The baseline specification is:

$$EVS_{i,t+\tau} = \alpha + \beta_1 EVS_{Ind,i,t} + \beta_2 EVS_{IndSize,i,t} + \beta_3 EVS^{(K)}_{Phoenix,i,t} + \epsilon_{i,t} \quad (4.1.1)$$

where:

- $EVS_{i,t+\tau}$ is the realized multiple for firm i at horizon $\tau \in \{1, 2, 3\}$ years.
- EVS_{Ind} is the mean multiple of the SIC 2-digit industry peers (excluding firm i).
- $EVS_{IndSize}$ is the mean multiple of the four closest size-matched peers within the industry.
- $EVS^{(K)}_{Phoenix}$ is the mean multiple of the top K structural peers identified via Accounting-Graph Distance.

All valuation multiples are winsorized at the 1% and 99% levels annually to mitigate the influence of outliers. Standard errors are adjusted using the Newey-West procedure with a lag of 3 to account for temporal autocorrelation in the time-series averages of the coefficients.

5 Main Results

5.1 Baseline Valuation Accuracy

Table 2 presents the results of the baseline regressions for forecast horizons of one, two, and three years. Consistent with prior literature, traditional Industry and Industry-Size peers load significantly, explaining approximately 13% of the cross-sectional variation in future multiples (Column 2).

The inclusion of Phoenix Peers (Column 3) yields a dramatic improvement in predictive power. The Adjusted R^2 nearly triples, rising from 0.131 to 0.388 at the $t+1$ horizon. The coefficient on $EVS_{Phoenix}$ is economically large (0.714) and statistically highly significant (t-stat = 14.54), absorbing much of the explanatory power previously attributed to industry classification.

This predictive advantage persists at longer horizons. At $t+3$, the model retains an Adjusted R^2 of 0.253, suggesting that structural accounting similarity captures persistent economic fundamentals rather than transitory pricing anomalies.

5.2 Orthogonality and Peer Overlap

A potential concern is that Phoenix Peers essentially replicate industry classifications using a more complex methodology. To address this, we compute the Jaccard similarity coefficient between the peer sets generated by the different methods.

As shown in Table 3, the overlap is strikingly low. The mean Jaccard similarity between Phoenix Peers and Industry-Size peers is only 0.0004 (and 0.006 for industry). On average, out of the top 4 Phoenix peers identified for a focal firm, virtually none are the same firms selected by the Industry-Size heuristic. This confirms that accounting-graph distance identifies a dimension of economic comparability that is fundamentally orthogonal to traditional industry and size groupings.

5.3 Sensitivity to Peer Set Size (K)

We examine the optimal number of peers (K) required to capture structural similarity. Table 4 shows a concave relationship: predictive power increases rapidly as K moves from 1 to 4, effectively stabilizing around $K = 4$ to $K = 10$ (Adj. $R^2 \approx 0.39 - 0.42$). Extending the set beyond 10 peers yields diminishing marginal returns. This supports our choice of $K = 4$ as a parsimonious baseline that balances idiosyncratic noise reduction with high structural relevance. Figure 5 plots the Adjusted R^2 of the valuation model as K varies from 1 to 200.

6 Mechanisms and Robustness

6.1 Constrained vs. Unrestricted Search

Traditional peer selection is rigid, typically limiting matches to firms within the same SIC codes. A distinct advantage of Phoenix Peers is the capacity to identify operationally similar firms across nominal industry boundaries. In Table 5, we formally test this flexibility by comparing our baseline unrestricted search (Column 1) against algorithms constrained to select peers only within the same SIC-2 or SIC-3 digit codes (Columns 2–5).

The results highlight a severe trade-off between strict industry constraints and sample coverage. The Unrestricted Phoenix method (Baseline) successfully identifies structural peers for the full sample of 78,063 firm-year observations. In contrast, restricting the search to the same SIC 2-digit code reduces the sample to 53,397 observations (a 32% loss), as many firms do not have four valid peers within their broad industry bucket that also meet the strict data requirements.

The attrition becomes catastrophic for finer partitions: requiring peers to share the same SIC 3-digit code and Size Quartile reduces the analyzable sample to just 8,758 observations—an 89% loss of coverage relative to the baseline. While the highly constrained models achieve slightly higher Adjusted R^2 values (ranging from 0.399 to 0.411 compared to 0.387), this marginal performance boost comes at the cost of excluding the vast majority of the market. The Unrestricted model remains robust and highly predictive (t-stat = 14.54) while providing valuation benchmarks for the tens of thousands of firms that are "orphaned" by strict industry-size binning. This demonstrates that accounting-graph distance effectively finds relevant peers even when traditional industry buckets fail to do so.

6.2 Sub-period Stability (1996–2016)

To ensure our results are not driven by sample period selection and to facilitate direct comparison with the leading study in the field, we re-estimate our baseline models on the restricted 1996–2016 period employed by Hoitash et al. (2023). By aligning our temporal window with this benchmark study, we can isolate the performance gains attributable to our structural methodology rather than differences in the underlying market environment.

The results, reported in Table 6, confirm that the predictive advantage of accounting-graph similarity is not only persistent but significantly outperforms established benchmarks. In absolute terms, the inclusion of Phoenix Peers increases the Adjusted R^2 by 28%, 20%, and 17% across the one-, two-, and three-year horizons, respectively. These gains are remarkably large when compared to the incremental improvements of approximately 7%, 6%, and 5.7% (6%) typically documented for outcome-based peers (such as FSB) over industry-size baselines.

7 Conclusion

Selecting appropriate peer firms is fundamental to valuation, performance benchmarking, and risk assessment, yet existing methods remain constrained by categorical industry classifications, factor exposures optimized for risk decomposition, or outcome-based similarities that conflate correlation with causation. This paper introduces a structural alternative grounded in double-entry bookkeeping.

We deconstruct financial statements to their atomic elements—individual account relationships—then reconstruct them as network graphs. Like a phoenix rising from ashes, we burn down the traditional "table of numbers" view and rebuild it as an interconnected structure. By measuring similarity through Jensen–Shannon distances between these accounting-graph distributions, we establish *Phoenix Peers*: a methodology capturing operational comparability through the topology of economic value flows rather than disclosed characteristics or observed outcomes.

The empirical results validate both novelty and economic content. Phoenix Peers achieve 39% adjusted R^2 in predicting enterprise-value-to-sales ratios, nearly tripling industry-size benchmarks (13%) and substantially exceeding the 5–7 percentage point gains documented for leading outcome-based measures. This advantage persists at multi-year horizons (30% and 25% at $t+2$ and $t+3$), indicating persistent economic fundamentals rather than transitory pricing anomalies.

Critically, Phoenix Peers exhibit less than 1% overlap with industry-size peers (Jaccard similarity of 0.0004), confirming orthogonal information content. Firms sharing similar accounting structures—and therefore operational constraints and capital allocation trade-offs—often operate in different nominal industries. By measuring similarity in structural flow space rather than categorical codes, Phoenix Peers reveal operational comparabilities invisible to traditional boundaries.

Practical advantages extend beyond accuracy. Unlike methods requiring shared SIC codes or size quartiles—excluding tens of thousands of observations—Phoenix Peers use unrestricted search across all firms. Our specification analyzes 78,063 observations; requiring SIC-3 and size quartile matches reduces coverage by 89% to 8,758 observations.

Three features distinguish our approach: **(1) Theoretical grounding**—derivation from mathematical laws of double-entry bookkeeping rather than empirical regularities; **(2) Transparency**—reliance on standardized financial data and well-defined information-theoretic metrics; **(3) Universality**—applicability across industries, geographies, and firm types through the common double-entry framework.

Promising extensions include: integrating with disclosure-based (TNIC, LLM) or outcome-based (FSB) distances to capture both operational structure and product positioning; subgraph decomposition to isolate which accounting components (operating flows vs. financing flows) drive peer quality; temporal smoothing over rolling windows to reduce transitory noise; and applications where traditional comparability fails (spin-offs, M&A targets, compensation benchmarking).

Our findings suggest that structural similarities encoded in double-entry topology predict valuation multiples more accurately than product descriptions or analyst judgments. If financial statement *structure*—independent of reported magnitudes—contains economically meaningful signals, then standard analysis focusing on ratio levels may overlook crucial information about business models and operational architectures.

Phoenix Peers demonstrate that centuries-old bookkeeping principles encode rich information about modern operational comparability. By measuring similarity in the mathematical space of accounting flows rather than categorical codes, we establish a peer framework grounded in the fundamental architecture through which firms record economic activity—examining not *what* firms report, but *how* the accounting system structures their reporting.

References

Alford, A. W. (1992). The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research*, 30(1):94–108.

Bastianello, F., Décaire, P. H., and Guenzel, M. (2025). Mental models and financial forecasts. Working paper.

Bender, J. and Nielsen, F. (2010). The fundamentals of fundamental factor models. Technical report, MSCI Research. SSRN Abstract ID 1707661.

Bhojraj, S. and Lee, C. M. C. (2002). Who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, 40(2):407–439.

Bradshaw, M., Ertimur, Y., and O'Brien, P. (2016). Financial analysts and their contribution to well-functioning capital markets. *Foundations and Trends in Accounting*, 11(3):119–191.

Covas, E. (2023). Named entity recognition using gpt for identifying comparable companies.

De Franco, G., Hope, O.-K., and Larocque, S. (2013). Analysts' choice of peer companies. *Review of Accounting Studies*, 20(1).

Dickinson, V. (2011). Cash flow patterns as a proxy for firm life cycle. *The Accounting Review*, 86(6):1969–1994.

Easton, P. D., Kapons, M. M., Monahan, S. J., Schütt, H. H., and Weisbrod, E. H. (2023). Forecasting earnings using k-nearest neighbors. Technical report, SSRN. SSRN Abstract ID 3752238.

Fairfield, P. M., Ramnath, S., and Yohn, T. L. (2009). Do industry-level analyses improve forecasts of financial performance? *Journal of Accounting Research*, 47(1):147–178.

Franco, G. D., Kothari, S., and Verdi, R. S. (2011). The benefits of financial statement comparability. Technical report, SSRN. SSRN Abstract ID 1266659.

Hirshleifer, D. and Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36(1–3):337–386.

Hoberg, G. and Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. Technical report, SSRN. SSRN Abstract ID 1181022.

Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465.

Hoitash, R., Hoitash, U., Kurt, A. C., and Verdi, R. S. (2023). A measure of financial statement benchmarking. *The Accounting Review*, 98(6):253–281.

Hou, K., Xue, C., and Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*.

Kaustia, M. and Rantala, V. (2021). Common analysts: Method for defining peer firms. *Journal of Financial and Quantitative Analysis*, 56(5):1505–1536.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524. Earlier version: NBER Working Paper No. 24540, 2018.

Liang, P. J. (2023). Bookkeeping graphs: Computational theory and applications. *Foundations and Trends in Accounting*, 17(2):77–172.

Liang, P. J. (2025). Three entropy measures of double-entry bookkeeping graph classification structure. Position paper.

Michaely, R. and Womack, K. L. (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *The Review of Financial Studies*, 12(4):653–686.

Pyo, J. J., Zhang, G., and Liang, P. J. (2025). Measuring information content of financial statement structure. Technical report, Carnegie Mellon University. Working Paper.

Rosenberg, B. and Marathe, V. (1976). Common factors in security returns: Microeconomic determinants and macroeconomic correlates. Technical Report 44, Research Program in Finance, University of California, Berkeley.

Vamvourellis, D., Tóth, M., Mehta, D., Bhagat, S., Pasquali, S., and Desai, D. (2023). Company similarity using large language models.

Vorst, P. and Yohn, T. L. (2018). Life cycle models and forecasting growth and profitability. *The Accounting Review*, 93(6):357–381.

Zhang, Y., Lu, Y., Mao, H., Huang, J., Zhang, C., Li, X., and Dai, R. (2023). Company competition graph.

Tables

Table 1: Practitioner Peer-Identification Approaches: Inputs, Objectives, and Outputs

Method	Inputs	Objective / Use Case	Output / Peer Definition
Barra fundamental factor models	Standardized descriptors from financial statements, analyst forecasts, and market data (e.g., size, value, growth, leverage, momentum, volatility, currency sensitivity, industry membership)	Portfolio risk control and attribution; monitor exposures and unintended bets; performance attribution	Factor-space neighbors: firms with similar exposures in a high-dimensional risk-factor space (systematic comparability)
Analyst- or vendor-selected comparables	Analyst judgment, Bloomberg comps, investment bank reports	Valuation and target-price setting; anchoring multiples and relative performance	Discrete, hand-selected peer lists; often influenced by incentives and institutional context
LLM-based embeddings (e.g., BlackRock)	Embeddings of 10-K Item 1 business descriptions (BERT, SBERT, GPT, PaLM, etc.)	Improve clustering of firms for return comovement and risk modeling; move beyond rigid taxonomies (GICS/SIC)	Continuous similarity scores in text-embedding space; clusters explain higher share of return variation than GICS buckets
Our approach: Bookkeeping-graph entropy peers (AG-D)	Double-entry bookkeeping graphs; node, entropy summaries; Jensen–Shannon distance	Identify structurally comparable firms for valuation, forecasting, and return prediction	Continuous similarity in accounting-system structure (stocks, flows, and topology); transparent peer sets grounded in accounting laws

Notes: Table compares key practitioner and emerging approaches to peer identification. Barra models focus on systematic risk attribution, analyst/vendor sets on subjective comparability, LLM-based embeddings on text-driven similarity, and our AG-D based Phoenix Peers on structural accounting comparability.

Table 2: **Baseline Valuation Regressions (1987–2024).** This table reports results from Fama-MacBeth regressions of future Enterprise-Value-to-Sales ($EVS_{t+\tau}$) on peer multiples. Horizons $\tau = 1, 2, 3$ are reported. t-statistics are reported in parentheses.

	Horizon $t+1$			**Horizon $t+2$**			**Horizon $t+3$**		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.3475	0.3527	-0.1609	0.4176	0.4206	-0.0843	0.5235	0.5266	0.0628
	(2.88)	(2.98)	(-1.49)	(4.26)	(4.32)	(-0.74)	(5.18)	(5.22)	(0.59)
EVS_{Ind}	0.7968	0.6486	0.2881	0.7679	0.6455	0.3279	0.7400	0.6082	0.3146
	(14.91)	(13.39)	(11.56)	(13.86)	(13.96)	(15.98)	(13.90)	(12.97)	(8.94)
$EVS_{IndSize}$		0.1459	0.0822		0.1199	0.0633		0.1291	0.0776
		(7.76)	(5.21)		(6.54)	(3.88)		(4.92)	(3.52)
$EVS^{(4)}_{Phoenix}$			**0.7140**			**0.6664**			**0.6173**
			(14.54)			(11.38)			(10.24)
Avg. R^2	0.124	0.131	0.388	0.112	0.117	0.300	0.101	0.106	0.254
Adj. R^2	0.123	0.130	**0.387**	0.111	0.115	**0.298**	0.100	0.105	**0.253**
Years	37	37	37	36	36	36	35	35	35
Obs.	78,063	78,063	78,063	70,495	70,495	70,495	63,782	63,782	63,782

Table 3: **Peer Similarity Matrix.** This table reports the mean Jaccard similarity coefficient between peer sets constructed using different methodologies. Low values indicate that the methods select fundamentally different firms.

Method	Industry	Ind-Size	Phoenix (AG)
Industry	1.0000	0.0949	0.0006
Ind-Size	0.0949	1.0000	**0.0004**
Phoenix (AG)	0.0006	0.0004	1.0000

Table 4: **Sensitivity to Peer Set Size (K).** Regression of EVS_{t+1} on controls and Phoenix peers constructed using the top K neighbors.

	Number of Phoenix Peers (K)			
Variable	$K=3$	$K=4$	$K=5$	$K=10$
Intercept	-0.1129	-0.1363	-0.1561	-0.2344
	(-1.04)	(-1.26)	(-1.45)	(-2.17)
EVS_{Ind}	0.2950	0.2776	0.2614	0.2223
	(11.73)	(11.44)	(11.09)	(10.39)
$EVS_{IndSize}$	0.0781	0.0767	0.0728	0.0709
	(5.00)	(4.99)	(4.87)	(4.71)
$EVS^{(K)}_{Phoenix}$	0.6801	0.7175	0.7553	0.8490
	(14.73)	(14.65)	(14.70)	(15.69)
Adj. R^2	0.377	0.388	0.398	0.422
Obs.	75,479	75,479	75,479	75,479

Table 5: **Robustness to Search Constraints and Sample Coverage.** This table compares the performance of Phoenix Peers ($K = 4$) under different search constraints. Column (1) reports the **Baseline** unrestricted model (from Table 2). Columns (2)–(5) are reconstructed by applying the relative performance ratios from the robustness tests to the baseline figures. t-statistics are in parentheses.

Constraint	(1) **None (Baseline)**	(2) **Same SIC-2**	(3) **Same SIC-3**	(4) **SIC-2 + Size Q**	(5) **SIC-3 + Size Q**
Intercept	-0.1609	-0.0164	-0.0971	-0.2581	-0.5179
	(-1.49)	(-0.17)	(-0.72)	(-1.34)	(-1.77)
EVS_{Ind}	0.2881	0.0860	0.1422	0.2667	0.4855
	(11.56)	(3.61)	(3.30)	(7.83)	(6.80)
$EVS_{IndSize}$	0.0822	0.0654	0.0760	0.0310	0.1003
	(5.21)	(3.77)	(3.07)	(1.00)	(1.92)
$EVS^{(4)}_{Phoenix}$	**0.7140**	**0.7509**	**0.7454**	**0.7520**	**0.6700**
	(14.54)	(14.41)	(13.49)	(12.51)	(11.24)
Avg. R^2	0.388	0.406	0.415	0.403	0.416
Adj. R^2	0.387	0.405	0.411	0.399	0.408
Total Obs.	**78,063**	**53,397**	**26,864**	**19,832**	**8,758**
% Sample Lost	–	*-31.6%*	*-65.6%*	*-74.6%*	*-88.8%*

Table 6: **Sub-period Stability (1996–2016)**

	Horizon $t+1$			Horizon $t+2$			Horizon $t+3$		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.2322	0.2406	-0.2606	0.3601	0.3628	-0.1076	0.4131	0.4151	-0.0748
	(1.57)	(1.64)	(-1.50)	(2.55)	(2.60)	(-0.59)	(3.86)	(3.97)	(-0.49)
EVS_{Ind}	0.8326	0.6859	0.2947	0.7831	0.6390	0.2971	0.7522	0.6066	0.2773
	(10.67)	(9.35)	(8.57)	(10.12)	(10.73)	(12.08)	(10.43)	(10.23)	(6.48)
$EVS_{IndSize}$		0.1432	0.0726		0.1416	0.0822		0.1421	0.0847
		(7.96)	(5.27)		(5.98)	(4.20)		(3.50)	(2.53)
$EVS^{(4)}_{Phoenix}$			0.7565			0.6804			0.6784
			(9.90)			(8.44)			(7.14)
Avg. R^2	0.118	0.125	0.407	0.101	0.106	0.305	0.090	0.096	0.261
Adj. R^2	0.117	0.124	0.406	0.100	0.105	0.304	0.089	0.095	0.260
Obs.	50,004	50,004	50,004	46,401	46,401	46,401	43,248	43,248	43,248

Notes: This table reports results from Fama-MacBeth regressions of future Enterprise-Value-to-Sales ($EVS_{t+\tau}$) on peer multiples for the restricted sample period 1996–2016. $EVS^{(4)}_{Phoenix}$ denotes the mean multiple of the top 4 peers identified via Accounting-Graph Distance. t-statistics (in parentheses) are Newey-West adjusted with a lag of 3.

Figures

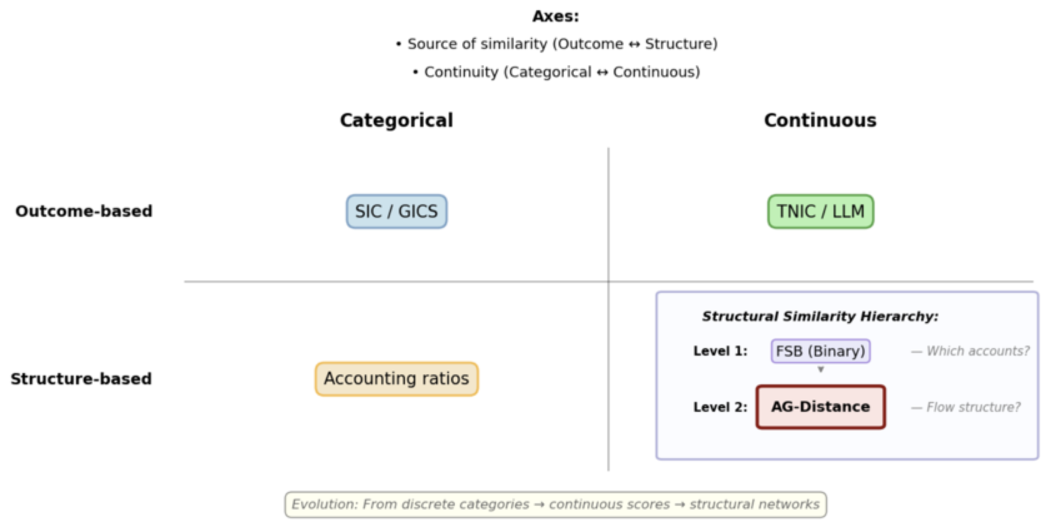

Figure 1: Taxonomy of Peer-Selection Methods

Notes: This figure classifies peer-selection approaches along two dimensions: the *source of similarity* (outcome-based vs. structure-based) and *continuity* (categorical vs. continuous). The outcome-based row captures methods grounded in observable market or product-space characteristics: SIC/GICS codes provide discrete industry bins, while TNIC and LLM-based embeddings yield continuous similarity scores derived from textual disclosures. The structure-based row captures methods grounded in accounting system characteristics: traditional accounting ratios (e.g., ROE, leverage, asset turnover) are inherently continuous but are typically discretized into bins or screens for peer selection, placing them in the categorical column. The lower-right quadrant represents continuous structural approaches organized as a hierarchy: FSB (Hoitash et al., 2023) measures similarity in *which* line items firms report (Level 1), while our AG-Distance measure captures similarity in *how* value flows through the double-entry system (Level 2). The arrow at the bottom illustrates the methodological evolution in peer selection—from discrete taxonomies toward continuous scores and, ultimately, toward structural network representations.

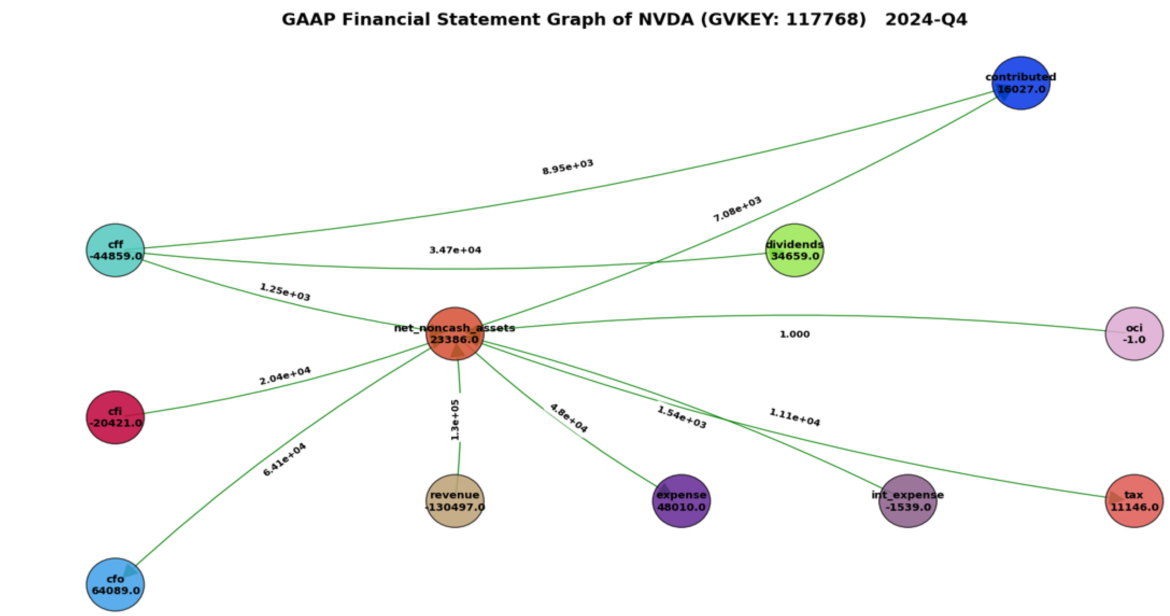

Figure 2: **GAAP Financial Statement Graph (NVIDIA Corp, 2024-Q4)**

Notes: This figure visualizes the accounting graph $G = (V, E, W)$ for NVDA. Nodes represent financial statement accounts (e.g., Revenue, CFO, CFI), and directed edges represent the flow of value (weights) satisfying the double-entry bookkeeping identity.

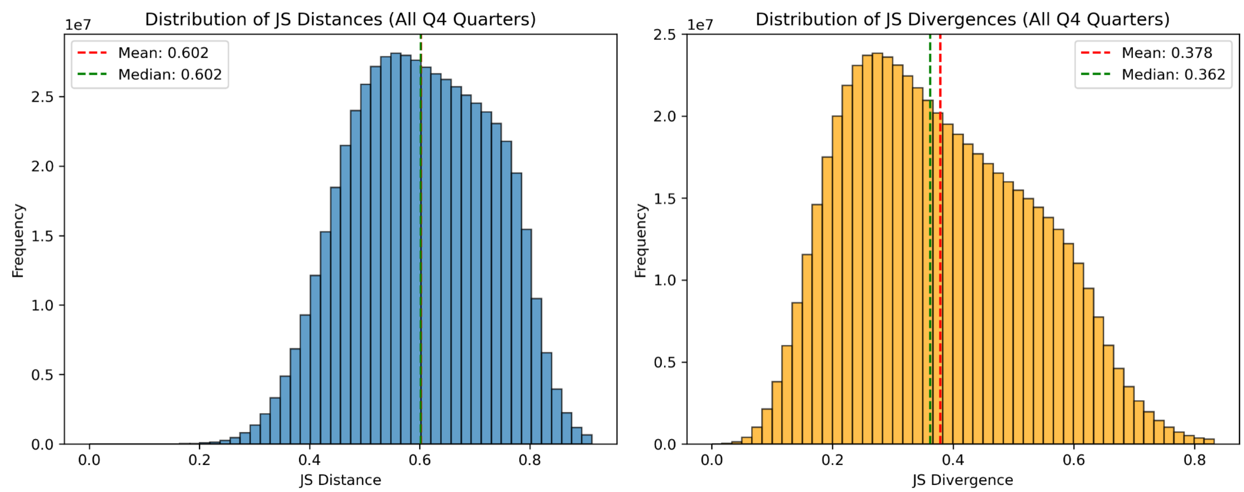

Figure 3: **Distribution of Pairwise Jensen-Shannon Distances and Divergences**

Notes: This figure plots the empirical distribution of pairwise structural comparisons across all Q4 firm-quarters in the sample. The left panel displays the Jensen-Shannon Distance (the metric used for Phoenix Peer selection), which exhibits a near-symmetric distribution centered around a mean of 0.602. The right panel displays the underlying Jensen-Shannon Divergence. The wide dispersion confirms that accounting-graph distance effectively captures significant cross-sectional variation in operational structure.

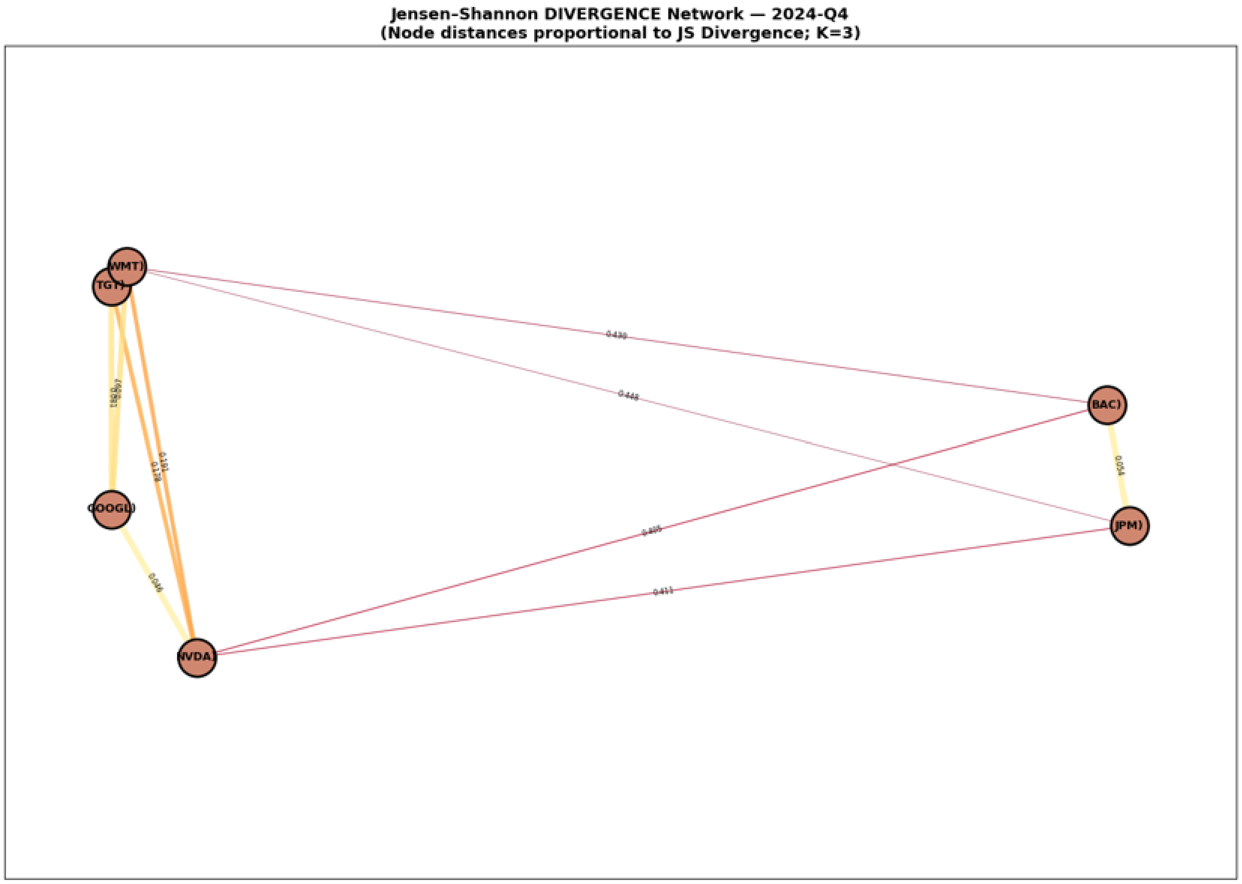

Figure 4: **Jensen-Shannon Divergence Network Visualization (2024-Q4)**

Notes: This network diagram visualizes the structural proximity between six major firms across different sectors. Edges represent the pairwise Jensen-Shannon distance between accounting graphs ($K = 3$). Shorter, thicker edges indicate high structural similarity (low divergence). The layout demonstrates that the accounting-graph distance naturally recovers economic clusters—grouping Retail (WMT, TGT), Tech (NVDA, GOOGL), and Banking (BAC, JPM)—based solely on the topology of their financial statements.

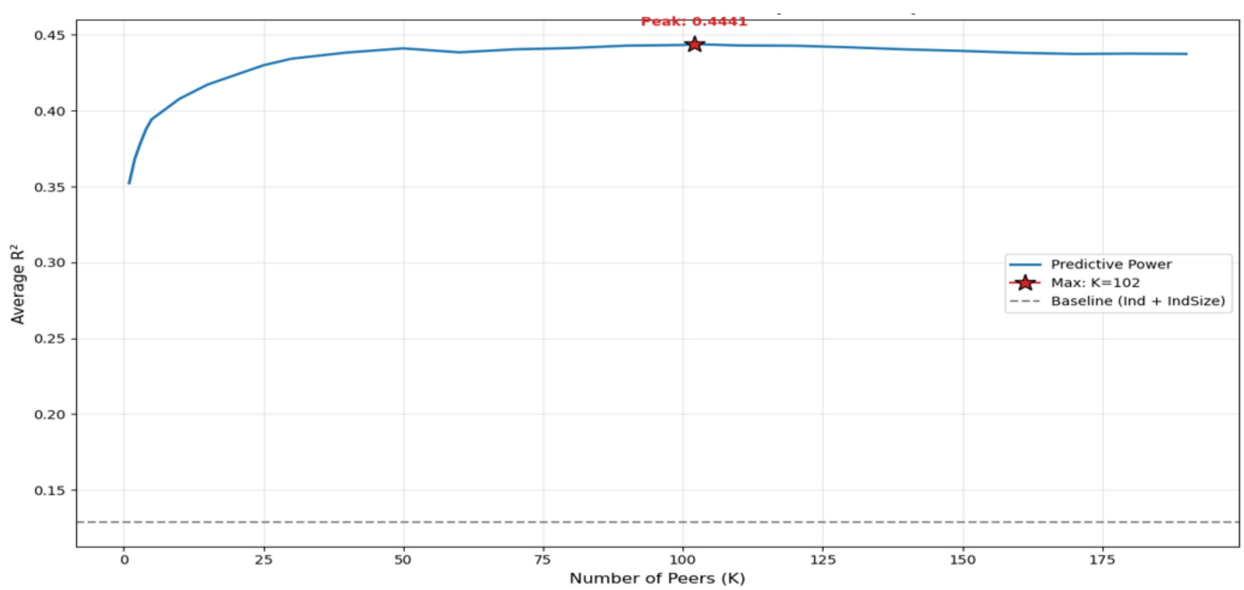

Figure 5: **Sensitivity of Valuation Accuracy to Peer Set Size (K)**

Notes: This figure plots the Average R^2 from the baseline valuation regressions as the number of Phoenix Peers (K) varies from 1 to 200. The gray dashed line represents the baseline performance of the standard Industry + Industry-Size model ($R^2 \approx 13\%$). The solid blue line shows the performance of the Accounting-Graph Distance model. While the global maximum is reached at $K = 102$ (Red Star, $R^2 \approx 44.4\%$), the curve exhibits strong concavity with sharply diminishing marginal returns after the first few neighbors. This trajectory supports the selection of a parsimonious $K = 4$ for the main analysis, capturing the majority of predictive power without over-expanding the peer set.