Forecasting with News-Implied Firm Networks

V. Hilt & G. Schwenkler*

October 6, 2026†

Abstract

Firms are connected through multiple economic networks, but whether different forms of connectivity contain distinct information about aggregate conditions is unknown. Using AI and four decades of financial news, we measure credit, equity, leadership, peer, and production networks on a common basis. Credit connectivity is particularly informative about interest rates, equity connectivity about financial markets, and production connectivity about real activity. Across networks, density and concentration generally forecast weaker conditions, while bridges forecast stronger activity. These signals predict the directions of consumption growth and long-term yields and support profitable market-timing strategies using previously published news.

Keywords: firm networks, aggregate fluctuations, asset pricing, predictability, artificial intelligence. JEL codes: C80, D20, E32, G12, G32.

*Schwenkler is at the Department of Finance, Santa Clara University Leavey School of Business. Hilt is at Wellington Management. Schwenkler is corresponding author. Email: gschwenkler@scu.edu, website.

†This paper previously circulated under the title "The Different Networks Of Firms Implied By The News." We are grateful to Don Bowen (discussant), Will Cong, Sanjiv Das, Teodor Dyakov (discussant), Daniel Giamouridis, Gerard Hoberg, Alexandr Kopytov (discussant), Christos Makridis, Alireza Tahbaz-Salehi, Rossen Valkanov (discussant), and Alexander Zentefis; seminar participants at City University of Hong Kong, Qube Research & Technologies, Santa Clara University, and Stanford University; and conference participants at the 2024 GenAI In Finance Conference, UNSW Asset Pricing Workshop, Boca Finance & Real Estate Conference, the 2025 Contemporary Issues in Financial Markets and Banking, JOIM Spring Conference, FIRS Conference, SoFiE Annual Meeting, Future Finance Fest, Santiago Finance Workshop, the 2026 Network Science in Economics Conference, EDHEC AI and Finance Workshop, and QFFE Conference for useful comments and suggestions.

1 Introduction

Firms are linked by credit, ownership, leadership, competition, production, and other types of relationships. These links support financing, monitoring, and information diffusion, but they can also transmit distress, create common exposures, and weaken competition. A large literature studies whether particular firms are connected and what those bilateral relationships imply for firm outcomes and asset prices. Table 1 summarizes important channels.

The aggregate question is different. Two economies can have the same number of firm links yet very different architectures. Links can be broadly dispersed, concentrated around a few hubs, or organized around firms that bridge otherwise separate clusters. Acemoglu et al. (2012) show that these features govern whether firm-level shocks dissipate or become aggregate fluctuations. Empirically measuring the firm network architecture through time is therefore essential for forecasting the conditions that matter to investors, regulators, and policymakers. A central obstacle is data availability. Existing datasets typically record only one relationship, cover sparse or changing sets of firms, or update infrequently. Subsidiary and board databases, industry classifications, customer disclosures, and input-output tables use different firm universes and conventions. Networks inferred from prices offer broader coverage but can be difficult to interpret economically.[1] Comprehensive public data are also unavailable for some important relationships, especially firm-level credit exposures. These limitations make it difficult to compare several forms of connectivity within a common information set or test whether changes in their aggregate architecture predict future conditions.

We close this measurement gap by exploiting a simple idea: financial news often describes firm relationships. We build an AI-based method that constructs economically labeled firm networks from financial news over flexible aggregation intervals, and use these to compare how different relationship types and network architectures predict future aggregate outcomes. Credit connectivity is most informative about bond market conditions, equity connectivity about returns and volatility, and production connectivity about consumption and industrial production. Across several types of firm relationships, networks that become denser or more concentrated tend to forecast adverse economic outcomes, while networks with stronger bridging across firm clusters forecast better out-

[1] See Diebold and Yılmaz (2014), Li et al. (2025), and others for related methods.

comes. These signals enable profitable market timing and accurate directional predictions. Firm connectivity therefore has several dimensions that matter when forecasting aggregate outcomes. Our findings highlight news as a valuable source of information about firm relationships. To facilitate adoption, we share our data and code in our online data repository.

An important innovation of our paper is a common measurement framework for comparing economically distinct firm networks. Our method combines modern natural language processing with supervised deep learning. We transform news sentences into semantic embeddings and use a neural-network classifier to estimate probabilities that the text describes a *credit*, *equity*, *leadership*, *peer*, *production*, or a residual *other* relationship. Aggregating these classifications yields multiple networks from the same source, over the same firm universe, at the same frequency. Existing text-based methods infer product-market proximity from 10-K disclosures using similarity scores (Hoberg and Phillips (2016)), or supply-chain links using large language models (LLMs) trained on existing relationship databases (Sun et al. (2025)). Other methods construct networks from firms' joint appearance in news (Scherbina and Schlusche (2015), Chen et al. (2021)). Complementing these approaches, we directly classify the economic relationship described by sentence-level content.

Although modest by modern AI standards, our classifier is high dimensional and contains more than 230,000 free parameters. We develop a computationally efficient training approach that combines scalable synthetic data with targeted human verification and runs on a standard laptop. We proceed in three steps. We initially train the classifier on a large, balanced sample of synthetic sentences generated by several LLMs. We then fine-tune the model using human-verified sentences from actual financial news, including difficult cases misclassified by the initial model. This design follows emerging work that combines scalable artificial-intelligence measurements with targeted human validation; see Ludwig et al. (2025) and Duan and Pelger (2026). In the last step, we round weak links down and strong links up to induce sparsity and reduce measurement errors. We provide conditions under which the aggregated link strengths converge to population reporting intensities as news coverage and training samples grow large, and present finite-sample evidence supporting those conditions. The resulting model achieves an out-of-sample F1 score on human-verified sentences of 84%, with an out-of-sample R^2 of 74%. The design is auditable because every link remains trace-

able to its source text. It can also be extended to other relationships and used to update measures of economic dependencies as news becomes available. We provide the methodological foundation underlying Schwenkler and Zheng (2026), who apply a version of our approach to study production networks, aggregate output, and firm systemic importance. Here, we measure economically distinct networks on a common basis to examine how predictive content varies with relationship type and network architecture. This comparative question cannot be addressed using a single production network. A related method has also been used by Schwenkler and Zheng (2025*a*) to establish return cross-momentum among cryptocurrency peers.

We apply the method to nearly 850,000 articles published between 1981 and 2025, combining the New York Times archive with major business outlets in the News on the Web corpus, one of the largest continuously updated collections of English-language online news. We identify more than 2.2 million firm mentions involving over 16,000 firms and more than 40,000 unique links. Sentence examples, keyword analyses, and comparisons with WRDS subsidiaries, BoardEx, the Text-based Network Industry Classifications of Hoberg and Phillips (2016), Compustat customer segments, and Bureau of Economic Analysis input-output tables show that the labeled networks capture their intended economic content rather than news co-coverage alone. The common measurement framework lets us answer an empirical question that existing data cannot cleanly address: do the number of firm connections, their concentration around a few firms, and the extent to which they bridge otherwise separate groups predict different aggregate outcomes across relationship types? Holding the firm universe, aggregation window, and information source fixed, we aggregate links over rolling 12-month windows and update *density* (how many links there are), *concentration* (how centralized links are), and *intermediation* (how much bridging there is) each month. These measures are standard summary statistics of a network's architecture; see Gofman et al. (2026) for a survey.

Looking at individual predictive regressions, we find that density and concentration generally predict lower subsequent returns and real activity. Leadership and peer concentration also predict higher volatility. Intermediation behaves differently. It predicts stronger real activity in non-credit networks, while credit intermediation predicts a steeper yield curve. Our inference uses bootstraps to account for predictor persistence that could otherwise generate spurious predictability

(Boudoukh et al. (2008), Valkanov (2003)) and p-value adjustments to address multiple-testing concerns (Harvey et al. (2016)). The predictive associations generally align with the domain of the different networks. The contrasting signs help explain how the beneficial and adverse channels in Table 1 can coexist. Density, concentration, and intermediation describe distinct economic states, so changes in firm connectivity need not have a single aggregate implication. Our findings extend the work of Herskovic et al. (2020), Pellegrino (2025), and Schwenkler and Zheng (2026) on the aggregate implications of production and peer network structure.

To assess economic significance, we construct real-time market-timing strategies and directional forecasts. Inspired by Breen et al. (1989) and Campbell and Thompson (2008), the trading strategies allocate to the three-month Treasury bill or the S&P 500 according to a connectivity metric's position relative to its five-year historical median. They maintain unit leverage while allowing index positions to remain open over multiple months. Density strategies based on all five networks achieve statistically significant Sharpe Ratios after adjustment across networks, and the best-performing strategy reaches 1.3-times the Sharpe ratio of the S&P 500. Credit, equity, and production density strategies also generate statistically positive CAPM alphas. Intermediation trading strategies perform strongly but do not survive the multiple-testing adjustment, while concentration is a weak real-time signal. Directional tests extend this evidence. Lower density forecasts positive cumulative consumption growth across all five networks, while greater density forecasts increases in the ten-year Treasury yield. Production and peer intermediation also predict positive consumption growth at distinct horizons. These tests further validate the contrast between adverse density signals and beneficial intermediation signals.

Taken together, our findings show that information contained in the news may predict aggregate outcomes at long horizons of several months to years. This may at first appear surprising. In a final step, we evaluate why this is possible. News-implied connectivity measures can reflect underlying firm relationships (Acemoglu et al. (2012), Liu and Tsyvinski (2024)), the newsworthiness of those relationships (Chahrour et al. (2021), Mullainathan and Shleifer (2005)), or information that news records before existing databases (Beaudry and Portier (2006)). Each of these channels may offer predictive power on its own. We disentangle the informational content of our news-implied

measures by using matched conventional networks to capture relationships already recorded in existing databases, and coverage frequencies to capture shifts in news reporting. When regressing our news-implied measures on conventional ones and coverage frequencies, we find that both components account for substantial but not exhaustive variation. We also find that some variation in the news-implied measures precedes subsequent changes in the existing databases. When repeating our predictive regressions while also controlling for the information reflected in existing networks and coverage frequencies, we find that the news-implied connectivity measures continue to offer predictive power. These results suggest that news-implied connectivity measures predict future aggregate outcomes because they reflect fundamental relationships between firms while also capturing coverage activity that anticipates changes in existing databases and drives attention and salience effects. They highlight financial news as a timely source of structured information about firm links that matters for aggregate outcomes, complementing Bybee et al. (2024), Cong et al. (2025), Gentzkow et al. (2019), Manela and Moreira (2017), and others who use text and news to measure economic conditions and forecast markets. There is, however, one important caveat: our objective is prediction rather than causal identification. Our results do not establish that changes in network architecture cause subsequent changes in aggregate outcomes.

The paper proceeds as follows. Sections 2–4 describe the data and methodology. Section 5 validates and characterizes the resulting networks. Section 6 studies aggregate predictability, and Section 7 concludes. The Online Appendix and data repository are available at `www.news-networks.net`.

2 Data

We construct a composite dataset from The New York Times, NOW corpus, CRSP, Compustat, and I/B/E/S. We provide a summary of our data collection approach in this section and offer details in the Online Appendix.

2.1 Firm population

We construct our firm population from CRSP, Compustat, and I/B/E/S. Online Appendix A provides details of our data collection. We only consider US corporations with equity securities and otherwise apply no filters. We obtain a sample consisting of 27,103 unique firms. Table 2 reports summary statistics and Online Appendix Figure A.1 displays the industry distribution.

2.2 News data

We collect news articles from two datasets that provide broad coverage of business-relevant news. For both datasets, we end the samples on December 31, 2025 because our CRSP data are updated only annually, and we exclude any company announcements.

The first news dataset is the New York Times, for which we download business news articles published after January 1, 1981 using the Developer API. The New York Times Archive contains many articles that are not about business. Because business articles are not clearly identified in the Archive and metadata properties changed over time, we develop a matching and scoring methodology to identify the subset of business-relevant articles in the Archive. We summarize the approach in Online Appendix B. It yields a sample containing 466,253 business news articles from the New York Times published during our sample period. Note that the Terms of Use of the New York Times Developer API restricts using these data for the training of AI models. As a result, we never use New York Times data to train our model. We only apply our methodology to New York Times data to construct news-implied networks.

Figure 1 shows the number of monthly business articles in our New York Times sample. There, we observe that the number of monthly articles has significantly decreased since the mid 2010s. To counteract this effect, we collect additional news data from the News on the Web (NOW) corpus, which we purchase from corpusdata.org. It contains the bodies of articles written in English in online magazines and newspapers across the world since January 2010. Covering more than 25 billion words, this dataset represents the largest full-text corpus of online news. One benefit of the NOW corpus is that it permits the usage of its data for the training of AI models. Thus, whenever we need to train or update our model using published news data, we use data from the NOW

corpus. As for the New York Times corpus, the NOW corpus does not allow sharing of the data. Thus, we do not make any raw news data available in our online repository.

We impose certain filters to reduce the size of the NOW corpus data we consider. First, we restrict the sample to articles published in reputable outlets that report on business matters in an editorially rigorous way. The set of outlets we consider includes Barron's, Bloomberg, Business Insider, CNBC, Forbes, Fortune, MarketWatch, Reuters, and The Wall Street Journal. Second, we remove any ads and opinion pieces that may be biased or inaccurate. We also filter out any content that is deemed to be advertisement, affiliated or sponsored, calling to action, or related to opinion pieces or editorial notes. The two filters fulfill an important requirement of our approach posed by Theorem 3.1, whereby we need news reporting about firm links to be truthful and unbiased. We describe how we filter through the NOW Corpus and collect articles from these sources in Online Appendix C. We obtain a sample containing 382,826 articles.

Online Appendix Table A.1 reports the distribution of news articles and their temporal span across the ten outlets we consider. While the New York Times is our largest source and spans the full sample period, Forbes, CNBC, and Business Insider together contribute nearly 280,000 more articles through the end of 2025. We attribute this to paywalls that may have prevented NOW from scraping for some sources, causing coverage to end before our sample. Figure 1 shows the time series of coverage from the different news outlets. We observe that our sample grew large starting in the late 2010s, when online news articles from Forbes, CNBC, and MarketWatch proliferated. There was a temporary spike in coverage from The Wall Street Journal around the COVID pandemic.

2.3 Firm identification in news data

We apply the methodology of Schwenkler and Zheng (2026) to identify firm mentions in our news data and construct sentence-based links.[2] The methodology proceeds in four steps. It first uses a Named Entity Recognition algorithm (NER) to identify and classify any entities named in the articles. For those entities labeled as organizations, it follows heuristics to separate firms from non-

[2]Our methodology does not hinge on Schwenkler and Zheng (2026). Any method that extracts sentences with two firm mentions can be used in this step, such as Chen et al. (2021) and Scherbina and Schlusche (2015). However, Theorem 3.1 requires that any firm identification algorithm used satisfies certain accuracy conditions.

firms. In the third step, it matches the sample with the CRSP database. In the fourth step, the methodology establishes a link between two firms when they are co-mentioned in a sentence and the sentence contains only two firm mentions.[3] We adopt minor modifications to address peculiarities of the NOW Corpus, primarily to remove embedded photo attribution lines that would otherwise be misidentified as firm mentions, correct tokenization artifacts, and ensure proper sentence boundary delineation before named entity recognition is applied. We also strip sequences of "@" symbols that the corpus inserts every 200 words in place of ten removed words for copyright compliance (see here). Finally, we remove repeated sentences within a day.

Table 2 provides summary statistics and Online Appendix Figure A.1 shows the industry distribution for the firms that are identified in our news sample. Over the course of our 45-year-long sample, we identify a total of 2,269,961 mentions of 16,387 distinct firms. We observe that firms covered in our news sample are mentioned around 0.74 times per month. They tend to be larger than the average firm in the population, with higher net incomes, sales, and costs of goods sold. They also have higher monthly returns and are covered by more earnings analysts. Online Appendix Figure A.1 shows that the manufacturing and retail industries are overrepresented in our news sample, while the finance industry is underrepresented.

Moving to firm links, we find 166,556 sentences with two firm mentions in our sample.[4] We establish a total of 42,959 unique links among 9,852 distinct firms. While the aggregate link count is large, links are identified only infrequently in our sample. On average, we establish 0.11 links per firm per month, indicating that only one in seven firm mentions is associated with a firm link.

Online Appendix Figure A.3 shows the implied uncategorized firm network, restricted to links that are mentioned at least 100 times in the news sample. We observe that the news-implied network has multiple clusters, each showcasing star architectures as in Acemoglu et al. (2012) while being connected through only a few intermediary firms. Goldman Sachs (GS) connects the banking cluster at the core of the network to the technology cluster through a link with Apple

[3]The analyses in Schwenkler and Zheng (2026) show that sentence-level co-mentions are sufficient to identify firm connections.

[4]The methodology finds 207,932 sentences with two or more firm mentions. Because sentences with more than two firm mentions may be more complex and 80% of the recognized sentences with firm links mention only two firms, we restrict the analysis to two-firm mentions for simplicity.

(AAPL), and the automotive cluster through a link with Tesla (TSLA). There are multiple disjoint clusters, such as the pharmaceutical cluster containing Pfizer (PFE) and Merck (MRK), and an oil and energy cluster containing Exxon Mobil (XOM) and Chevron (CVX).

3 Labeling methodology

We develop a methodology to construct labeled networks of firms. We begin by classifying sentences describing different types of firm relationships. To each sentence with two firm mentions, our methodology assigns probabilities for six possible labels: *credit*, *equity*, *leadership*, *peer*, *production*, and *other*. We define these categories as follows:

- **Credit**: Link through a claim on future payment or contingent financial support involving debt, loans, revolving facilities, notes, bonds, trade credit, or similar transactions. This category excludes any equity-like financing arrangements.
- **Equity**: Link that arises through the ownership of shares, voting power, or a transaction that creates, changes, or removes that ownership, as would occur in parent-subsidiary relationships, private placements, or M&A transactions. It explicitly excludes credit relationships.
- **Leadership**: Link due to the same person currently or recently holding leadership or governance roles at two firms. It excludes co-mentions of leaders from different firms and focuses on one leader shared across firms.
- **Peer**: Link between market rivals, competitors, substitutes, or benchmark peers. It does not consider analyst co-coverage unless an analyst explicitly links the two firms as peers.
- **Production**: Link that arises through the flow of goods, inputs, services, logistics, manufacturing, or operational dependency. It explicitly excludes trade credit or professional services not linked to output.
- **Other**: Any residual link that does not explicitly fall into any of the five categories above.

Online Appendix D provides detailed definitions and examples of our link categories.

We proceed as summarized in Figure 2 to classify the content of two-firm sentences. We first take the sentences and run them through an open-source sentence transformer algorithm called "all-mpnet-base-v2" provided by Hugging Face in Python. The transformer converts the text of the sentence into a numerical representation in the form of a 768-dimensional vector embedding that captures its most relevant semantic features. The transformer model was trained for the purpose of information retrieval and sentence similarity tasks using over 1 billion paired sentences from Reddit, WikiAnswers, Yahoo Answers, and others. Because our goal is to identify sentences that describe similar types of business relationships, this transformer model works well for our purposes. In unreported experiments, the performance of our method did not improve when we used more complex embeddings, such as "bge-large-en-v1.5" by the Beijing Academy of Artificial Intelligence (BAAI). We therefore present all results based on the "all-mpnet-base-v2" embedding.

Once we obtain the embeddings, we run them through a deep learning classification model that assigns link category probabilities. The classifier is a feed-forward neural network. Its first hidden layer contains 256 units and its second hidden layer contains 128 units. Both use ReLU activation, and each is followed by dropout regularization at a rate of 0.3. A six-unit output layer uses a softmax function to compute a probability for each link category. The classifier contains 230,534 trainable weights and biases. More precisely, for Sentence i contained in an article published at Time t covering Firms f and g, the classifier computes the probability $p_{i,t,c}(f,g)$ that the sentence, as represented by its embedding, describes a relationship between f and g that falls into Category c, where c can be "credit," "equity," "leadership," "peer," "production," or "other." The output probabilities are constrained to sum to one:

$$p_{i,t,\text{credit}} + p_{i,t,\text{equity}} + p_{i,t,\text{leadership}} + p_{i,t,\text{peer}} + p_{i,t,\text{production}} + p_{i,t,\text{other}} = 1. \tag{1}$$

The probabilities can therefore be interpreted as the model's confidence that the sentence describes that type of relationship. The training of our deep learning model is intricate. We provide an overview in Section 4.

Table 3 provides examples of how our methodology works, with the tickers of the two identified firms marked in parentheses. Our methodology correctly labels the first sentence as almost cer-

tainly describing a credit relationship, matching the information about JPMorgan's loan to Lehman Brothers. It also correctly infers a production relationship from the second sentence, which mentions in passing that Sony is a customer of Intel when reporting on Mr. Abary's press conference. Finally, when faced with a sentence stating that analysts at JPMorgan upgraded their forecasts for Teva, the methodology assigns the label "other."

3.1 Labeled firm networks

To construct a network for a given Category c (e.g., $c = \textit{credit}$), we first choose a time horizon $[t_0, t_1)$ with $t_0 < t_1$ over which we aggregate the network. We collect all classified sentences contained in articles published during the time interval $[t_0, t_1)$ and recover their link category probabilities $p_{i,t,c}(f, g)$. We construct a proxy $\tilde{p}_{i,t,c}(f, g)$ for the sentence link strength by rounding whenever the estimated link probability is sufficiently high or low as follows:

$$\tilde{p}_{i,t,c}(f,g) = \begin{cases} 1, & \text{if } p_{i,t,c}(f,g) \geq p^+, \\ 0, & \text{if } p_{i,t,c}(f,g) \leq p^-, \\ p_{i,t,c}(f,g), & \text{otherwise.} \end{cases} \tag{2}$$

Here, p^- and p^+ with $0 < p^- < p^+ < 1$ are additional parameters that we specify in Section 4. Finally, we aggregate sentence link strength proxies by averaging them:

$$\tilde{p}_{c,[t_0,t_1)}(f,g) = \frac{1}{N_{[t_0,t_1)}(f,g)} \sum_{i \in \mathcal{S}_{[t_0,t_1)}(f,g)} \tilde{p}_{i,t,c}(f,g), \tag{3}$$

where $\mathcal{S}_{[t_0,t_1)}(f, g)$ is the set of sentences mentioning Firms f and g published during the time interval $[t_0, t_1)$ that were run through our model, and $N_{[t_0,t_1)}(f, g) = |\mathcal{S}_{[t_0,t_1)}(f, g)|$ is the size of this set. We set $\tilde{p}_{c,[t_0,t_1)}(f, g) = 0$ whenever $N_{[t_0,t_1)}(f, g) = 0$ to ensure well-defined measurements. The network for Category c that applies for the time horizon $[t_0, t_1)$ contains all Firms f and g linked this way, with link strengths given by $\tilde{p}_{c,[t_0,t_1)}(f, g)$. Importantly, our method does not impose a calendar frequency; users choose the time interval. Theorem 3.1 shows that estimated link strengths

converge as the amount of relevant news and the sizes of the training samples grow large.

While we could also construct networks directly based on the estimated link probabilities $p_{i,t,c}$, rounding serves two purposes. First, it enforces network sparsity by discarding co-mentions where the model assigns low confidence to any specific link type. Second, it makes the aggregated link strength robust to classification noise. By replacing near-certain predictions with hard assignments of zeroes or ones, the average in Eq. (3) is driven by sentences where the model is confident, and small errors in the estimated probabilities for those sentences have little impact on the final link strength estimate.

3.2 Convergence

The link strength in Eq. (3) approximates the population reporting intensity with which a firm link is reported in the news. The reporting intensity may be related to the true strength of the underlying firm relationship, and we provide evidence in Sections 5.4 and 6.4 that this is indeed the case. In finite samples, however, Eq. (3) is a noisy estimate. There are several sources of noise. First, our links are based on news reporting and the news is well-known to exhibit biases towards certain types of firms (Fang and Peress (2009)). Thus, if the news does not report about all firm linkages, our networks may exhibit biases. Our network construction approach may also introduce errors. For one, we use the methodology of Schwenkler and Zheng (2026) to identify firm mentions in sentences. We also use our own deep learning methodology to classify the contents of the different sentences. These methodologies are naturally noisy. The following theorem provides conditions for the convergence of our methodology. The proof can be found in Online Appendix J.

Theorem 3.1. *Consider Firms f and g and fix the Link Category c. For times $0 \leq t_0 < t_1 < \infty$, write $\pi_{c,[t_0,t_1)}(f,g) \in [0,1]$ as the population frequency with which the news reports that Firms f and g share a relationship of Type c during time interval $[t_0,t_1)$. Let N_{firm} and N_{class} denote the sizes of the data samples used to train the firm identification algorithm and the link classification model, respectively. Assume that the following conditions hold:*

(A1) The news reports truthfully with stable intensity. This means that, conditional on Firms f and g actually sharing a relationship of Type c during $[t_0,t_1)$, the empirical frequency with

which the news reports this relationship among sampled sentences that actually mention both Firms converges in probability to $\pi_{c,[t_0,t_1)}(f,g)$ *as the number of sampled sentences grows large* ($N_{[t_0,t_1)}(f,g) \to \infty$). *If Firms* f *and* g *do not share a Type-c relationship, then the empirical reporting frequency converges in probability to zero and* $\pi_{c,[t_0,t_1)}(f,g) = 0$.

(A2) There is vanishing ambiguity in the way the news reports about firm relationships. This means that the share of sampled sentences that actually mention both Firms f *and* g *and report more than one relationship type converges in probability to zero as the number of sampled sentences grows large* ($N_{[t_0,t_1)}(f,g) \to \infty$).

(A3) The firm identification algorithm is consistent in the sense that the share of sentences incorrectly included in or excluded from the set of sentences mentioning both Firms f *and* g *becomes negligible in probability as the size of the training sample grows large* ($N_{firm} \to \infty$).

(A4) The labeling methodology is consistent. This means that, for any fixed Link Type c *and almost every non-ambiguous Sentence* i, *the probability estimate* $p_{i,t,c}(f,g)$ *converges in probability to one if the underlying Sentence* i *actually reports a relationship of Type* c *and it converges to zero otherwise as the training sample grows large* ($N_{class} \to \infty$).

(A5) The thresholds p^- *and* p^+ *are bounded away from zero and one. That is, there exists an* $\epsilon_0 > 0$, *so that* $0 < \epsilon_0 \leq p^- < p^+ \leq 1 - \epsilon_0 < 1$.

Then, the discrepancy between the link strength approximation in Eq. (3) and the population reporting intensity becomes vanishingly small as the number of sampled sentences and the training samples grow large. More precisely, for any $\epsilon > 0$:

$$\lim_{N_{[t_0,t_1)}(f,g)\to\infty} \lim_{N_{class},N_{firm}\to\infty} \mathbb{P}\left(\left|\tilde{p}_{c,[t_0,t_1)}(f,g) - \pi_{c,[t_0,t_1)}(f,g)\right| > \epsilon\right) = 0$$

Subject to the thresholds p^- and p^+ falling inside the unit interval (Assumption (A5)), Theorem 3.1 decomposes the accuracy of our network estimates into three sets of conditions: the informativeness and clarity of the news sources we use (Assumptions (A1) and (A2)), the consistency of the firm identification algorithm (Assumption (A3)), and the consistency of the link classifier

(Assumption (A4)). Assumptions (A3)–(A5) are standard. Assumption (A1) formalizes the documented selectivity of news reporting (Mullainathan and Shleifer (2005), Chahrour et al. (2021)), while Assumption (A2) requires that sentences reporting more than one firm relationship, such as those describing peers merging or entering production partnerships, are asymptotically rare.

Theorem 3.1 is asymptotic in nature and its conditions can only be verified approximately in finite samples. Online Appendix J presents an empirical validation exercise, which we summarize here. Assumption (A1) justifies focusing on New York Times articles and reputable, fact-checked business publications such as the Wall Street Journal and CNBC, which makes systematic misreporting of non-existent firm links unlikely. Assumption (A2) supports our choice of a multi-class classifier whose class probabilities are complementary and satisfy Eq. (1). If violated, independent binary classifiers as in Section 4.4 would be more suitable. Online Appendix Table A.6 shows that strong multi-label classifications are rare, supporting our approach. Assumption (A3) motivates our choice of the firm identification algorithm of Schwenkler and Zheng (2026), which achieves a precision of 96.5% and a recall of 48.9%. The recall estimate is based only on ticker-coded mentions and therefore excludes other valid mentions across the corpus. Imperfect recall does not create false co-mentions, but it reduces and may change the composition of the sentences available for estimating each link. Assumption (A3) requires these omissions to become asymptotically negligible. Finally, Online Appendix J provides finite-sample evidence supporting Assumption (A4) using the training and testing accuracy analysis of the next section.

4 Model training, testing, and extensions

Our methodology is high dimensional. The neural classifier contains 230,534 trainable weights and biases. Two additional thresholds, p^+ and p^-, govern the rounding in Eq. (2). Fitting such a complex model requires a fine-tuned approach. We proceed in three stages to train our model. In the first stage, we run an initial calibration of our model using synthetic sentences constructed using large language models (LLMs). In the second stage, we update the model using human-verified sentences. In the third stage, we fix the rounding parameters p^+ and p^-. All codes used to train and validate our model are available in our data repository.

4.1 Stage 1: Synthetic data

In the first stage, we train our model using synthetic data generated via large language models (LLMs). For this, we create synthetic sentences that clearly describe the different types of firm relationships using several LLM models: GPT-4o, GPT-5.4, Gemini 1.0, and Gemini 1.5. For each firm link type (credit, equity, leadership, peer, production, and other), we prompt the LLMs to generate 1,500 sentences that describe such a relationship between two fictional firms. We intentionally ask to construct sentences for fictional firms to not contaminate the learning with the names of firms that may normally be associated with a certain type of relationship. For example, we do not want "Citi" to always be associated with a credit relationship just because it is a bank or "Morningstar" to always be associated with the "other" label just because it provides market research. We ask to construct diverse sentence structures that resemble the way The Wall Street Journal, Financial Times, or The New York Times would describe relationships between firms. We also ask to consider different types of relationships for each label. For example, for a credit relationship, a sentence could describe one firm getting a loan from a bank or it could describe a customer getting trade credit from a supplying firm. We share our synthetic dataset in our data repository. In total, our synthetic training dataset contains 9,000 sentences that are equally distributed across the six firm relationships.

Training with synthetic data has advantages over relying exclusively on human-labeled sentences. Human labeling may produce unbalanced samples that confound learning because some link types, such as peer links, are reported more frequently. Using an LLM allows us to create balanced samples. Human labeling is also costly. LLMs can generate precise sentences quickly and cheaply (albeit at a possibly large environmental cost). Human labeling is also subjective because it is based on how individual people interpret different sentences. Despite its advantages, training based on synthetic data has one major weakness. Synthetic sentences are likely simple and may not fully represent the complexity of journalistic reporting. We address this weakness in Stage 2 by updating the model through few-shot learning on a smaller sample of published news sentences with manually verified link labels. Thus, the model estimated in Stage 1 is only a starting point and not the final model we use to construct networks.

We proceed as follows to train, validate, and test our model in Stage 1. We first create a categorical value for each link type and associate it with the corresponding embedded sentences. We randomly split the synthetic sentence universe into a training set (60% of the sample), a cross-validation set (20%), and a test set (20%), each containing a balanced number of sentences from every link type. Given the high complexity of the model, we use advanced machine-learning techniques to train it with reasonable computational resources (training runs on a standard MacBook laptop). We fit the model by minimizing the categorical cross-entropy loss and repeatedly evaluate its performance in the validation set to determine certain hyperparameters. We use the RMSprop adaptive learning-rate algorithm to scale how quickly the model learns and process randomly ordered mini-batches of 32 sentences at each epoch. To prevent the model from overfitting to the training data, we apply dropout regularization and randomly deactivate a share of the network's connections during training. We allow at most 50 training epochs but halt earlier when the validation loss fails to improve for three consecutive epochs. We then restore the parameters from the best-performing epoch. The full technical specification of our training procedure is described in Online Appendix E. Finally, we evaluate the out-of-sample accuracy of the model in two test sets: one containing the remaining 1,800 synthetic sentences that were not used for training or cross-validation and another containing 300 actual news sentences from the NOW corpus with human-verified link labels. We describe our approach to constructing the human-verified dataset in Section 4.2.

Table 4 presents the test root mean squared error, R^2, precision, recall, and F1 score of the model estimated in Stage 1.[5] We observe that the Stage 1 model is highly accurate when labeling synthetic sentences but less accurate for actual news sentences. While the overall F1 score is 98% in the synthetic test set, it is only 79% in the human-verified test set. The decline in accuracy arises primarily from difficulty labeling credit, peer, production, and other links. We observe that the Stage 1 model has low recall but high precision for peer, production, and other links in actual news data. This means that the Stage 1 model has difficulty identifying sentences that describe these links between two firms. However, once it recognizes such a link, it is usually correct. In

[5]Precision (recall) is the share of true positives among predicted (actual) positives. F1 is their harmonic mean.

contrast, the Stage 1 model has low precision but high recall for credit sentences. This means that the model does well at identifying actual credit links but only because it tends to overstate the number of sentences that describe credit. This can also be observed through the confusion matrix in Table 4, which shows that the Stage 1 model overpredicts credit links in the human-verified test sample. The low precision rate for credit links also leads to a low out-of-sample R^2. Still, the Stage 1 model does well at labeling equity and leadership links, with F1 scores above 97%.

4.2 Stage 2: Few-shot update

We augment the Stage 1 model with a few-shot update using human-verified sentences. The general idea is as follows. We first select sentences that were published in our news sample and have human reviewers verify their labels. This set is used to further train our model starting from Stage 1. We intentionally fine-tune the Stage 1 model rather than completely recalibrating it in Stage 2 to preserve the knowledge gained from the large synthetic sample while also incorporating domain-specific knowledge through human verification. Duan and Pelger (2026), Ludwig et al. (2025), and others show that this approach can improve the performance of AI models trained on synthetic data. Next, we evaluate the performance of the updated model in a separate set of human-verified sentences that is held out for testing purposes and never used for training. Finally, we take the updated model resulting from Stage 2 and use it to label all sentences published in both the NOW and New York Times corpora.

We construct our sample of human-verified news sentences as follows. We first randomly pick sentences from the NOW corpus and manually assign the most appropriate label until 100 sentences are collected per relationship type.[6] We focus on sentences that only mention two firms and describe a clear relationship. This gives us an initial random sample of 600 actual news sentences with human-verified labels. We split this initial set equally into a training set containing 300 sentences used to run a few-shot update of the Stage 1 model and a test set containing 300 sentences used to evaluate the resulting Stage 2 model. We ensure that the training and test sets have balanced category memberships.

[6] Given the licensing restrictions we face, we only consider sentences published in the NOW corpus and neglect the New York Times corpus for this exercise.

Because the preceding sentences are selected randomly, they pose average difficulty for our labeling methodology. To ensure that the model is updated using particularly difficult sentences, we complement the initial set with human-verified false positives from the Stage 1 model. To do this, we take the sentences labeled by the Stage 1 model and sort them by their predicted link probabilities. For each category, we search for 25 false positives in which the model assigns a high link probability but the sentence actually describes a different type of relationship. We ensure that the sentences are sourced from the NOW corpus to adhere to our licensing restrictions, that they are not included in the initial human-verified set, and that equal numbers of sentences across link categories are selected. For credit sentences, we are able to find only 8 false-positive sentences, although we successfully find 25 false positives per category for the remaining categories. This results in a false-positive set containing 133 human-verified sentences. These sentences are included in the human-verified training sample used to update the Stage 1 model. Overall, the human-verified training sample contains 433 sentences with uneven category membership because it has fewer credit sentences, while the human-verified test sample contains 300 sentences with equal proportions.

We implement the model update in Stage 2 by running backpropagation on the Stage 1 model parameters using the same adaptive learning-rate algorithm with randomly ordered mini-batches of 32 sentences. However, unlike in Stage 1, we apply an inverse weighting scheme to ensure that the credit category, which has fewer sentences, receives a proportionally higher weight. We use a learning rate that is one-tenth smaller to prevent the model from forgetting what it learned in Stage 1. At the same time, we run backpropagation over many more epochs to enable sufficient learning from the human-verified sentences. We pick the number of epochs to maximize the out-of-sample F1 score, which balances precision and recall. We evaluate this score in the test set of human-verified sentences. Online Appendix Table A.2 reports the out-of-sample F1 score of the resulting Stage 2 model over different numbers of epochs. We find that 100 epochs deliver the highest out-of-sample F1 score for the Stage 2 model. We fix this number of epochs for Stage 2 going forward.[7]

[7]Ideally, we would select the number of epochs on a separate set of human-verified sentences. Instead, we reuse the test set for both epoch selection and final evaluation because of the large costs associated with human verification. Reassuringly, the test F1 score stays within about two percentage points of its peak for any epoch count between 75 and 250. The results are therefore not overly sensitive to this choice.

Table 5 reports the out-of-sample precision and recall rates, the out-of-sample R^2, and the confusion matrix for the fine-tuned Stage 2 model. We observe that the few-shot update with human-verified sentences improves accuracy among human-verified news sentences. The overall F1 score of the model increases from 79 to 84%, and the out-of-sample R^2 increases from 59.7 to 73.6%. The improvement arises from greater precision when labeling credit sentences as well as greater recall when labeling peer, production, and other sentences. The out-of-sample R^2 for credit sentences becomes positive and large after the Stage 2 update. The update also produces nearly balanced recall and precision rates across categories. The model therefore does not systematically overpredict or underpredict any link type.

Despite the improvements in accuracy, the Stage 2 model still exhibits the lowest precision when labeling credit links, aligning with the relative scarcity of credit sentences in the NOW sample (see Section 5.1). This lower precision implies that the credit network is measured with more classification noise than the other networks. Such noise generally should attenuate rather than inflate predictive associations. The forecasting power we document for the credit network in Section 6 should therefore be read as a conservative estimate of its true predictive content. Improving credit-link precision, for instance by supplementing our sample with specialized credit-market trade publications, is a natural direction for future work.

4.3 Stage 3: Rounding parameters

In the final Stage 3, we pick the parameters p^+ and p^- governing the rounding in Eq. (2). We follow a heuristic-based approach to specify their values to ensure that the resulting networks are economically meaningful. To specify the round-down threshold p^-, we consider the probability of correctly guessing a link label at random. Given that there are six link categories, a random guess would be correct with a probability of one-sixth. Thus, we set $p^- = 1/6$ and round down to zero whenever the predicted link probability falls below this threshold.

To specify the round-up threshold p^+, we consider the impact this threshold has on the predicted link indicators. Any link category with predicted probability above p^+ gets rounded up to one. Thus, the argmax (i.e., the category with the highest predicted probability) for sentences with such

link probabilities naturally coincides with any category that is rounded up (admitting that more than one category may be rounded up per sentence if p^+ is small). Given this logic and given that the F1 score balances recall and precision when predicting link indicators, we calibrate p^+ by choosing the round-up threshold that maximizes the out-of-sample F1 score of our model on the human-verified test sentences while keeping $p^- = 1/6$ fixed. Figure A.2 in the Online Appendix displays the implied out-of-sample F1 scores for different values of the round-up threshold p^+. While the F1 differences across thresholds are modest given the small test sample, the peak at $p^+ = 0.494$ is present not just in the human-verified sentences analyzed here, but also in an unreported hybrid test set that also includes an equal amount of synthetic sentences picked at random and balanced across link categories. Based on these results, we fix $p^+ = 0.494$.

4.4 Extensions

Our approach can easily be extended to identify additional link types, such as analyst coverage. The third example in Table 3 shows how the "other" category can include such links. Adding a category requires generating new synthetic and human-verified training sentences and re-running Stages 1 through 3 of Section 4. Online Appendix F details the procedure.

Our model can also be extended to deal with instances in which Assumption (A2) is violated and the news sample contains many sentences that describe multiple firm link types simultaneously. In those cases, it is recommended to use independent binary classifiers for each link type rather than the single multi-class classifier of Figure 2. The resulting link probabilities may then no longer add up to one (i.e., $\sum_c p_{i,t,c}(f,g) > 1$ is possible), but the proof of Theorem 3.1 can be extended to handle this case with more complex notation.

5 Output of the methodology

We analyze the output of our methodology. We consider both the content of the sentences in which links are recognized and the implied labeled networks, in aggregate over the full sample and dynamically over time.

5.1 Sample sentences

We begin by reporting representative links and sample sentences in which they are identified. The sample sentences are independently verifiable against well-documented corporate history. We proceed as follows. We compute link strengths over the full sample as outlined in Eq. (3). We then take all firm pairs that are linked at least 100 times over the sample, and sort them by the link strength. Tables 6 and 7 show the three strongest links for each category. The tables also report the number of sentences in which the link strength is exactly equal to one (i.e., the link probability assigned by our model is above the round-up threshold p^+).

The tables document that our model adequately identifies the different types of firm links. It correctly identifies direct credit exposures (Delta's loan to PanAm) and M&A-related liabilities (NLT's delayed payment for American General shares, Texas Air assuming People Express's debts). In unreported analyses, credit also captures indirect credit exposures and alternative credit relationships such as asset leasing. For production links, the model captures customer-supplier relationships (Nissan and Chrysler) and joint production ventures (Apple and Motorola, Hitachi and Nippon Electric). Some of these sentences reference more than the two firms that the firm identification methodology of Schwenkler and Zheng (2026) recognizes (e.g., the Apple-Motorola sentence also mentions IBM). This shows that we do not discard such sentences even though we may only assign links between two firms in those cases.

Peer links clearly capture competitive relationships, while equity links pick up direct M&A investments. Leadership links capture overlapping shareholders (Jos. A. Bank and Men's Wearhouse), close relationships among employees, directors, or shareholders (Merrill Lynch and Enron), one company assigning directors or managers to another (Chris-Craft and Warner, Columbia and Sony), and, in unreported sentences, employees who worked at different companies over time. Notably, our model allows two firms to share multiple relationship types: for Chris-Craft and Warner Communications, it assigns a 56.7% likelihood of an equity link and 30.2% of a leadership link. The representative sentences for "other" links show that this category may subsume relationships that cannot be cleanly identified from the text. All of the strongest other links are between well-known peers. Yet, unlike the peer link sentences, none use explicit competitive language.

Tables 6 and 7 also document that link types differ in abundance. Judging by the number of strong sentences (those scoring above our round-up threshold p^+), production links are most abundant and credit links are sparsest, while many sentences are classified as "other." Online Appendix Figure A.4 shows the distribution of news sources among these strong sentences by link category. The New York Times dominates, with as many as 91% of strong credit sentences coming from this sample alone. This confirms that the difficulty we faced in Stage 2 finding false-positive credit sentences outside the New York Times reflects a structural feature of the data: credit sentences rarely appear in the NOW Corpus.

5.2 Keyword analyses

We complement Tables 6 and 7 by running keyword analyses on the sample of strong sentences. Figure 3 displays the most frequently mentioned nouns in these sentences across link categories, and Online Appendix Figure A.5 displays the analogous most frequently mentioned verbs. We ignore the words *be*, *company*, *include*, *month*, *percent*, *say*, *share*, *week*, and *year* because they are frequently mentioned across link categories without informing about their differences.

The keyword evidence further confirms that each network captures information related to the link type it represents. For example, nouns like *credit*, *debt*, and *loan*, as well as verbs like *repay* and *yield*, are among the most frequently mentioned keywords in sentences in which we identify credit relationships. In contrast, words like *acquire*, *merger*, *shareholder*, and *stake* are among the most frequent keywords for equity links, highlighting the different nature of these types of financing links. Sentences in which we identify peer links frequently mention the keywords *compete*, *industry*, *market*, and *rival*, while production link sentences contain the keywords *customer*, *product*, and *service*. Reflecting their more person-oriented nature, sentences in which we identify leadership links often mention position keywords like *director*, *executive*, and *president*, or employment keywords like *hire*, *serve*, and *work*. Sentences in which we identify other links tend to mention keywords related to performance reporting like *earning*, *fall*, and *rise*; analyst recommendations like *analyst*, *estimate*, *recommend*, and *target*; and event reporting like *case*, *court*, and *lawsuit*. Together with the sample sentences, these keywords show that other links capture structural relationships between

firms that our methodology does not assign to one of the five economic categories.

5.3 Labeled news-implied networks

We construct labeled firm networks, aggregated over the full time series, using the methodology of Section 3.1. We plot networks for each category, including only firm links that were identified at least 100 times in our sample and for which the resulting link strength is not zero. Figure 4 shows the six resulting networks.

Compared to the unlabeled network in Online Appendix Figure A.3 that is based purely on sentence co-mentions, the labeled networks of Figure 4 are interpretable and economically meaningful. For example, the breakdown by link category explains why Goldman Sachs connects with Tesla and Apple. Despite the fact that Goldman Sachs is a bank, these links are not classified as credit by our methodology. The model classifies the Goldman-Apple connection primarily as a production link and the Goldman-Tesla connection primarily as an "other" link. When reviewing the sentences in which Apple and Goldman are co-mentioned, we observe that our methodology most strongly identifies a production link because Goldman partnered with Apple in 2019 to offer a credit card, with Goldman providing the financial back-end and Apple providing the customer-facing front-end. While this may not qualify as a manufacturing production link, it involves a customer (Apple) using the services of a supplier (Goldman). Our model is trained to classify such relationships as production. In contrast, our methodology most strongly identified the other type of link between Goldman and Tesla because Goldman analysts often provide forecasts and ratings for the Tesla stock.

Looking at the general architecture of the different networks, we observe that the credit network has a core-periphery structure, with the banking sector being highly central and interconnected, and industry clusters surrounding the core. A core-periphery structure arises naturally for credit networks; see Craig and von Peter (2014), Farboodi (2023), and others. The equity and peer networks tend to have stronger bilateral linkages, consistent with the intuition that these networks capture pairwise transactions (equity) or comparisons (peer). The production network has a strong star architecture, consistent with the theory of Acemoglu et al. (2012). It showcases a strong core

of large technology firms that connect to different industries through long chains of intermediation. The leadership and other networks appear to be most dispersed, though the other network reflects a strong cluster around the big banks.

Figure 5 shows time series of the credit and production networks over subsequent 15-year sub-samples. Time series graphs for the remaining networks are available in Online Appendix G. The graphs show that our methodology captures changes in news-implied firm links over time. For example, the evolution of the credit network shows a spike in intensity in the banking cluster in the intermediate sub-sample. This coincides with the global financial crisis of 2008 and 2009 during which the news more strongly reported about credit relationships between banks because they were distressed. This does not mean that credit links between banks were only weak before or after the financial crisis. It is just that our methodology picks them up more strongly during the financial crisis because the news increased its reporting intensity about these types of links. In contrast, the evolution of the production network documents greater interconnection over time. The production network for the earlier period shows many disjoint production clusters, while the graphs for the latter periods show more connections across clusters. The evolution of the production network aligns with Acemoglu and Azar (2020), Carvalho and Voigtländer (2014), and others, who document that the US production network has become more interconnected over time.

As these observations indicate, changes in the links estimated by our methodology over time can arise due to actual changes in the structure of these links (as when the production network becomes more interconnected) or due to changes in how the news reports about the links (as when the banking cluster became more prominent in the credit network during the financial crisis). We exploit both of these channels in our predictive analyses in Section 6.

5.4 Comparison to existing networks

We compare our labeled networks to five benchmark networks assembled from existing databases. These include an equity network from WRDS company subsidiary data (1996–2018), a leadership network from common executive and director positions in BoardEx (from 2000), a peer network from the Text-based Network Industry Classifications (TNIC) of Hoberg and Phillips (2016) (1988–2021),

a production network from Compustat customer segments (from 1997-Q4), and an inter-sectoral production network from BEA industry-by-industry total requirement tables (1997–2021).[8] Online Appendix H details how we construct and match each benchmark network and how we standardize the universe of firm links for comparison. In all cases, we restrict ourselves to the 2,000 largest firms by average market cap over the relevant sample period. The Online Appendix shows that a smaller or larger firm sample yields analogous results.

We run two exercises to compare our networks to the existing alternatives. We first compute the cosine similarity between our link strengths and those in an alternative network. Cosine similarity measures how aligned two vectors of link strengths are and is commonly used to compare networks; see Craig et al. (2023) and others. In a second analysis, we run regressions of the link strength in an alternative network on the link strengths in our six labeled networks, using quasibinomial regressions with a logit link when the alternative strength is bounded between zero and one (TNIC, BoardEx, Subsidiary) and Gamma regressions with a log link when it is non-negative and effectively unbounded (Compustat, BEA). When running firm-level regressions, we include the log market capitalization of the linked firms as regressors to capture unobserved heterogeneity in how firms are covered in the existing databases. We do not include this control in the sector-level regressions for the BEA data because the network is complete. We compute standard errors via a cluster bootstrap at the firm or sector level, whichever is appropriate.[9]

Table 8 presents our results, where we mark in green (yellow) the statistically strongest (second strongest) associations. The cosine similarities show that links in the alternative networks correlate strongly with our matching network. The equity network from WRDS subsidiary data and the production network from Compustat customer segments both achieve their highest cosine similarity with our corresponding network. For the other alternative networks, our corresponding networks come in second or third. These similarities are nonetheless generally low, reflecting sparse overlap and many zero entries. For example, our equity network and the WRDS subsidiary network combine

[8] We are unable to construct benchmarks for the credit and other networks. Credit link data are practically unavailable outside of central banks (Craig et al. (2023)), and the other network is a residual category combining a range of potential alternative links, so neither admits a clean benchmark.

[9] Standard clustering requires a fixed assignment of each observation to a cluster, but a firm (or sector) can appear on either side of a link, so we resort to bootstrap instead; see the caption of Table 8 for details.

for 8,255 active links, but WRDS subsidiary data report only 168 of them. Conversely, TNIC's peer network reports 34,004 active links against our 1,578. This makes cosine similarity a noisy measure of the true correlation.

The link regressions offer sharper evidence by controlling for coverage variation through the log market cap regressor, clustered standard errors, and the covariance structure across our six networks. The alternative networks are generally positively and significantly associated with their matching network, with the strongest and most significant associations occurring for the WRDS equity, BoardEx leadership, and TNIC peer networks. Two associations are weaker at first pass but resolve once we address a specific data limitation. The BEA production association is initially weak because of within-sector aggregation but becomes significant once we exclude self-links. The association between Compustat segments and our production network is positive but insignificant, reflecting sparse Compustat coverage (only 462 of more than 6,700 combined links). Online Appendix G shows that the diagonal associations often become stronger and more significant once we expand the firm universe beyond the top 2,000 firms.

Several off-diagonal associations across network types are also statistically significant. For example, the WRDS equity and TNIC peer networks both correlate with our credit network. We show in Online Appendix G that these kinds of off-diagonal associations tend to be spurious, driven by low-frequency links. Their significance weakens and often disappears once we exclude links identified fewer than ten times in our news sample. However, the strong association between the TNIC peer network and our other network remains. This finding confirms our observation from Section 5.1 that the other category captures implicit competitive relationships not explicitly communicated in the news. To avoid confounding and because the other category is ambiguous, we exclude it from our empirical analysis in Section 6.

All in all, the results of this section demonstrate that our labeled networks genuinely capture the types of firm links they target, combining external validation from conventional databases with internal validation from news-sentence contents. While some of our networks could also be constructed from existing databases, we offer two advantages. We are not tied to a predetermined update schedule. Users can choose the aggregation window and recompute the networks as news

arrives, although precision depends on the volume of relevant coverage within that window. TNIC, Compustat segments, and BEA data instead update infrequently or have historical cutoff dates. Our networks also cover a much larger universe of firms than the comparatively sparse Compustat segments and WRDS subsidiary data. These advantages make our network data useful for empirical applications beyond the one in this paper. For convenience and to enable replication of our aggregate forecasting exercises, the online repository reports networks updated monthly.

6 Forecasting aggregate outcomes

We ask whether the aggregate architecture of firm relationships predicts subsequent financial and real conditions. The analysis is deliberately predictive, addressing practical questions that investors, regulators, and policy-makers may face. We measure what the networks reveal before an outcome is realized, but we do not measure the causal effect of changing a network.

6.1 Empirical framework

We consider monthly outcomes from WRDS and FRED: the three-month, two-year, and ten-year Treasury rates; the CRSP value-weighted market log return; the S&P 500 log return; the VIX; and log growth in real personal consumption expenditures and industrial production. We use the 3-month Treasury rate as a proxy for the level of the yield curve and the spread between the ten and two-year Treasury rates as a proxy for its slope. These variables represent bond-market, equity-market, and real conditions. Table 9 reports summary statistics.

For each labeled network, we compute summary statistics of the network architecture that capture changes in the degree of firm connectivity. We follow Acemoglu et al. (2012) and measure:

- The density, which is the average number of active (i.e., non-zero) links of a node in the network. The density captures how many links are present in a network.

- The concentration, which is given by the coefficient of variation of the node degree distribution. Here, the degree of a node is the sum of all its link strengths. The concentration of a

network is large whenever a few nodes have many strong links, while many nodes have only a few weak links.

- The intermediation, given by the weighted degree covariation of two firms that are indirectly connected through a third firm. It is large whenever many firms act as intermediaries connecting two strongly connected nodes.

The metrics closely relate to alternative network metrics used by Herskovic (2018).[10] More broadly, Schwenkler and Zheng (2026) provide an overview of how density, concentration, and intermediation relate to other network topology metrics.

Figure 6 shows the time series of the connectivity metrics for the different networks, also showcasing the exact formulas for how the metrics are computed. For this application, we update the measures monthly. At each update, we collect all news links identified in the previous 12 months, excluding the current month to avoid forward-looking bias, and compute news-implied networks using the methodology of Section 3.1. Following Section 5.4, we restrict ourselves to the 2,000 largest firms by average market cap to ensure a stable firm universe over time, and only consider links that we identify at least 10 times over that sample period to exclude noisy low-frequency observations. Figure 6 shows that the credit network has the highest concentration, aligning with the strong core-periphery structure documented in Section 5.3. On average across the sample, the peer network is densest while the production network has the highest intermediation intensity.

We run regressions of the cumulative change in the aggregate variables over future periods on our network metrics constructed only using contemporaneously available data. For Month t and Horizon h, we run regressions of the type:

$$Y_{t+h-1} = \alpha_{n,h} + \gamma_{n,h}^{\top} X_{t-1} + \beta_{D,n,h} D_{n,t} + \beta_{C,n,h} C_{n,t} + \beta_{I,n,h} I_{n,t} + \epsilon_{n,t,h}. \tag{4}$$

Here, $D_{n,t}$, $C_{n,t}$, and $I_{n,t}$ are the density, concentration, and intermediation metrics for Network n measured at the start of Month t. The outcome variable Y_t and controls X_t are measured at the end of Month t. The controls include the lagged monthly article count and AIC-selected lags of the

[10] On average across network types, we find that the correlation between our density and the sparsity measure of Herskovic (2018) is -61.3%, while the correlation between our and his concentration metrics is 97.8%.

outcome. Return regressions also include lagged VIX, and VIX regressions include the lagged S&P 500 return to account for leverage-effect dynamics (Black (1976)). We standardize the outcome and all regressors, so each regression coefficient is the conditional change in the future monthly outcome, in standard deviations, associated with a one-standard-deviation change in a connectivity metric.

Boudoukh et al. (2008) and Valkanov (2003) show that the persistence induced by rolling-window construction can distort conventional predictive-regression inference. We therefore compute p-values through a bootstrap procedure. We fit a separate AR(1) model to each connectivity metric and generate 1,000 placebo paths that preserve its marginal persistence and innovation variance but are unrelated to the outcomes. For every draw, we re-estimate Eq. (4) with the same controls and Newey–West covariance estimator, using lag $\max\{12, h-1\}$. Two-sided empirical p-values compare the realized coefficient with its placebo distribution. Because we do not know ex-ante what to expect and to address multiple-hypothesis testing concerns raised by Harvey et al. (2016), we apply p-value adjustments using the false-discovery-rate procedure of Benjamini and Hochberg (1995). The dimension along which we apply these adjustments depends on the application.

Note that our classification procedure uses language models whose training corpora may include information from later periods (Kong et al. (2026)), a potential source of look-ahead bias. This exposure is limited, however. Stage 1 trains only on sentences about fictional firms. As a result, the classifier only learns the generic language of a relationship type by construction. Stage 2 fine-tunes with published news sentences but using a ten-times-smaller learning rate. Section 4.2 shows that the resulting model's accuracy on synthetic sentences remains high, consistent with Stage 2 sharpening decision boundaries rather than displacing what Stage 1 learned. The networks themselves are built exclusively from contemporaneously published news. The remaining concern is only whether the generic language of relationship-description has itself shifted over time, a second-order risk relative to firm-specific look-ahead.

6.2 Predictive regression coefficients

Figures 7 and 8 display the estimates of the regression coefficients $\beta_{D,n,h}$, $\beta_{C,n,h}$, and $\beta_{I,n,h}$ in (4), together with symbols that mark their statistical significance. We keep the y-axis range fixed across

graphs to enable easy visual comparison. We seek to identify the connectivity metric that most strongly predicts an outcome at a given horizon. As a result, we apply p-value adjustments for the three connectivity metrics per network and prediction horizon. In Online Appendix I, we document similar results if we adjust the p-values by prediction horizon rather than metric, answering the different question whether a given metric is more strongly predictive over one horizon than another given the possibility of false positives.

The regression coefficients reveal three main patterns. First, predictive content is concentrated in economically related outcomes. Credit connectivity is most informative about the yield curve, equity connectivity about stock returns and volatility, and production connectivity about consumption and industrial production. Leadership and peer networks also predict both equity market and real outcomes. Cross-domain associations exist, but many weaken when the false-discovery-rate adjustment is instead applied across horizons (Online Appendix I). This domain alignment is an external validation of the labels because the outcome data play no role in constructing the networks.

Second, density and concentration precede weaker conditions. Greater density in the equity, leadership, peer, and production networks predicts lower subsequent market returns and real growth. For example, a one-standard-deviation increase in leadership density is associated with a monthly market return roughly 0.2 standard deviations lower at many horizons through 18 months. Production density predicts monthly consumption growth about 0.1 standard deviations lower at horizons through one year. Concentration has similar implications for real activity. Production concentration predicts consumption growth about 0.1 standard deviations lower and industrial-production growth about 0.2 standard deviations lower beginning around five months ahead. These results align with Acemoglu et al. (2012), Herskovic (2018), Herskovic et al. (2020), and Schwenkler and Zheng (2026), who show that shocks to concentration within production networks can predict lower output and higher volatilities. By measuring several networks comparably, we establish that the negative forecasts associated with density and concentration extend beyond production links. Leadership and peer concentration also predict lower returns and higher VIX, with leadership coefficients for VIX reaching roughly 0.6 standard deviations.

Third, intermediation often has the opposite sign. Credit intermediation predicts a steeper

yield curve, with the slope coefficient reaching approximately 0.3 standard deviations around 18 months. Across the non-credit networks, greater intermediation generally predicts stronger subsequent real activity. Equity intermediation predicts consumption and industrial-production growth up to roughly 0.2 standard deviations higher around six months ahead. The corresponding leadership, peer, and production signals tend to peak beyond 18 months. The contrast with density and concentration provides empirical evidence that additional direct links, greater centralization, and stronger bridges represent different economic states.

Many of the strongest associations emerge over horizons of several months rather than immediately. The timing suggests that news-implied networks may capture persistent changes in economic organization and in the salience of firm relationships, rather than transitory news shocks. This interpretation aligns with Beaudry and Portier (2006), who show that news about future fundamentals can precede its realized aggregate effects. Section 6.4 provides supporting evidence by showing that several news-implied measures lead changes in lower-frequency databases. Our bootstrap inference preserves the persistence of each predictor and uses Newey–West adjustments for overlapping outcomes, which reduces the concern that long-horizon significance arises mechanically.

6.3 Predicting in real-time

The previous section establishes in-sample predictability. We now ask whether investors and forecasters can use these relationships as information arrives. We evaluate market-timing strategies and real-time directional forecasts of aggregate outcomes.

6.3.1 Trading

We build a market-timing strategy in the spirit of Breen et al. (1989) and Campbell and Thompson (2008). The strategy invests in three-month Treasury bills unless a connectivity metric forecasts higher S&P 500 returns. Guided by Figures 7 and 8, we enter the index when density or concentration is below its median over the previous 60 months, or when intermediation is above its trailing median. Positions remain open for h months, where h ranges from one to 60, and only a fraction $1/h$ of the portfolio turns over each month. The strategy therefore maintains unit leverage.

Because the threshold uses only past observations and the networks use news published before the trading month, every signal can be constructed in real time.[11]

Persistence and overlapping positions complicate conventional inference. For each metric, we therefore fit an AR(1) model, generate 1,000 unrelated sample paths, and rebuild the portfolios. We compare the realized Sharpe Ratios and CAPM alphas with their placebo distributions at each holding horizon, then adjust the p-values across networks for a fixed metric and horizon. Table 10 reports performance at the holding period with the highest Sharpe Ratio. Figure A.10 in the Online Appendix plots the associated cumulative returns and the reverse-signal strategies. The online data repository provides the complete strategy results and underlying connectivity metrics.

Only strategies based on density achieve statistically significant Sharpe Ratios. Entering the market when the credit, equity, and production networks are less dense also produces statistically positive CAPM alphas. The reverse-signal strategies in Figure A.10 perform considerably worse. The preferred holding period is one month for credit density and more than 30 months for the other density signals. Annualized alphas reach 3.3%, and the moderate market betas document active timing rather than passive exposure. These results complement Acemoglu et al. (2015), Miranda-Pinto et al. (2023), and others, who show that additional links can amplify adverse shocks. We show that lower (higher) density forecasts higher (lower) stock returns in real time.

Intermediation and concentration do not produce trading performance that survives the multiple-testing adjustment. Intermediation appears significant under standard Newey-West inference, but the significance disappears after adjusting across networks at each holding horizon. Across the non-credit networks, intermediation strategies nevertheless generate strong returns in Figure A.10 and outperform their reverse-signal counterparts. Intermediation therefore provides a positive but statistically weaker real-time signal. Concentration is the weakest real-time predictor. Its strategies rarely achieve significant performance, and their reverse-signal counterparts often perform better.

[11] We use a long window to estimate the median because the connectivity metrics are highly autocorrelated.

6.3.2 Directional predictions

We next apply the same real-time signals to outcomes that cannot be traded directly in conventional financial markets. Our starting point is that low density and concentration predict an increase in the outcome, while high intermediation predicts an increase. We test whether these signals predict changes in Treasury yields or the sign of cumulative consumption and industrial-production growth over horizons from one to 60 months. These tests are relevant not only for professional forecasters, but also for emerging macroeconomic prediction markets (Diercks et al. (2026)).

Because cumulative growth is positive in most multi-month windows, raw accuracy can exceed 50% without predictability. We therefore report excess directional accuracy relative to the accuracy implied by the unconditional frequencies of predicted and realized increases. We initially calculate signed excess accuracy. A negative value identifies an informative reverse signal. For comparability, Table 11 orients every reported signal in its predictive direction, so all displayed excess accuracies are positive. In particular, the ten-year Treasury result reports high density because it reverses the original low-density signal. We reuse the AR(1) placebo design from the trading exercise and evaluate every placebo against the realized outcome over the same months as the news-implied signal. Two-sided empirical p-values compare absolute excess accuracy with its placebo distribution. We again adjust across the five networks for each metric and horizon, now separately for each outcome. The online data repository reports all 4,500 results in a spreadsheet.

Table 11 summarizes the economically meaningful patterns. Density provides the broadest directional information. Lower density forecasts positive cumulative consumption growth across all five networks, reinforcing the evidence in Section 6.2 that denser networks precede weaker real conditions. Density also predicts changes in the ten-year Treasury yield across all five networks. Intermediation produces narrower but economically meaningful forecasts of consumption growth in the production and peer networks. Their positive sign reinforces the earlier finding that networks with stronger bridges precede stronger real activity. The directional tests align with the results of Section 6.2 without using outcome magnitudes or a linear forecasting specification in Eq. (4). As in the trading exercise, concentration produces isolated results but no stable directional pattern.

Across our tests, density consistently predicts adverse outcomes, intermediation predicts favor-

able outcomes, and concentration provides more limited predictive information. These findings establish empirically that aggregate connectivity is multidimensional. A denser network, a more hub-concentrated network, and a network with more bridges are not equivalent economic states. Their distinct forecasting patterns align with the simultaneous operation of the beneficial and adverse channels in Table 1 across different architectures and relationships.

6.4 Why is there predictability?

We ask why measuring economic relationships through news provides predictive information for aggregate outcomes. News-implied networks potentially combine two observable sources of variation. First, their topology changes as the underlying economic relationships among firms change (Liu and Tsyvinski (2024), Schwenkler and Zheng (2026)). Second, selectivity in what is covered in the news may itself be informative about future outcomes (Bybee et al. (2024), Chahrour et al. (2021), Schwenkler and Zheng (2025*b*)). In addition, news-implied network may capture information that is absent from existing databases. We use independent network measures and news-coverage intensity to separate variation associated with observable sources from residual variation specific to our news-implied measures.

For each Network n and Metric m, we estimate

$$M_{n,m,t} = a_{n,m} + b_{n,m,A}^{\top} A_{n,t} + b_{n,m,N}^{\top} N_t + e_{n,m,t}, \tag{5}$$

where $A_{n,t}$ contains density, concentration, and intermediation from the matched network in Section 5.4. The vector N_t contains the log number of articles and the numbers of firm mentions and links per article. All three objects use the same 12-month information window ending in $t-1$. We estimate Eq. (5) over the sample available for each conventional database and test $b_{n,m,A} = 0$ and $b_{n,m,N} = 0$ with HAC-robust Wald tests. We apply p-value adjustments across networks within metric category because we seek to answer whether the observable channels consistently explain variation for a given metric. Credit is excluded because it has no matched database.

Table 12 shows high adjusted R^2 values for density and concentration across networks and moderate values for intermediation. The observable sources therefore span substantial variation

in the connectivity metrics, while leaving meaningful residual variation. News-implied connectivity generally loads significantly on coverage features, as expected for measures constructed from news reporting. Three out of the four networks load significantly on the connectivity metrics of the existing databases, though not consistently across connectivity metrics. Leadership does not load significantly on conventional topology. We do not interpret this to mean that the news-implied leadership network does not contain information about true leadership connectivity for two reasons. First, we only observe the start and end dates of a directorship but not the date when this information became public in the BoardEx data, which may weaken its contemporaneous alignment with the news-implied leadership network. Second, Figure 10 establishes that news-implied leadership intermediation leads its conventional counterpart at horizons of around half a year, documenting a leading rather than a contemporaneous relationship. All in all, Table 12 establishes that the news-implied connectivity measures combine variation represented in conventional databases, variation associated with coverage choices, and orthogonal variation not captured by either source.

We next ask whether the news-implied metrics add predictive information after controlling for both sources. We estimate

$$Y_{t+h-1} = a_{n,h} + G_{n,h}^{\top} X_{t-1} + B_{A,n,h}^{\top} A_{n,t} + B_{N,n,h}^{\top} N_t + B_{M,n,h}^{\top} M_{n,t} + E_{n,t,h}. \tag{6}$$

We evaluate each component of the coefficient $B_{M,n,h}$ using the same AR(1) placebo procedure as before, replacing the news-implied metrics while holding the realized outcomes, controls, conventional metrics, p-value adjustments, and coverage features fixed.

Figure 9 shows that the controls absorb some, but not all, baseline predictability. Production concentration and intermediation retain information about later real activity, while leadership and peer concentration retain information about future VIX. Leadership density also predicts the yield-curve level and slope in these restricted samples. These coefficients should not be compared mechanically with the full-sample estimates because each matched database covers a different subperiod. The relevant conclusion is narrower: conventional topology and observed coverage intensity do not span all of the aggregate predictive information in the news-implied measures.

We finally test whether the faster updating of continuously published news contributes to the

residual predictive power relative to conventional network databases. The subsidiary, BoardEx, and TNIC measures are based largely on annual filings, and customer segment data are quarterly. News can report relationship changes between filing dates and therefore enter our networks earlier. We test this timing channel metric by metric. First, we residualize each news-implied metric with respect to its matched contemporaneous metric from the existing network. We then regress the matched metric of the existing network h months ahead on its current value and the news-implied residual. Significance is assessed with a 1,000-draw persistence-matched placebo bootstrap and false-discovery-rate adjustment across the three metrics at each horizon. A significant coefficient means that contemporaneously distinct news-implied variation leads the matched database.

Figure 10 finds selective rather than universal lead-lag evidence. Equity intermediation leads its subsidiary counterpart at horizons up to roughly one year. Leadership intermediation weakly leads BoardEx around six months, and peer concentration begins to lead TNIC around 18 months. No production metric leads the quarterly customer segment measure. The absence of a production lead aligns with the smaller timing advantage available against the quarterly update frequency of the Compustat database. These results reflect only predictive associations, not causal relationships.

Taken together, the exercises of this section establish that the news-implied metrics reflect conventional topology and coverage choices but are not exhausted by either. Some residual variation forecasts aggregate outcomes, and some precedes changes in lower-frequency counterparts. Financial news therefore supplies both a common measurement framework and timely information about firm-network architecture.

7 Conclusion

We show that different firm relationships contain distinct information about future aggregate conditions, and that this information depends on both relationship type and network architecture. Across five economically labeled networks extracted from financial news, greater network density and concentration generally forecast adverse outcomes while greater intermediation predicts beneficial results. The relationship represented by a network often determines which outcomes its architecture predicts most strongly. We develop market-timing strategies that profitably exploit

this predictability. Directional forecasts extend the evidence to consumption growth and long-term Treasury yields. Our findings are predictive rather than causal. Still, they advance empirical network analysis by establishing that different dimensions of network architecture carry distinct predictive signals for aggregate outcomes.

To understand the origins of the predictability, we show that news-implied measures contain information about underlying firm relationships and selective news coverage, yet neither source exhausts their predictive power. Several signals continue to forecast aggregate outcomes after controlling for conventional networks and coverage intensity, and some precede changes in lower-frequency databases. Financial news therefore provides a common measurement framework and timely information about economic connectivity.

Establishing these results requires comparable measures of economy-wide firm connectivity. We develop a novel AI-based methodology that extracts credit, equity, leadership, peer, production, and other links from more than four decades of financial news. It combines scalable synthetic training data with targeted human verification, preserves text-level traceability, and measures these relationships over a common firm universe and information set. We validate the accuracy of our networks both theoretically and empirically through direct sentence validation and comparisons with existing networks. To enable applications beyond those studied in this paper, we provide monthly time series of our network data and the accompanying code in our online repository.

References

Acemoglu, Daron, Asuman Ozdaglar and Alireza Tahbaz-Salehi (2015), 'Systemic risk and stability in financial networks', *American Economic Review* **105**(2), 564–608.

Acemoglu, Daron and Pablo D. Azar (2020), 'Endogenous production networks', *Econometrica* **88**(1), 33–82.

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar and Alireza Tahbaz-Salehi (2012), 'The network origins of aggregate fluctuations', *Econometrica* **80**(5), 1977–2016.

Arora, Navneet, Priyank Gandhi and Francis Longstaff (2012), 'Counterparty credit risk and the credit default swap market', *Journal of Financial Economics* **103**(2), 280–293.

Azar, José, Martin C. Schmalz and Isabel Tecu (2018), 'Anticompetitive effects of common ownership', *The Journal of Finance* **73**(4), 1513–1565.

Barrot, Jean-Noël and Julien Sauvagnat (2016), ' Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks', *The Quarterly Journal of Economics* **131**(3), 1543–1592.

Beaudry, Paul and Franck Portier (2006), 'Stock prices, news, and economic fluctuations', *American Economic Review* **96**(4), 1293–1307.

Benjamini, Yoav and Yosef Hochberg (1995), 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(1), 289–300.

Bernard, Andrew B., Andreas Moxnes and Yukiko U. Saito (2019), 'Production networks, geography, and firm performance', *Journal of Political Economy* **127**(2), 639–688.

Black, Fisher (1976), Studies of stock price volatility changes, *in* 'Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Section, Chicago, IL', pp. 177–1181.

Boudoukh, Jacob, Matthew Richardson and Robert F. Whitelaw (2008), 'The myth of long-horizon predictability', *The Review of Financial Studies* **21**(4), 1577–1605.

Breen, William, Lawrence R. Glosten and Ravi Jagannathan (1989), 'Economic significance of predictable variations in stock index returns', *The Journal of Finance* **44**(5), 1177–1189.

Bybee, Leland, Bryan T. Kelly, Asaf Manela and Dacheng Xiu (2024), 'Business news and business cycles', *The Journal of Finance* **79**(5), 3105–3147.

Cai, Jing and Adam Szeidl (2018), 'Interfirm relationships and business performance', *The Quarterly Journal of Economics* **133**(3), 1229–1282.

Cameron, A. Colin, Jonah B. Gelbach and Douglas L. Miller (2008), 'Bootstrap-based improvements for inference with clustered errors', *The Review of Economics and Statistics* **90**(3), 414–427.

Campbell, John Y. and Samuel B. Thompson (2008), 'Predicting excess stock returns out of sample: Can anything beat the historical average?', *The Review of Financial Studies* **21**(4), 1509–1531.

Carvalho, Vasco M. and Nico Voigtländer (2014), Input diffusion and the evolution of production networks. NBER Working Paper 20025.

Chahrour, Ryan, Kristoffer Nimark and Stefan Pitschner (2021), 'Sectoral media focus and aggregate fluctuations', *American Economic Review* **111**(12).

Chen, Zilin, Li Guo and Jun Tu (2021), 'Media connection and return comovement', *Journal of Economic Dynamics and Control* **130**, 104191.

Chiu, Peng-Chia, Siew Hong Teoh and Feng Tian (2013), 'Board interlocks and earnings management contagion', *The Accounting Review* **88**(3), 915–944.

Cohen, Lauren and Andrea Frazzini (2008), 'Economic links and predictable returns', *The Journal of Finance* **63**(4), 1977–2011.

Cong, Lin William, Tengyuan Liang, Xiao Zhang and Wu Zhu (2025), 'Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information', *Management Science* **71**(12), 10727–10739.

Craig, Ben and Goetz von Peter (2014), 'Interbank tiering and money center banks', *Journal of Financial Intermediation* **23**(3), 322–347.

Craig, Ben R., Madina Karamysheva and Dilyara Salakhova (2023), Do market-based networks reflect true exposures between banks? ECB Working Paper No. 2023/2867.

Diebold, Francis X. and Kamil Yılmaz (2014), 'On the network topology of variance decompositions: Measuring the connectedness of financial firms', *Journal of Econometrics* **182**(1), 119 – 134.

Diercks, Anthony M., Jared Dean Katz and Jonathan H. Wright (2026), Kalshi and the rise of macro markets, Finance and Economics Discussion Series 2026-010, Board of Governors of the Federal Reserve System.
URL: *https://doi.org/10.17016/FEDS.2026.010*

Duan, Junting and Markus Pelger (2026), Inference with ai-generated covariates. Working Paper.

Duffie, Darrell, Andreas Eckner, Guillaume Horel and Leandro Saita (2009), 'Frailty correlated default', *Journal of Finance* **64**, 2089–2123.

Edmans, Alex, Doron Levit and Devin Reilly (2018), 'Governance Under Common Ownership', *The Review of Financial Studies* **32**(7), 2673–2719.

Fang, Lily and Joel Peress (2009), 'Media coverage and the cross-section of stock returns', *The Journal of Finance* **64**(5), 2023–2052.

Farboodi, Maryam (2023), 'Intermediation and voluntary exposure to counterparty risk', *Journal of Political Economy* **131**(12), 3267–3309.

Fracassi, Cesare (2017), 'Corporate finance policies and social networks', *Management Science* **63**(8), 2420–2438.

Gao, George P., Pamela C. Moulton and David T. Ng (2017), 'Institutional ownership and return predictability across economically unrelated stocks', *Journal of Financial Intermediation* **31**, 45–63.

Gentzkow, Matthew, Bryan T. Kelly and Matt Taddy (2019), 'Text as data', *Journal of Economic Literature* **57**(3), 535–74.

Gofman, Michael, Bernard Herskovic and Gill Segal (2026), 'Networks in finance: Foundations and frontiers', *Annual Review of Financial Economics* . Forthcoming.

Harvey, Campbell R., Yan Liu and Heqing Zhu (2016), '... and the Cross-Section of Expected Returns', *The Review of Financial Studies* **29**(1), 5–68.

Herskovic, Bernard (2018), 'Networks in production: Asset pricing implications', *The Journal of Finance* **73**(4), 1785–1818.

Herskovic, Bernard, Bryan Kelly, Hanno Lustig and Stijn Van Nieuwerburgh (2020), 'Firm volatility in granular networks', *Journal of Political Economy* **128**(11), 4097–4162.

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

Kong, Yaxuan, Hoyoung Lee, Yoontae Hwang, Alejandro Lopez-Lira, Bradford Levy, Dhagash Mehta, Qingsong Wen, Chanyeol Choi, Yongjae Lee and Stefan Zohren (2026), 'Evaluating llms in finance requires explicit bias consideration'.

Leary, Mark T. and Michael R. Roberts (2014), 'Do peer firms affect corporate financial policy?', *The Journal of Finance* **69**(1), 139–178.

Li, Yingying, Guoli Liu, Changlei Lyu and Xinghua Zheng (2025), Uncovering stock risk linkages via mixed membership co-jump networks. Working Paper.

Liu, Ernest and Aleh Tsyvinski (2024), 'A dynamic model of input–output networks', *The Review of Economic Studies* **91**(6), 3608–3644.

Ludwig, Jens, Sendhil Mullainathan and Ashesh Rambachan (2025), Large language models: An applied econometric framework. NBER Working Paper 33344.

Manconi, Alberto, Massimo Massa and Ayako Yasuda (2012), 'The role of institutional investors in propagating the crisis of 2007–2008', *Journal of Financial Economics* **104**(3), 491–518.

Manela, Asaf and Alan Moreira (2017), 'News implied volatility and disaster concerns', *Journal of Financial Economics* **123**(1), 137 – 162.

Miranda-Pinto, Jorge, Alvaro Silva and Eric R. Young (2023), 'Business cycle asymmetry and input-output structure: The role of firm-to-firm networks', *Journal of Monetary Economics* **137**, 1–20.

Mullainathan, S. and A. Shleifer (2005), 'The market for news', *American Economic Review* **95**(1), 1031–1053.

Pellegrino, Bruno (2025), 'Product differentiation and oligopoly: A network approach', *American Economic Review* **115**(4), 1170–1225.

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert and Christopher T Stanton (2020), 'Workplace knowledge flows', *The Quarterly Journal of Economics* **135**(3), 1635–1680.

Scherbina, Anna and Bernd Schlusche (2015), Economic linkages inferred from news stories and the predictability of stock returns. Working Paper.

Schwenkler, Gustavo and Hannan Zheng (2025*a*), 'News-driven peer co-movement in crypto markets', *Journal of Corporate Finance* **93**, 102772.

Schwenkler, Gustavo and Hannan Zheng (2025*b*), Why does news coverage predict returns? Working Paper.

Schwenkler, Gustavo and Hannan Zheng (2026), News-implied production networks and aggregate output. Working Paper.

Shue, Kelly (2013), 'Executive networks and firm policies: Evidence from the random assignment of mba peers', *The Review of Financial Studies* **26**(6), 1401–1442.

Sufi, Amir (2007), 'Information asymmetry and financing arrangements: Evidence from syndicated loans', *The Journal of Finance* **62**(2), 629–668.

Sun, Qianyou, Jiexin Zheng, Bohan Jin, Lihua Chen and Yijie Peng (2025), 'Intercorprel-llm: Enhancing financial relational understanding with graph-language models'.

Valkanov, Rossen (2003), 'Long-horizon regressions: theoretical results and applications', *Journal of Financial Economics* **68**(2), 201–232.

Valta, Philip (2012), 'Competition and the cost of debt', *Journal of Financial Economics* **105**(3), 661–682.

Bilateral relationship	Beneficial channels	Adverse channels
Credit	Repeated credit relationships can strengthen monitoring and improve financing arrangements (Sufi (2007)).	Counterparty credit risk can transmit losses across contracting firms and affect credit prices (Arora et al. (2012)).
Equity	Common ownership may facilitate information diffusion (Gao et al. (2017)) and strengthen governance (Edmans et al. (2018)).	Common ownership may weaken firms' competitive incentives (Azar et al. (2018)) and transmit losses across stocks through portfolio rebalancing (Manconi et al. (2012)).
Leadership	Information diffusion through managerial networks can improve corporate decision-making and firm performance (Fracassi (2017), Sandvik et al. (2020)).	Common information in leadership networks can lead to herding (Chiu et al. (2013)) and worse outcomes (Shue (2013)).
Peer	Information dissemination through direct communications with peers (Cai and Szeidl (2018)) or through indirect learning from competitors (Leary and Roberts (2014)).	Competition may raise firms' cost of debt (Valta (2012)), while information contagion can transmit credit risk across peer firms (Duffie et al. (2009)).
Production	Production links give firms access to specialized suppliers and generate productivity gains (Bernard et al. (2019)).	Customer-supplier links transmit idiosyncratic shocks across firms (Barrot and Sauvagnat (2016)) and information into asset prices (Cohen and Frazzini (2008)).

Table 1: *Firm-level and asset-pricing channels of bilateral firm relationships.* The table summarizes effects documented for linked firms and their securities. It intentionally excludes aggregate implications, which are the focus of our empirical analysis. The list is not exhaustive. See Gofman et al. (2026) for a broader survey.

	Mean	Median	Std dev.	Min.	Max.
Whole population:					
Market capitalization (million USD)	1,323.82	101.18	10,308.68	0.04	659,615.06
Total assets (million USD)	3,917.07	143.58	44,092.36	0.08	2,171,196.98
Total debt (million USD)	3,157.84	61.56	41,145.78	0.00	2,024,344.69
Book leverage (%)	58.03	53.21	92.26	0.00	7,418.05
Cash holdings (million USD)	233.52	11.16	3,816.33	-37.10	383,256.08
Quarterly net income (million USD)	17.25	-0.03	182.55	-5,829.59	7,438.06
Quarterly sales (million USD)	369.68	20.48	2,343.60	-191.45	97,842.59
Quarterly cost of goods sold (million USD)	263.61	12.66	1,897.36	-0.25	87,911.08
Average monthly return (%)	0.32	1.03	5.17	-69.38	224.48
Average monthly volatility (%)	4.10	3.62	2.66	0.00	137.35
Quarterly dividends per share (USD)	0.06	0.00	1.08	0.00	166.42
Monthly analyst coverage	3.39	2.21	3.22	1.00	40.93
Sample mentioned in news data:					
Market capitalization (million USD)	2,004.73	176.40	13,160.44	0.11	659,615.06
Total assets (million USD)	5,581.58	224.10	54,607.55	0.09	2,171,196.98
Total debt (million USD)	4,525.21	99.96	50,983.97	0.00	2,024,344.69
Book leverage (%)	56.18	52.66	79.81	0.00	7,418.05
Cash holdings (million USD)	336.84	17.90	4,755.57	-37.10	383,256.08
Quarterly net income (million USD)	25.69	0.13	223.15	-5,829.59	7,438.06
Quarterly sales (million USD)	518.37	38.37	2,845.21	-90.02	97,842.59
Quarterly cost of goods sold (million USD)	368.42	22.76	2,302.11	-0.25	87,911.08
Average monthly return (%)	0.80	1.23	3.45	-32.95	75.71
Average monthly volatility (%)	3.87	3.44	2.06	0.08	34.82
Quarterly dividends per share (USD)	0.08	0.00	1.35	0.00	166.42
Monthly analyst coverage	3.77	2.51	3.52	1.00	40.93
Monthly firm mentions in news	0.74	0.09	4.56	0.00	221.13
Monthly firm links in news	0.11	0.01	0.70	0.00	30.17

Table 2: *Summary statistics of firm population & sample covered in news data.* We consider a firm to be mentioned in the news if it has at least one monthly mention. The population includes 27,103 distinct firms while the news-covered sample covers 16,387 distinct firms. The data span the time between January 1, 1981, and December 31, 2025. The above statistics are time series moments over months in which a firm was alive during our sample. Market data come from CRSP, fundamentals from Compustat, and analyst following from I/B/E/S, all of which we access through the WRDS API. Total debt is total liabilities. Book leverage is the ratio of total debt over total assets. Analyst coverage counts the number of analysts following a firm.

"JPMorgan (JPM) extended the credit using an inaccurate evaluation of Lehman's worth (LEH), improperly counting Lehman's customer money as belonging to the firm."

Credit:	92.3%	Equity:	0.7%
Leadership:	0.2%	Other:	4.0%
Peer:	0.2%	Production:	2.6%

"Sony (SONY) executive Mike Abary told a press conference that the Intel-powered (INTEL) Internet TVs, are aimed at a young, tech-savvy audience; that's a demographic trade-off."

Credit:	0.1%	Equity:	0.2%
Leadership:	0.1%	Other:	1.1%
Peer:	33.1%	Production:	65.4%

"JPMorgan (JPM) upgrades Teva (TEVA) to neutral from underweight after JPMorgan said it sees a 'favorable catalyst path' for the pharmaceutical company."

Credit:	7.0%	Equity:	0.3%
Leadership:	1.7%	Other:	79.3%
Peer:	5.9%	Production:	5.8%

Table 3: *Examples of sentence labeling.* We mark in green the link category that is assigned the highest probability based on our link labeling methodology.

		Accuracy					Confusion matrix (predicted label)					
		RMSE	R^2	Precision	Recall	F1 Score	Credit	Equity	Leadership	Peer	Production	Other
Synthetic test set:												
True label	Credit	0.0460	0.9873	0.9933	0.9933	0.9933	298	0	0	0	0	2
	Equity	0.0340	0.9931	0.9967	0.9933	0.9950	1	298	0	1	0	0
	Leadership	0.0083	0.9996	1.0000	1.0000	1.0000	0	0	300	0	0	0
	Peer	0.0869	0.9547	0.9732	0.9667	0.9699	0	1	0	290	1	8
	Production	0.0647	0.9749	0.9833	0.9833	0.9833	1	0	0	0	295	4
	Other	0.1010	0.9388	0.9538	0.9633	0.9585	0	0	0	7	4	289
	Overall	0.0649	0.9747	0.9834	0.9833	0.9833						
Human-verified test set:												
True label	Credit	0.4105	-0.0113	0.4889	0.8800	0.6286	44	2	1	1	0	2
	Equity	0.0726	0.9683	0.9434	1.0000	0.9709	0	50	0	0	0	0
	Leadership	0.0793	0.9623	0.9800	0.9800	0.9800	1	0	49	0	0	0
	Peer	0.2564	0.6057	0.8372	0.7200	0.7742	12	0	0	36	2	0
	Production	0.2772	0.5390	0.8824	0.6000	0.7143	14	1	0	4	30	1
	Other	0.2839	0.5163	0.9000	0.5400	0.6750	19	0	0	2	2	27
	Overall	0.2593	0.5967	0.8386	0.7867	0.7905						

Table 4: *Accuracy of Stage 1 model in test sample.* We construct a synthetic dataset of sentences that describe one of the different link types between two fictional firms using LLMs. In Stage 1, we train our model using 60% of this sample as training data, which we also use to measure the in-sample accuracy rate. We determine the number of epochs to train our models based on the minimum categorical cross-entropy loss, which we evaluate on 20% of the sample. Finally, we measure the out-of-sample error of the Stage 1 model in two ways. The first way is by considering the 20% of our synthetic data that was not used for training or cross validation (i.e., the *synthetic test set*). The second way is by considering 300 human-verified sentences describing relationships between two firms and published in the NOW corpus (i.e., the *human-verified test set*). We evaluate errors as follows. For the root mean squared error (RMSE) and R^2, we consider the difference between the link category indicator and the predicted link probability. For the precision, recall, and F1 score, we consider the difference between the link category indicator and the predicted category indicator that marks the argmax across the different categories. Here, "Precision" measures the ratio of true positives over all predicted positives, "Recall" measures the ratio of true positives out of all actual positives, and "F1 Score" gives the harmonic mean between precision and recall. For the confusion matrices, we evaluate categorical indicators using the argmax function across the predicted probabilities for each category. Then, we count the number of predicted categories in the different test samples, broken down by true categories in the data.

			Accuracy		Confusion matrix (predicted label)					
		R^2	Precision	Recall	Credit	Equity	Leadership	Peer	Production	Other
Synthetic test set:										
True label	Credit	0.9351	0.9792	0.9400	282	0	0	0	8	10
	Equity	0.9838	0.9966	0.9800	3	294	0	0	0	3
	Leadership	0.9950	1.0000	0.9900	0	0	297	0	1	2
	Peer	0.9104	0.9818	0.8967	1	1	0	269	3	26
	Production	0.9492	0.9455	0.9833	1	0	0	1	295	3
	Other	0.8642	0.8683	0.9667	1	0	0	4	5	290
	Overall	0.9396	0.9619	0.9594						
Human-verified test set:										
True label	Credit	0.5788	0.7049	0.8600	43	2	1	2	1	1
	Equity	0.9755	0.9615	1.0000	0	50	0	0	0	0
	Leadership	0.9663	0.9800	0.9800	0	0	49	0	0	1
	Peer	0.6908	0.8478	0.7800	4	0	0	39	3	4
	Production	0.6317	0.7917	0.7600	6	0	0	2	38	4
	Other	0.5721	0.7674	0.6600	8	0	0	3	6	33
	Overall	0.7359	0.8422	0.8400						

Table 5: *Out-of-sample accuracy of Stage 2 model.* We evaluate the accuracy of the Stage 2 model fine-tuned over 100 epochs using the same synthetic and human-verified news sentences as in Table 4. For the R^2, we consider the difference between the link category indicator and the predicted link probability. For the precision and recall, we consider the difference between the link category indicator and the predicted category indicator that marks the argmax across the different categories.

Firm 1	Firm 2	Link strength	Link count	Sample sentence
Credit: 9,033 sentences covering 2,610 distinct firms with link strength above p^+.				
NLT	American General	43.3%	139	"*The decision frees NLT, a Nashville insurance company, to make early payment for up to 9.9 percent of the shares already tendered by American General stockholders, an amount that would not require regulatory approval.*"
PanAm	Delta Airlines	41.2%	108	"*To help Pan Am get to that point, Delta will also provide $80 million of temporary financing, pending closing of the asset purchase.*"
People Express	Texas Air	33.0%	104	"*If it acquires People Express, Texas Air would take on an additional $700 million in debt.*"
Equity: 18,731 sentences covering 4,182 distinct firms with link strength above p^+.				
Time Warner	Turner Broadcasting	70.1%	323	"*Time Warner said it would buy the 82 percent of Turner it did not already own in a stock deal worth about $7.5 billion.*"
Seagram	Conoco	57.9%	222	"*Analysts said Seagram's stake in the new DuPont-Conoco company could go as high as 20 percent, which would make it the second-largest shareholder after the duPont family and their trusts, which control about 22 percent.*"
Chris-Craft	Warner Comm.	56.7%	170	"*The transaction makes Chris-Craft the largest stockholder in Warner, replacing Mr. Murdoch, whose stake will be diminished to 5.6 percent from 7 percent.*"
Leadership: 19,770 sentences covering 3,801 distinct firms with link strength above p^+.				
Merrill Lynch	Enron	31.1%	106	"*Some securities lawyers, however, say Merrill and Mr. Tilney could have had a conflict of interest because of Mr. Tilney's relationship with Mr. Fastow and because Merrill Lynch advisers to Enron, like Mr. Tilney, may have had their judgment clouded by their investment in LJM2.*"
Chris-Craft	Warner Comm.	30.2%	170	"*In recent weeks, Chris-Craft and Warner have been feuding over Chris-Craft's request for more representation on the Warner board.*"
Jos. A. Bank	Men's Wearhouse	29.4%	102	"*Jos. A. Bank is probably hoping that shareholder pressure from overlapping shareholders and arbitrageurs will push Men's Wearhouse to a deal.*"

Table 6: *Most frequently mentioned links (Part I).* This table lists the three strongest links for different link categories. We measure link strength in accordance with Eq. (3). We restrict ourselves only to links that are identified at least 100 times in our news sample.

Firm 1	Firm 2	Link strength	Link count	Sample sentence
Peer: 33,368 sentences covering 3,413 distinct firms with link strength above p^+.				
Vanguard	Fidelity	70.9%	128	"*First, they say, much of the growth in retail fund flows is coming from 40(k) contributions at the workplace, a market dominated by big fund companies like Fidelity and Vanguard with impressive performance records and economies of scale.*"
Citigroup	Bankers Trust	70.0%	134	"*Even big commercial banks such as Bankers Trust and Citibank have been nibbling at the edges, and thousands of savings and loan institutions may not be far behind.*"
Amazon	Barnes & Noble	68.3%	167	"*The best example is Amazon.com, which continues to widen the gap between itself and its better-known book-selling competitor, Barnes & Noble.*"
Production: 40,261 sentences covering 5,531 distinct firms with link strength above p^+.				
Apple	Motorola	83.0%	321	"*The Macintosh uses a chip made by Motorola Inc., and some models will soon use the Power PC chip, the product of a joint development by Apple, Motorola and the International Business Machines Corporation.*"
Hitachi	Nippon Electric	74.8%	124	"*The Japanese electronics companies Hitachi and NEC said they would join forces to manufacture Internet routers and switches.*"
Nissan	Chrysler	72.3%	123	"*A separate agreement, in place since 2004, through which an affiliate of Nissan builds transmissions for Chrysler, will continue, the companies said.*"
Other: 32,923 sentences covering 5,007 distinct firms with link strength above p^+.				
Pfizer	Merck	59.2%	196	"*Nor have any documents yet emerged, as they did at Merck, suggesting that Pfizer had doubts about its drug.*"
Bank of America	Wells Fargo	57.9%	111	"*Bank of America has modified only 4 percent of the eligible mortgages, and Wells Fargo has modified 6 percent.*"
Morgan Stanley	Merrill Lynch	55.4%	222	"*Bank of America Merrill Lynch and Morgan Stanley advised JAB Holding.*"

Table 7: *Most frequently mentioned links (Part II).* This table lists the three strongest links for different link categories. We measure link strength in accordance with Eq. (3). We restrict ourselves only to links that are identified at least 100 times in our news sample.

		Alternative networks					
		Equity (WRDS subsidiary data)	Leadership (BoardEx company profiles)	Peer (TNIC of Hoberg and Phillips (2016))	Production (Compustat customer segments)	Production (BEA total requirements with self-links)	Production (BEA total requirements without self-links)
Cosine similarities:							
Our networks	Credit	0.0805	0.0830	0.0475	0.1715	0.2092	0.2288
	Equity	0.1658	0.1479	0.0654	0.1758	0.2750	0.3136
	Leadership	0.0598	0.1965	0.0645	0.2017	0.3941	0.4781
	Peer	0.0649	0.1971	0.0944	0.2716	0.4736	0.4497
	Production	0.0609	0.1800	0.0707	0.3941	0.4737	0.5165
	Other	0.0718	0.1997	0.1019	0.2809	0.4891	0.4871
Link regressions:							
Our networks	Credit	*** 1.9965	0.3880	* 0.4658	0.3163	-0.8547	-1.0368
		(3.3874)	(0.8399)	(2.2911)	(0.2529)	(-0.9655)	(-1.8717)
	Equity	*** 2.5833	* 0.7569	** 0.4247	-0.0719	-0.2145	0.2183
		(7.2077)	(2.5590)	(2.5950)	(-0.2773)	(-0.6112)	(0.6624)
	Leadership	* 1.0257	*** 1.0103	* 0.4572	0.1054	0.0971	1.0536
		(2.1195)	(4.2489)	(2.2066)	(0.3073)	(0.2116)	(2.4310)
	Peer	* 0.9320	* 0.5765	*** 0.5302	0.0094	** 0.9814	0.4461
		(2.0915)	(2.4085)	(4.4879)	(0.0816)	(3.4738)	(1.5805)
	Production	0.7413	° 0.3981	0.1262	0.1001	° 0.8351	* 1.0315
		(1.5922)	(1.7378)	(0.9630)	(1.1024)	(2.9076)	(3.7392)
	Other	* 0.9585	* 0.4251	*** 0.6270	0.0767	0.8174	0.3348
		(2.3617)	(2.0231)	(3.7718)	(0.4439)	(2.1597)	(0.9113)
Regression type		Quasibinomial	Quasibinomial	Quasibinomial	Gamma	Gamma	Gamma
Regression link		Logit	Logit	Logit	Log	Log	Log
Level		Firm	Firm	Firm	Firm	Sector	Sector
Self-links		N	N	N	N	Y	N
Total active links		8,255	7,626	38,842	6,707	276	253
Total linked nodes		1,410	1,411	1,847	1,421	23	23
Included active links		4,531	6,138	34,004	462	276	253
Included linked nodes		1,120	1,324	1,657	421	23	23
Active links (alt. net.)		168	981	34,004	462	276	253
Active links (our net.)		4,503	6,030	1,578	99	203	184

Table 8: *Link regressions & cosine similarities.* Cosine similarities are the inner product of the link-strength vectors from an alternative network and one of our six networks, divided by the product of their ℓ^2-norms. For the regressions, we regress each alternative network's link strength on the link strengths from our six networks, including log average market cap of the linked firms as a control for differential coverage across databases (omitted for the BEA sectoral regressions). Standard errors are clustered at the firm level via a bootstrap: we resample nodes with replacement, reconstruct all sampled pairs (dropping self-pairs), and re-estimate the regression. For the sectoral regressions based on the BEA data, we employ a wild bootstrap at the sector level to address the small number of sectors; see Cameron et al. (2008). The standard error is the standard deviation of the resulting coefficients across 1,000 repetitions. Values in parentheses are t-statistics, with significance assessed against a Student's t-distribution with degrees of freedom equal to the number of resampled nodes minus one. ***, **, *, and ° denote significance at the 99.9%, 99%, 95%, and 90% confidence levels, respectively. Cells marked in green (yellow) mark the strongest (second strongest) positive associations.

Variable	Start	End	Mean	Median	Std dev.	Min.	Max.
3-month Treasury rate (%)	1981-01	2025-12	3.88	3.94	3.28	0.01	16.30
2-year Treasury rate (%)	1981-01	2025-12	4.58	4.27	3.56	0.12	16.46
10-year Treasury rate (%)	1981-01	2025-12	5.50	4.70	3.21	0.62	15.32
Average yield curve slope (%)	1981-01	2025-12	0.92	0.83	0.89	-1.36	2.83
Monthly CRSP market log-return (%)	1981-01	2025-12	0.77	1.18	4.52	-25.98	12.15
Monthly S&P 500 log-return (%)	1981-01	2025-12	0.73	1.16	4.37	-24.54	12.38
Average monthly VIX level	1990-01	2025-12	19.46	17.67	7.42	10.13	62.67
Industrial production log-growth	1981-01	2025-12	0.13	0.18	0.97	-14.14	6.35
Consumption log-growth	1981-01	2025-12	0.24	0.24	0.86	-11.59	7.99

Table 9: *Summary statistics of our aggregate variables.* All factors are sampled at the monthly frequency. We take averages of levels within a month when necessary. Log-growth is computed from previous-month close to current-month close.

	Optimal holding period	Return	Volatility	Annualized: Sharpe Ratio	CAPM α	CAPM β
Passive S&P 500		9.33%	14.58%	0.47	0.00%	1.00
Passive 3M Treasury		2.45%	0.58%	0.00	0.00%	0.00
Credit:						
Density	1	8.58%	8.98%	* 0.63	* 3.26%	0.35
Concentration	59	5.67%	7.11%	0.38	-0.18%	0.42
Intermediation	60	6.22%	7.91%	0.41	0.11%	0.46
Equity:						
Density	35	6.88%	7.20%	* 0.54	° 1.48%	0.36
Concentration	60	5.36%	8.72%	0.28	-1.05%	0.51
Intermediation	59	7.35%	8.08%	0.54	1.10%	0.48
Leadership:						
Density	31	7.52%	7.54%	* 0.60	1.96%	0.38
Concentration	59	5.16%	6.64%	0.33	-0.40%	0.38
Intermediation	15	8.34%	9.48%	0.57	1.68%	0.54
Peer:						
Density	31	7.22%	7.31%	* 0.58	1.74%	0.37
Concentration	60	5.46%	7.62%	0.33	-0.44%	0.43
Intermediation	19	7.72%	8.7%	0.54	1.51%	0.48
Production:						
Density	32	7.38%	7.55%	* 0.58	° 1.73%	0.39
Concentration	60	5.75%	6.76%	0.41	0.16%	0.39
Intermediation	3	7.67%	8.83%	0.53	1.92%	0.41

Table 10: *Trading performance.* We implement the market-timing strategy described in the text and report the holding period that achieves the highest Sharpe Ratio. CAPM alphas and betas use the S&P 500 as the market portfolio. The monthly risk-free rate compounds to the lagged annualized three-month Treasury rate over one year. Significance follows from the AR(1) placebo exercise and a Benjamini-Hochberg adjustment across network types for each metric and holding horizon. "Passive S&P 500" and "Passive 3M Treasury" report buy-and-hold benchmarks over the identical sample period.

Signal	Network	Outcome	Horizon (months)	Accuracy (%)	Excess (p.p.)
Low density	All	Positive consumption growth	4 to 10[a]	55.0 to 62.2	4.6 to 8.0
High density	All	Rising 10-year Treasury yield	1 to 9	54.9 to 62.0	4.8 to 12.0
High intermediation	Production	Positive consumption growth	3 to 4	60.0 to 60.5	5.0 to 5.9
High intermediation	Peer	Positive consumption growth	18 to 20	64.7 to 65.3	5.2 to 5.8

Table 11: *Directional forecasting results.* Accuracy is the fraction of correctly predicted outcome directions. Excess accuracy subtracts the accuracy implied by the unconditional frequencies of predicted and realized increases. Reported density ranges use horizon-network combinations significant at the 10% level after adjusting across networks. Four of five consumption signals and all five Treasury signals include at least one horizon significant at the 5% level. Both intermediation patterns are significant at the 5% level. Not every network is significant at every horizon within a reported range. [a]The credit-density signal extends through 13 months. The online data repository provides all 4,500 results.

	Density	Concentration	Intermediation
Equity:			
Adjusted R^2	0.7881	0.7707	0.1260
Wald(existing network / subsidiary)	0.3270	* 3.8607	0.9086
	(0.0868)	(0.0868)	(0.0868)
Wald(news coverage)	*** 11.4057	* 2.7457	0.3663
	(0.0355)	(0.0355)	(0.0355)
Leadership:			
Adjusted R^2	0.7578	0.6872	0.2425
Wald(existing network / BoardEx)	1.1995	0.6785	0.7660
	(0.0428)	(0.0428)	(0.0428)
Wald(news coverage)	*** 9.8093	*** 9.7922	1.0091
	(0.0412)	(0.0412)	(0.0412)
Peer:			
Adjusted R^2	0.8238	0.7547	0.4387
Wald(existing network / TNIC)	*** 9.0215	0.2832	** 5.4737
	(0.0693)	(0.0693)	(0.0693)
Wald(news coverage)	*** 32.5660	*** 13.7087	*** 8.4552
	(0.1753)	(0.1753)	(0.1753)
Production:			
Adjusted R^2	0.8572	0.7224	0.4545
Wald(existing network / segments)	*** 6.3166	1.1277	° 2.6738
	(0.0299)	(0.0299)	(0.0299)
Wald(news coverage)	*** 33.4162	*** 5.9224	* 4.3446
	(0.1497)	(0.1497)	(0.1497)

Table 12: *Spanning tests for news-implied network metrics.* We estimate regressions as in Eq. (5) for each Network Type n and Connectivity Metric m, and test the null hypotheses $b_{n,m,A} \equiv 0$ and $b_{n,m,N} \equiv 0$. Reported values are F-statistics for Wald tests. We apply Newey-West standard-error adjustments with lag 12 and also adjust the corresponding p-values across networks within connectivity metric category using the false-discovery-rate procedure of Benjamini and Hochberg (1995). ***, **, *, and ° denote significance at the 99.9%, 99%, 95%, and 90% confidence levels, respectively. The values in parentheses report the incremental adjusted R^2 from relaxing the tested constraints (either $b_{n,m,A} = 0$ or $b_{n,m,N} = 0$).

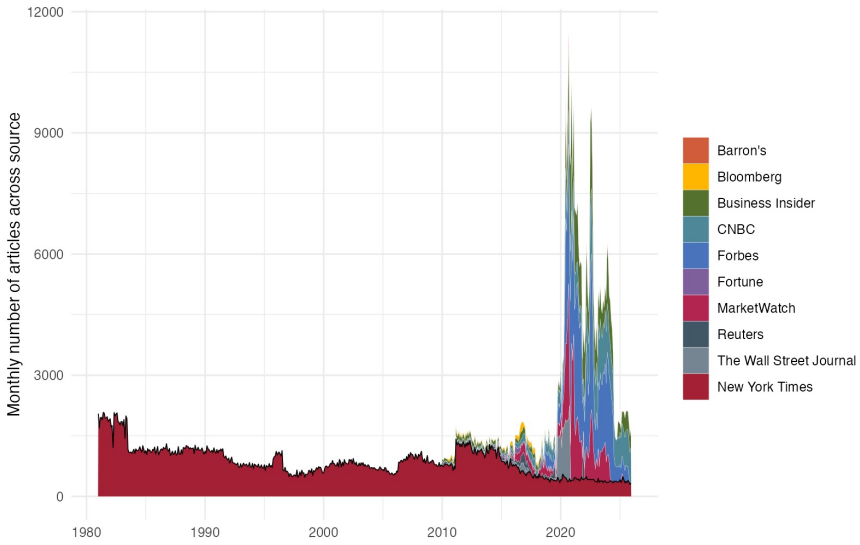


Figure 1: *Monthly number of articles.* The plot shows the monthly number of articles contained in our news sample across the different news outlet sources. The black line separates data collected from the New York Times Archive from those obtained from the NOW Corpus.

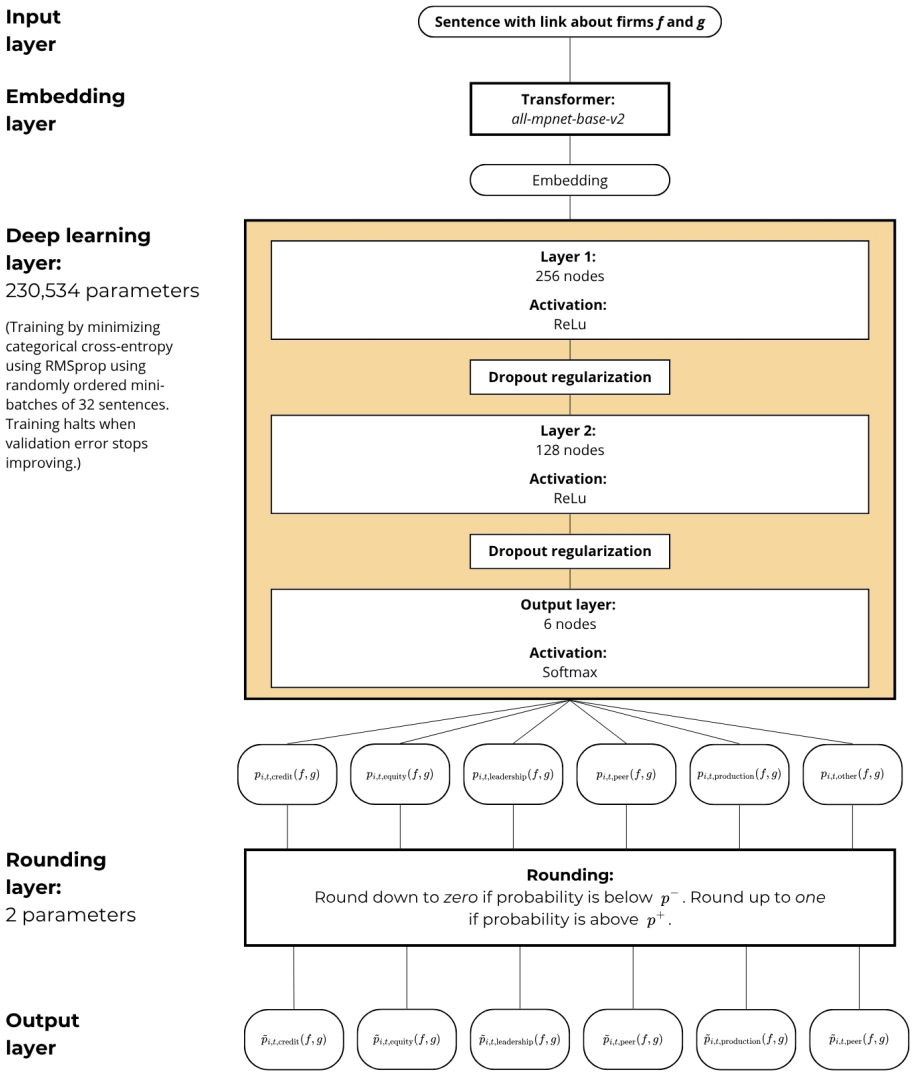


Figure 2: *Model architecture.* This figure shows the architecture of the model underlying the link-labeling methodology in Section 3.



(a) Credit.



(b) Equity.



(c) Leadership.



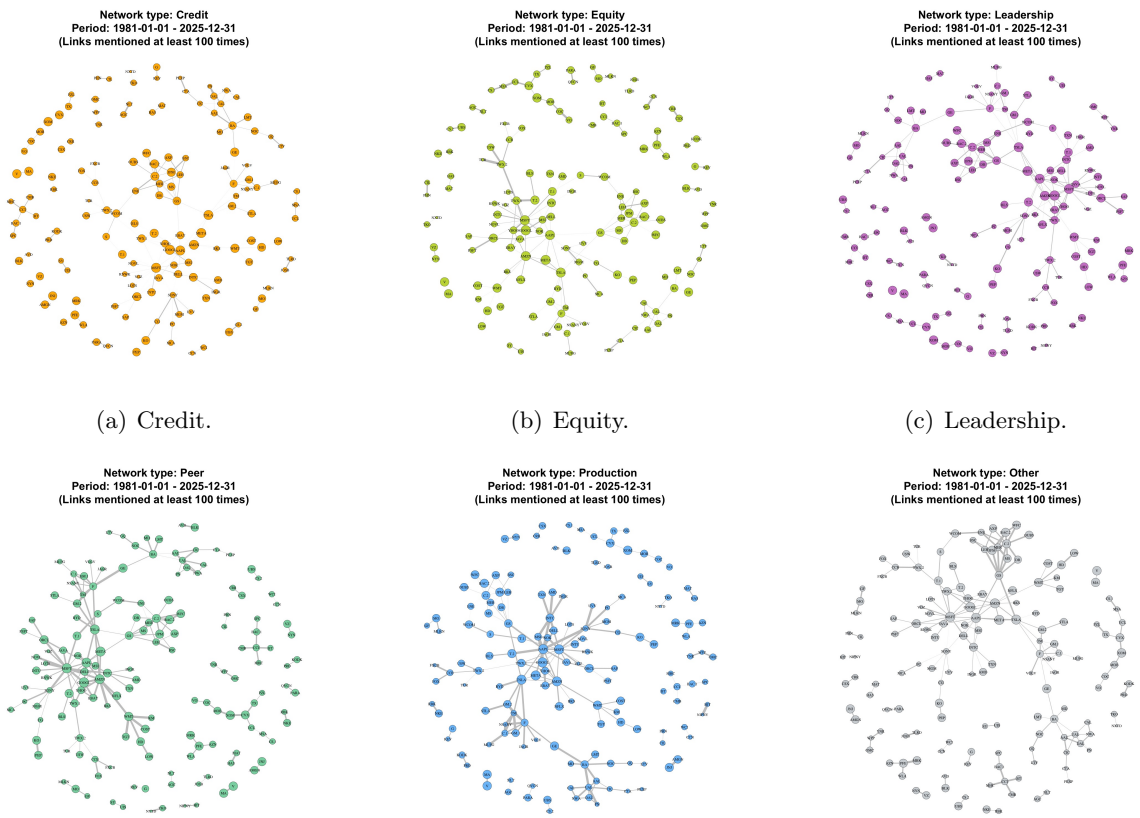
(d) Peer.



(e) Production.

(f) Other.

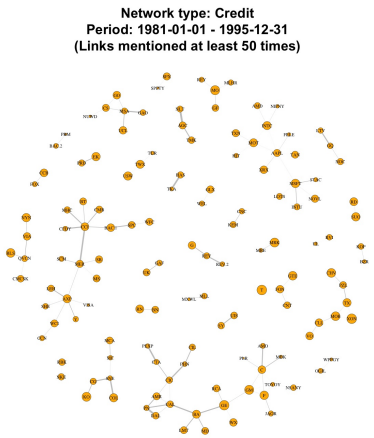
Figure 3: *Most frequent nouns.* These word clouds display the 100 most frequent nouns in strong sentences (i.e., those for which our methodology assigns a link probability above the round-up threshold p^+). We use the universal point-of-speech tagger from the cleanNLP package available in R. We restrict ourselves to the English language model from spaCy (*"en_core_web_sm"*). We lemmatize all nouns prior to displaying them in a word cloud. We exclude proper nouns as well as the nouns "company," "month," "share," "percent," "week," and "year."



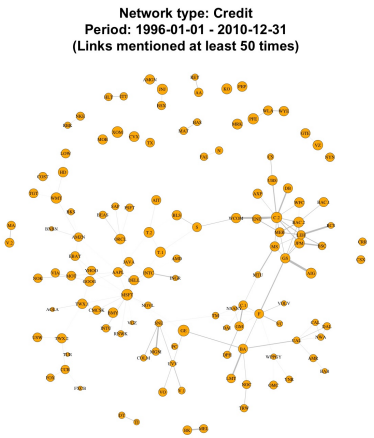
(a) Credit. (b) Equity. (c) Leadership.

(d) Peer. (e) Production. (f) Other.

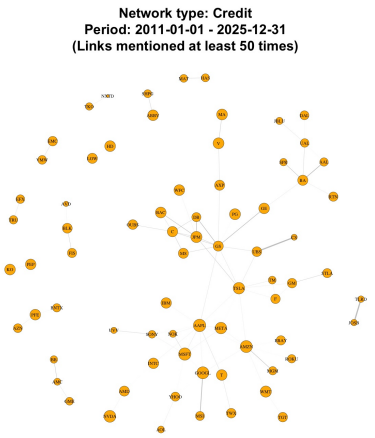
Figure 4: *Aggregate news-implied networks.* These graphs show the different news-implied networks constructed using our methodology, aggregated over the sample period. We only include links that were identified at least 100 times over the full sample. The size of a node is proportional to the logarithm of the firm's average market capitalization over the sample. The width of a link is proportional to the estimated link strength. The link widths are normalized across graphs so that a thicker line means a stronger link regardless of the network in which it appears.



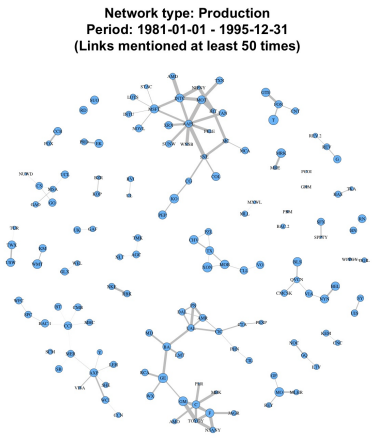
(a) Credit (1981-1995).



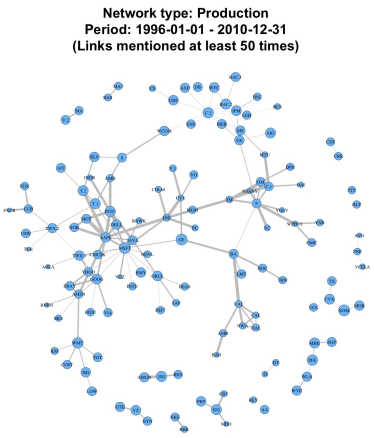
(b) Credit (1996-2010).



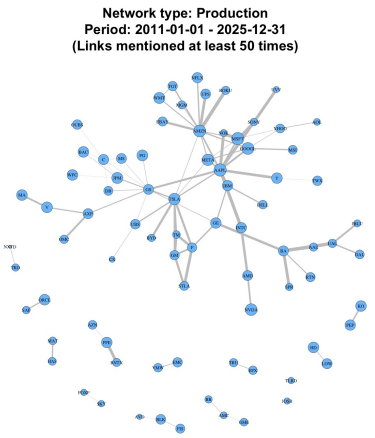
(c) Credit (2011-2025).



(d) Production (1981-1995).



(e) Production (1996-2010).

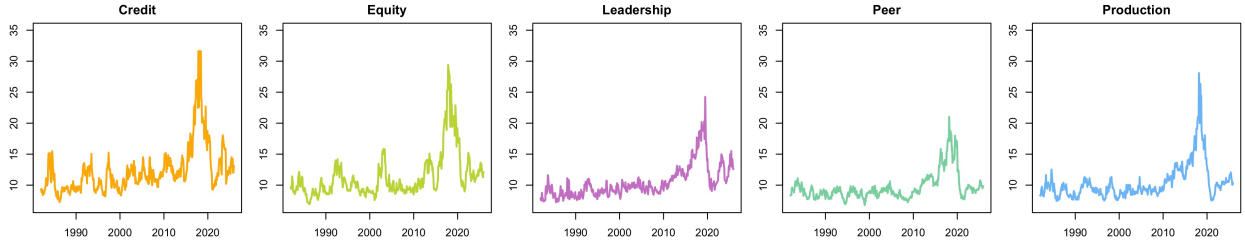


(f) Production (2011-2025).

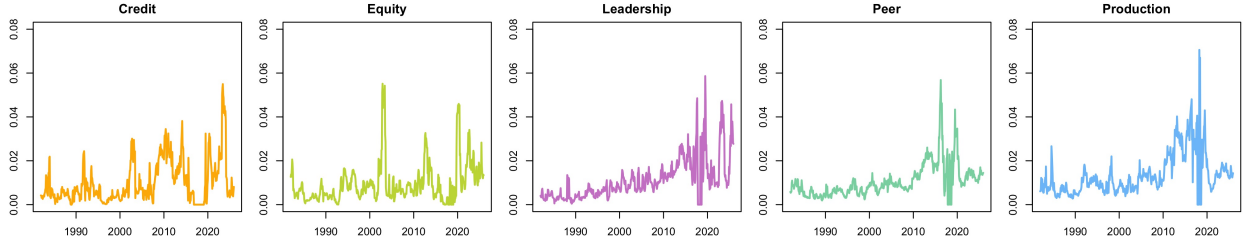
Figure 5: *Time series of news-implied networks.* These graphs show the different news-implied networks constructed using our methodology, aggregated over subsequent 15-year periods. We only include links that were identified at least 50 times over the respective sub-samples. The size of a node is proportional to the logarithm of average market cap of the firm over the sample. The width of a link is proportional to the estimated link strength. The link widths are normalized across graphs so that a thicker line means a stronger link regardless of which network this link lives in.



(a) Density.

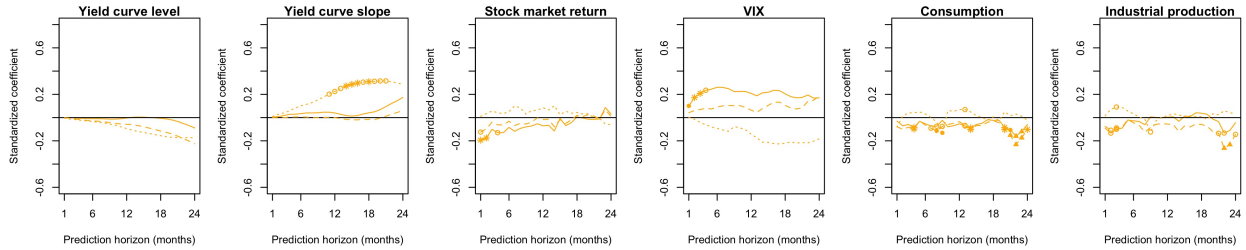


(b) Concentration.

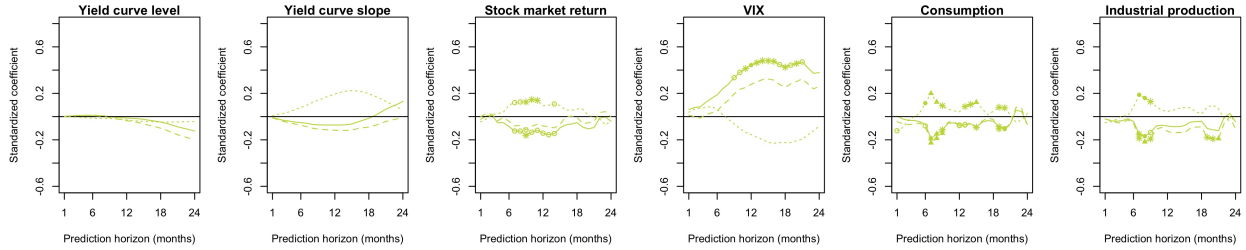


(c) Intermediation.

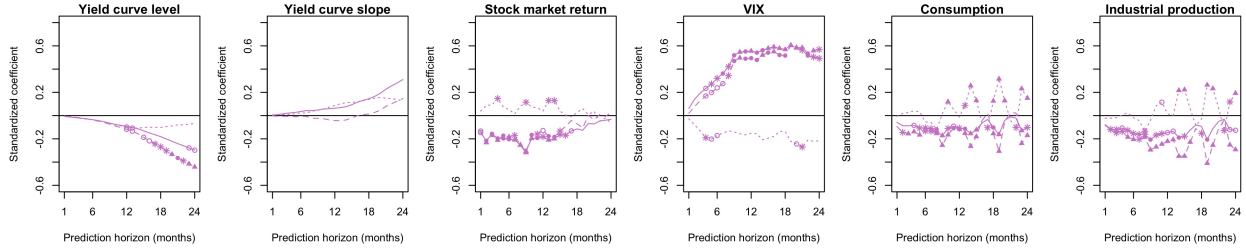
Figure 6: *Time series of connectivity metrics.* For our monthly forecasting application, we follow Schwenkler and Zheng (2026) and evaluate density as $\bar{d}_t = \frac{1}{N}\sum_{n=1}^{N} d_t^n$, where $N = 2,000$ is the number of nodes in the network, and $d_t^n = \sum_{j\neq n} \mathbf{1}_{\{w_t^{j,n}>0\}}$ is the number of active links of Node n in the network updated in Month t. Here, $w_t^{j,n}$ is the number of identified links that connect Nodes j and n during the rolling 12-month aggregation window ending before Month t. Furthermore, setting $D_t^n = \sum_{j\neq n} w_t^{j,n}$ as the weighted degree of a node and $\bar{D}_t = \frac{1}{N}\sum_n D_t^n$ as the average weighted degree, we evaluate concentration as $\sqrt{\mathrm{Var}(D_t^n)}/\bar{D}_t$ and intermediation as $\sqrt{\sum_{n=1}^{N}\sum_{j\neq n}\sum_{k\neq j,n} w_t^{j,n} D_t^j w_t^{k,n} D_t^k}/\bar{D}_t^2$. These formulas follow from Acemoglu et al. (2012) and Schwenkler and Zheng (2026).



(a) Credit network.

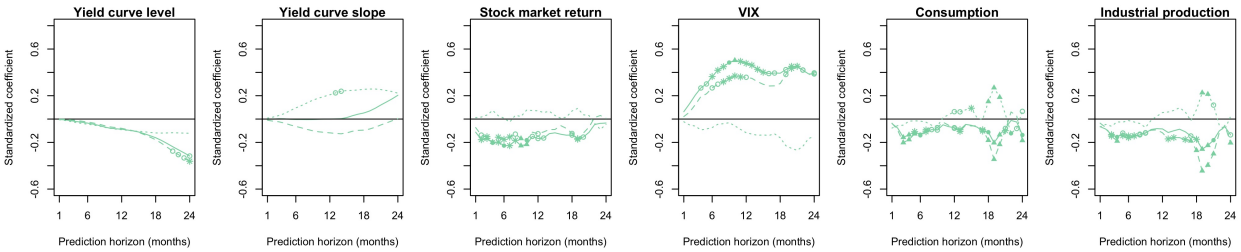


(b) Equity network.

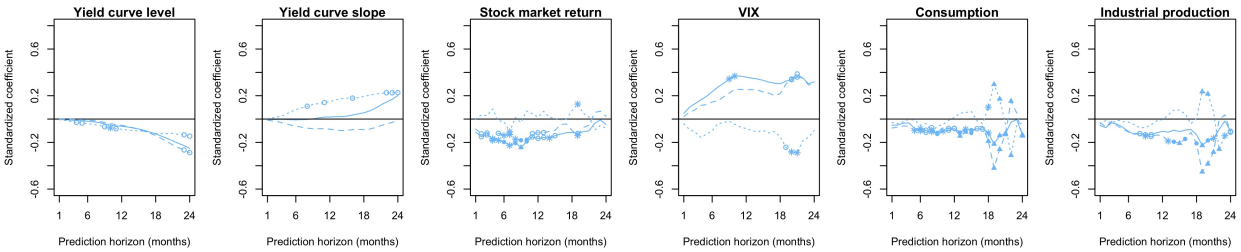


(c) Leadership network.

Figure 7: *Coefficient estimates for Regression* (4) *(Part I).* We fit Regression (4) for each network type and prediction horizon, and report the resulting estimates for the regression coefficients $\beta_{D,n,h}$ (density, solid lines), $\beta_{C,n,h}$ (concentration, dashed lines), and $\beta_{I,n,h}$ (intermediation, dotted lines). We mark statistical significance through different symbols. An empty circle, a star, a filled circle, and a filled triangle mark significance at the 90%, 95%, 99%, and 99.9% confidence levels, respectively. We apply the false-discovery-rate p-value adjustment of Benjamini and Hochberg (1995) across connectivity metrics for a fixed prediction horizon.

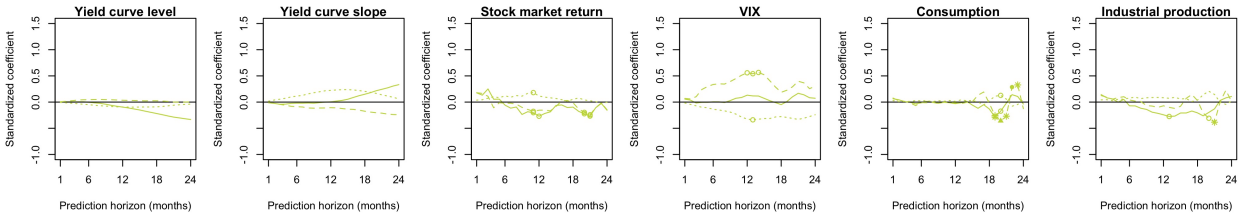


(a) Peer network.

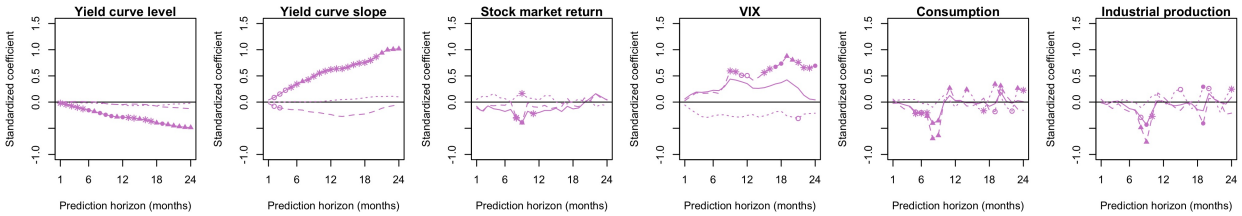


(b) Production network.

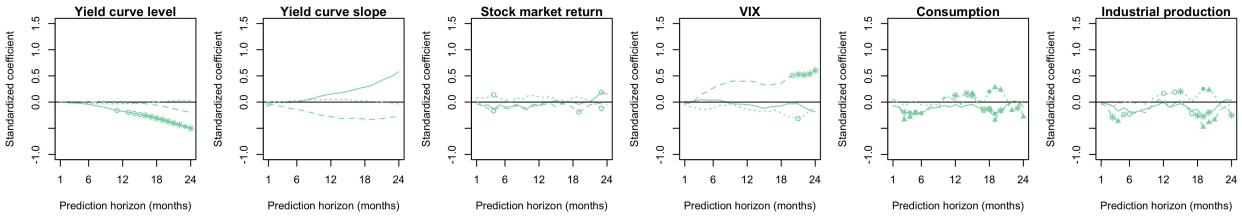
Figure 8: *Coefficient estimates for Regression* (4) *(Part II).* The notes to Figure 7 apply. This figure reports the peer and production networks.



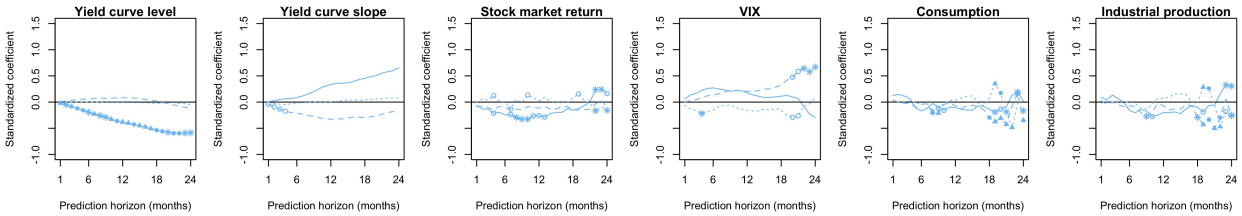
(a) Equity network.



(b) Leadership network.

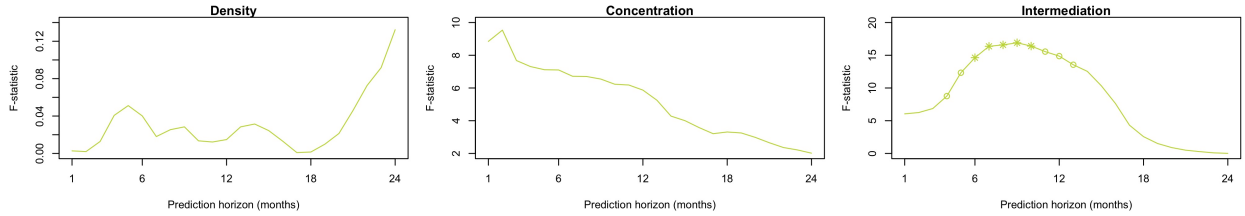


(c) Peer network.

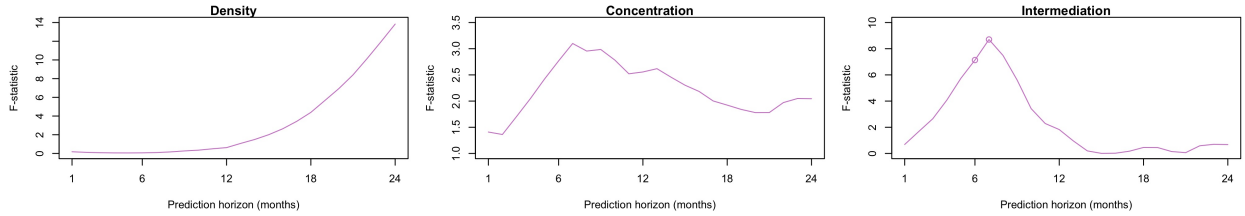


(d) Production network.

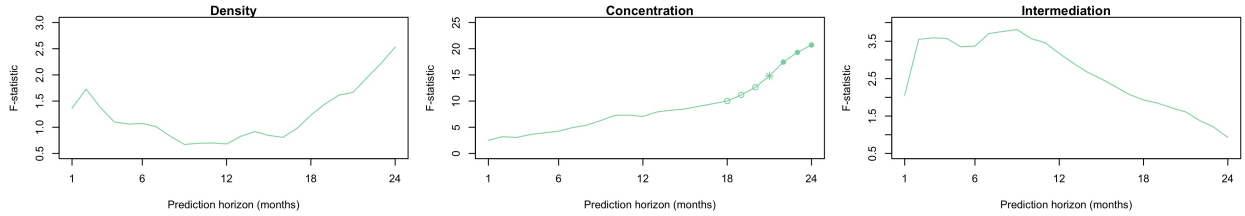
Figure 9: *Coefficient estimates for Regression* (6). Lines report standardized coefficients on density (solid), concentration (dashed), and intermediation (dotted). An empty circle, star, filled circle, and filled triangle denote placebo significance at the 10%, 5%, 1%, and 0.1% levels, respectively, after applying the false-discovery-rate adjustment of Benjamini and Hochberg (1995) across the three metrics at each horizon.



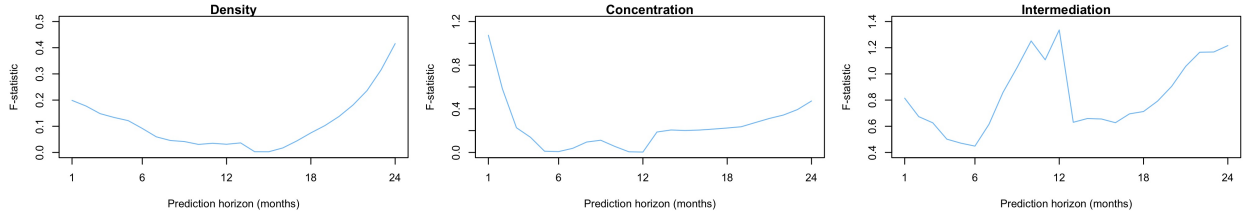
(a) Equity network.



(b) Leadership network.



(c) Peer network.



(d) Production network.

Figure 10: *Does news-implied connectivity predict alternative connectivity?* Lines report Newey–West F-statistics, using lag $\max\{12, h-1\}$, for whether the news-implied residual predicts the matched metric h months ahead beyond its own current value. Symbols denote placebo significance at the 10%, 5%, 1%, and 0.1% levels after the false-discovery-rate adjustment across metrics at each horizon. The p-values use 1,000 persistence-matched draws.