The Economic Consequences of Firm Networks

V. Hilt & G. Schwenkler*

August 13, 2026†

Abstract

Economic links can facilitate beneficial interactions among firms but also propagate shocks, making the aggregate consequences of firm connectivity theoretically ambiguous. We construct time-varying credit, equity, leadership, peer, and production networks from financial news to empirically distinguish these effects. Network topology strongly predicts future aggregate outcomes, particularly at longer horizons. Greater density and concentration predominantly forecast lower stock returns, industrial production, and consumption, while intermediation has distinct and often opposite implications. Credit intermediation predicts lower future credit spreads and market volatility. Our findings show that the aggregate consequences of interconnectedness depend critically on both the economic nature and topology of firm links.

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 connected through many different economic relationships. They borrow from one another, own stakes in one another, share managers and directors, compete in the same markets, and exchange goods and services through production relationships. A large literature shows that these connections can have important consequences for firm-level and aggregate outcomes. Production networks can propagate idiosyncratic shocks and generate aggregate fluctuations (Acemoglu et al. (2012), Carvalho (2010), Gabaix (2011), Herskovic (2018)), credit networks can transmit financial distress (Azizpour et al. (2018), Eisenberg and Noe (2001)), and ownership, leadership, and peer relationships can affect information diffusion, corporate policies, competition, and asset prices (Azar et al. (2018), Fracassi (2017), Leary and Roberts (2014), among others).

Yet the aggregate implications of firm connectivity are theoretically ambiguous. Economic links can be beneficial because they facilitate information diffusion, monitoring, financial intermediation, resource sharing, and specialization. The same links can be harmful because they create channels for contagion and shock propagation, generate common exposures and correlated behavior, or weaken competition. Table 1 summarizes these competing channels across different types of firm relationships. Importantly, their relative importance may depend not only on whether firms are connected, but also on how those connections are organized. A network in which firms are densely connected to one another may have very different aggregate implications from one in which otherwise separate groups of firms are connected through intermediaries.

Empirically distinguishing these mechanisms is difficult because firm networks are difficult to observe. Existing datasets typically measure one particular type of relationship, cover only a subset of firms, or are updated infrequently. Supply-chain databases, for example, provide valuable information about production relationships but have limited coverage. Other researchers have developed indirect measures of peer relationships (Hoberg and Phillips (2016), Lee et al. (2015)), common technologies (Lee et al. (2019)), or interconnectedness inferred from asset prices (Diebold and Yılmaz (2014), Duan et al. (2012), Hautsch et al. (2014)). Such measures have substantially advanced the literature, but their economic interpretation can be difficult because the underlying links are not always observed directly (Abbassi et al. (2017), Craig et al. (2023)). For some eco-

nomically important relationships, most notably firm-level credit exposures, comprehensive data are simply not publicly available. More fundamentally, differences in data sources and measurement make it difficult to compare different economic networks within a common empirical framework.

This paper develops such a framework. We use financial news to construct time-varying, labeled networks of firms. Our methodology reads sentences in which two firms are mentioned and estimates the probability that the sentence describes one of six relationships: *credit*, *equity*, *leadership*, *peer*, *production*, or *other*. We aggregate these sentence-level classifications into firm links and, in turn, into networks. Because the same methodology and underlying news data are used for every relationship type, the resulting networks provide a common measurement system through which economically different forms of firm connectivity can be studied and compared.

We apply the methodology to 849,079 articles published between 1981 and 2025. Our data combine the New York Times archive with the News on the Web (NOW) corpus, from which we collect articles published by major business news outlets including the Wall Street Journal, Bloomberg, Reuters, CNBC, Forbes, and others. We identify more than 2.2 million firm mentions involving over 16,000 distinct firms and more than 40,000 unique links. Unlike many existing network datasets, our approach can construct different types of firm networks at arbitrary frequencies and for a large common universe of firms. Moreover, the links are interpretable: each estimated connection can be traced to the news sentences from which it was inferred.

Our network construction methodology combines modern natural language processing with supervised deep learning. We first transform sentences into numerical embeddings using a sentence transformer and then use a neural-network classifier to estimate probabilities over the six relationship types. Training proceeds in two steps. We initially train the classifier on a balanced sample of 9,000 synthetic sentences generated by several large language models using fictional firm names. We then fine-tune the model using human-verified sentences drawn from actual financial news, including difficult cases that the initial model misclassifies. This few-shot update materially improves performance in real news data: the overall out-of-sample F1 score on human-verified sentences increases from 79% to 84%, while the out-of-sample R^2 increases from 59.7% to 73.6%. We additionally provide conditions under which the estimated news-implied links converge to the

underlying firm relationships as news coverage and training samples become sufficiently large.

We conduct several exercises to establish that the resulting networks capture economically meaningful relationships. Representative sentences and keyword analyses show that the classifier identifies the intended content of each network. Credit links are associated with language concerning loans, bonds, debt, and repayment; equity links with ownership, acquisitions, and shareholders; peer links with competition and markets; production links with customers, products, and services; and leadership links with executives, directors, and employment. We then compare our networks with independently constructed benchmarks based on WRDS subsidiary data, BoardEx, the Text-based Network Industry Classifications of Hoberg and Phillips (2016), Compustat customer segments, and the Bureau of Economic Analysis input-output tables. Despite being constructed from an entirely different information source, our labeled networks generally exhibit their strongest positive associations with external networks measuring the corresponding economic relationship. These comparisons provide external validation that the methodology recovers genuine firm links rather than merely patterns of news co-coverage.

We next use the networks to study the aggregate implications of firm connectivity. This exercise exploits a feature of our data that is difficult to obtain from existing sources: we can observe several economically distinct networks among a common universe of firms, using the same measurement methodology and over the same sample period. For each network, we construct rolling measures of three dimensions of network topology: *density*, which captures the prevalence of direct connections; *concentration*, which captures the extent to which connections are concentrated among a subset of firms; and *intermediation*, which captures the importance of firms that connect otherwise indirectly connected parts of the network. We ask whether these measures predict subsequent credit spreads, stock-market returns, volatility, industrial production, and consumption.

We find that network topology contains substantial information about future aggregate conditions, but that this information differs systematically across both network types and outcomes. The relationships are weak at short horizons and become substantially stronger over horizons of one to three years. At the three-year horizon, credit-network connectivity predicts both AAA and BAA credit spreads and future stock-market volatility, while the other networks exhibit com-

paratively little predictive power for credit spreads. All five economically labeled networks contain information about subsequent industrial production, and all five jointly predict future consumption growth. Equity, leadership, and peer networks also contain information about future stock-market outcomes. Thus, different economic networks contain distinct information about subsequent aggregate conditions rather than merely reflecting a common news or connectivity factor.

Our central empirical result is that the predictive implications of interconnectedness depend critically on network topology. Density and concentration predominantly predict worse subsequent aggregate outcomes. Denser and more concentrated leadership, peer, and production networks forecast lower industrial production and consumption and, in many specifications, lower stock-market returns. The magnitudes are economically large. At the three-year horizon, for example, a one-standard-deviation increase in leadership-network concentration predicts approximately a 0.7-standard-deviation decline in industrial production and a 0.55-standard-deviation decline in consumption. The negative relationship between production-network connectivity and subsequent real activity closely parallels Schwenkler and Zheng (2026), who obtain the same qualitative result using a different production-network dataset and methodology, providing external validation for the predictive content of our networks.

Intermediation behaves very differently. Its estimated effect generally lies to the right of those of density and concentration and in several cases reverses sign. Greater equity-network intermediation predicts higher subsequent stock returns, industrial production, and consumption. Greater credit-network intermediation predicts substantially lower future market volatility and narrower credit spreads. At the three-year horizon, a one-standard-deviation increase in credit-network intermediation predicts approximately a 0.5-standard-deviation decline in the VIX. Thus, an economy in which firms become more directly or more concentratively connected can exhibit very different subsequent outcomes from one in which firms increasingly intermediate relationships between otherwise distinct parts of the network.

These findings provide a way to reconcile competing views of economic interconnectedness. The literature has identified beneficial effects of firm links through information diffusion, monitoring, resource sharing, specialization, and financial intermediation (Sufi (2007), Greenwood and Jovanovic

(1990), Giroud and Mueller (2019), Fracassi (2017), Cai and Szeidl (2018), Bernard et al. (2019)), but also adverse effects through contagion, shock propagation, herding, and competition (Eisenberg and Noe (2001), Barrot and Sauvagnat (2016), Chiu et al. (2013), Azar et al. (2018), Pellegrino (2025)). Our evidence suggests that these mechanisms need not be contradictory. Greater direct and concentrated connectivity is associated with weaker subsequent aggregate outcomes, consistent with architectures that create greater direct or concentrated exposure to shocks. Intermediation, in contrast, is associated in several networks with more favorable subsequent outcomes, consistent with an architecture in which firms connect otherwise distinct parts of the economy and facilitate beneficial economic interactions without requiring pervasive direct exposure. Although our predictive analysis does not establish a causal mechanism, it yields a unified empirical conclusion: the aggregate implications of firm connectivity depend not only on the economic nature of the links, but critically on how those links are organized.

Our paper makes three contributions. First, we contribute to the literature on economic networks by developing a scalable methodology for measuring multiple types of firm relationships within a common framework. Existing research has typically studied production, credit, ownership, leadership, or peer networks separately. Our approach makes it possible to construct these networks from a common information source, at high frequency, and for a broad universe of firms. The methodology is also extensible: new relationship types can be introduced by generating training examples for the desired category and updating the classifier. We make the resulting monthly network data available to facilitate applications beyond those considered in this paper.

Second, we contribute to the literature studying the granular origins of aggregate fluctuations (Acemoglu et al. (2012), Gabaix (2011), Herskovic (2018), and others). Whereas much of this literature focuses on production networks, our common measurement framework allows us to compare the predictive content of several economically distinct networks. We show that production relationships are only one component of the connection between firm-level network structure and aggregate outcomes. Credit, equity, leadership, and peer networks contain distinct information about subsequent financial and real conditions.

Third, our results emphasize topology as a central dimension of economic interconnectedness.

The distinction between density, concentration, and intermediation shows that "more connected" is not a sufficient description of a network's economic implications. Architectures characterized by more direct or concentrated connections tend to precede weaker aggregate outcomes, while architectures characterized by greater intermediation can predict stronger real activity and more favorable financial conditions. This result provides empirical guidance for theories in which network structure determines whether firm connections diversify shocks and facilitate beneficial interactions or instead amplify and propagate disturbances.

The remainder of the paper is organized as follows. Section 2 describes the news and firm data. Section 3 develops the network-labeling methodology and its theoretical properties while Section 4 describes the training and validation of the classifier. Section 5 analyzes the resulting networks and compares them with existing network databases. Section 6 studies the relationship between network topology and subsequent aggregate financial and economic outcomes. There is an accompanying Online Appendix and Online Data Repository 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 population of firms 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 Figure 1 display the industry distribution in our firm population.

2.2 News data

We collect news articles from two different datasets that are likely to contain business-relevant news. For both datasets, we end the samples on December 31, 2025 because our CRSP data is only

updated annually, and 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 2 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 only 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 ourselves to articles published in reputable outlets that report about 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 we impose 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.

Table 3 shows the distribution of news articles across the different outlets, as well as their temporal span. While the New York Times represents our largest news datasource covering the full sample period, we also obtain a large number of articles from other sources. Forbes, CNBC, and Business Insider span through the end of 2025 and provide nearly 280,000 articles combined. Notably, we also include articles from The Wall Street Journal and Bloomberg, which are likely to be high quality and business-focused. For some sources, the data ends before the end of our sample as coverage in the NOW Corpus terminated. We assume that this is because the news outlets imposed paywalls that prevented NOW from scraping the required data. Figure 2 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.[1] 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-firms. In the third step, it matches the sample of possible firms 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.[2] We adopt a few minor modifications to address peculiarities of the NOW Corpus, primarily to remove embedded photo attribution lines

[1] Our methodology does not hinge on Schwenkler and Zheng (2026). Any method that extract sentences with two firm mentions can be used in this step. Alternative approaches include Chen et al. (2021), Ge et al. (2023), and Scherbina and Schlusche (2015).

[2] Analyses in Schwenkler and Zheng (2026) suggest that it suffices to identify firm connections by co-mentions in sentences versus whole body.

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), and remove repeated sentences within a day.

Table 2 provides summary statistics and Figure 1 shows the industry distribution for the firms that are identified in our news sample using the methodology of Schwenkler and Zheng (2026). 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. We observe that the manufacturing and retail industries are overrepresented in our news sample, while the finance industry is underrepresented.

Moving to firm links, the methodology of Schwenkler and Zheng (2026) finds 166,556 sentences with two firm mentions in our sample.[3] It establishes a total of 42,959 unique links among 9,852 distinct firms. While the aggregate link count is large, links tend to only be infrequently identified in our sample. On average, our approach establishes 0.11 links per firm and month, indicating that only one-in-seven firm mentions are associated with firm links.

Figure 3 shows the implied uncategorized firm network, restricted to links that are mentioned at least 100 times in the news sample. There, the link widths are proportional to the log link frequency and the node sizes are proportional to the log average market caps of the firms. We observe that the news-implied network has multiple clusters, each showcasing star architectures as in Acemoglu et al. (2012) while being connected only through 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 (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) to the left of the banking cluster, and an oil and energy cluster at the top left containing

[3]The methodology finds c sentences with two or more firm mentions. Given that sentences with more than two firm mentions may be more complex, and given that 80% of the recognized sentences with firm links anyways only mention two firms, we restrict ourselves to two-firm mentions for simplicity.

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: 1) credit, 2) equity, 3) leadership, 4) peer, 5) production, and 6) 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 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 necessarily consider analyst co-coverage unless an analyst explicitly links the two firms as peers.
- **Production**: Link arises through the flow of goods, inputs, services, logistics, manufacturing, or operational dependency. It explicitly excludes trade credit or professional services not linked to production.
- **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 4 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, S2ORC, 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 with two hidden layers of 128 units each, followed by an output layer that uses a softmax function to compute a probability for each of the six link categories. 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:

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

where we suppress the dependency on Firms f and g to imply that this holds for all firm pairs (f,g). 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 4 provides examples of how our methodology works, where we mark in parentheses the tickers of the two firms identified in the sentence. Our methodology correctly labels the first sentence as almost certainly describing a credit relationship, which is correct given the information

communicated about JPMorgan's loan to Lehman Brothers. It also correctly infers a production relationship from the second sentence, which in passing mentions Sony being a customer of Intel when reporting about Mr. Abary's press conference. Finally, when faced with a sentence stating that analysts at JPMorgan upgrade its forecasts for Teva, the methodology infers a label of "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 $t_0 < t_1$ (i.e., $t_0 =$ start of the current month and $t_1 =$ start of the following month) 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 up or down 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 thresholds that we specify in Section 4. Finally, we aggregate sentence link strength proxies by averaging them out:

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

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 $N_{c,[t_0,t_1)}(f, g) = |\mathcal{S}_{[t_0,t_1)}(f, g)|$ is the size of this set. We set $\tilde{p}_{c,[t_0,t_1)}(f, g) = 0$ whenever $N_{c,[t_0,t_1)}(f, g) = 0$ 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 $\tilde{p}_{c,[t_0,t_1)}(f, g)$.

While we could also construct networks directly based on the estimated link probabilities $p_{i,t,c}$

instead of the rounded proxies $\tilde{p}_{i,t,c}$, rounding is empirically relevant and 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 Consistency

The labeled networks we extract from the news are only approximations of the true underlying networks. 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; see García (2018), Fang and Peress (2009), Solomon and Soltes (2012), Schwenkler and Zheng (2025), and others. 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 consistency of our methodology. The proof can be found in Online Appendix E.

Theorem 3.1. *Consider Firms f and g and fix the Link Category c. For times $0 \leq t_0 < t_1 < \infty$, write $I_{c,[t_0,t_1)}(f,g)$ as the indicator that marks whether Firms f and g shared 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. This means that, given unlimited coverage capacity, the news would eventually report all existing links but never misreport non-existent links:

$$\mathbb{P}\left(\textit{News reports Link } c \textit{ between Firms } f \textit{ \& } g \textit{ during } [t_0,t_1) \,\middle|\, I_{c,[t_0,t_1)}(f,g)=1\right) \to 1 \textit{ and}$$
$$\mathbb{P}\left(\textit{News reports Link } c \textit{ between Firms } f \textit{ \& } g \textit{ during } [t_0,t_1) \,\middle|\, I_{c,[t_0,t_1)}(f,g)=0\right) \to 0$$

as the number of sampled sentences grows large ($N_{c,[t_0,t_1)}(f,g) \to \infty$).

(A2) There is vanishing ambiguity in the way the news reports about firm relationships so that

$$\mathbb{P}\left(\textit{Sentence } i \textit{ reports about Link types } c \textit{ and } c' \neq c \textit{ between Firms } f \textit{ \& } g\right) \to 0$$

as the number of sampled sentences grows large ($N_{c,[t_0,t_1)}(f,g) \to \infty$).

(A3) The firm identification algorithm is consistent so that:

$$\mathbb{P}\left(\textit{Firm } f \textit{ identified in Sentence } i \,|\, \textit{Firm } f \textit{ mentioned in Sentence } i\right) \to 1 \textit{ and}$$
$$\mathbb{P}\left(\textit{Firm } f \textit{ identified in Sentence } i \,|\, \textit{Firm } f \textit{ not mentioned in Sentence } i\right) \to 0$$

for almost all Sentences i *as the size of the training sample grows large* ($N_{firm} \to \infty$).

(A4) The labeling methodology is consistent so that

$$\mathbb{E}\left(p_{i,t,c}(f,g) \,|\, \textit{Sentence } i \textit{ reports Link } c \textit{ between } f \textit{ and } g\right) \to 1 \textit{ and}$$
$$\mathbb{E}\left(p_{i,t,c}(f,g) \,|\, \textit{Sentence } i \textit{ reports a link other than } c \textit{ between } f \textit{ and } g\right) \to 0$$

for almost all Sentences i *as the size of 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* $\epsilon > 0$, *so that* $0 < \epsilon \leq p^- < p^+ \leq 1 - \epsilon < 1$.

Then, the discrepancy between the link strength approximation in Eq. (3) and the true link indicator becomes vanishingly small as the number of sampled sentences and the training samples grow large:

$$\left|\tilde{p}_{c,[t_0,t_1)}(f,g) - I_{c,[t_0,t_1)}(f,g)\right| \to 0 \quad \textit{almost surely as} \quad N_{c,[t_0,t_1)}(f,g), N_{firm}, N_{class} \to \infty.$$

Subject to the thresholds p^- and p^+ falling inside of the unit interval (Assumption (A5)), Theorem 3.1 decomposes the accuracy of our network estimates into three independently verifiable 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 (Chahrour et al. (2021), Mullainathan and Shleifer (2005), Nimark and Pitschner (2019), and others), whereby paid news outlets only report the subset of events that are most newsworthy. Assumption (A2) requires that sentences that report more than one firm relationship, such as those that describe peers merging or entering production partnerships, become asymptotically rare.

While Theorem 3.1 is asymptotic in nature and its conditions can only be verified approximately in finite samples, it provides a formal basis to validate our approach. We present a validation exercise for the assumptions of Theorem 3.1 in Online Appendix F and provide a summary here. Assumption (A1) justifies our choice to focus on New York Times articles and reputable business publications, such as the Wall Street Journal and CNBC, given that these sources are regarded as fact-checked and editorially rigorous. This makes systematic misreporting of non-existent firm links unlikely. Assumption (A2) supports our choice to use a multi-class classifier for which the class probabilities are complementary and satisfy Eq. (1). If this assumption were violated, then it may be more suitable to use independent binary classifiers for each link type. We present such an extension in Section 4.4. We verify in Online Appendix F that the frequency of ambiguous sentences in our sample is rare, supporting our approach. Assumption (A3) justifies our choice to use the firm identification algorithm of Schwenkler and Zheng (2026). Table 3 of Schwenkler and Zheng (2026) shows that the algorithm achieves a precision of 96.5%, which supports the false positive condition in Assumption (A3). The recall rate is lower at 48.9%, though this figure is measured only against ticker-coded firm mentions and likely understates the algorithm's true recall across all firm mentions in the corpus. The practical effect of imperfect recall is to reduce the effective number of sentences $N_{c,[t_0,t_1)}(f,g)$, which slows convergence but does not introduce false co-mentions that would corrupt the link probability estimates. Finally, we formally validate Assumption (A4) in Online Appendix F. This requires an analysis of the training and testing accuracy of our methodology, which we present in the next section.

4 Model training, testing, and extensions

Our methodology is high dimensional. In total, there are 115,720 free parameters. 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 hyperparameters p^+ and p^- of Eq. (2). 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 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.

Our approach to train using synthetic data has advantages compared to human-labeled sentences. First, human labeling may result in unbalanced panels that may confound the learning because some link types (like peers) are more frequently reported. Using an LLM allows for the

creation of balanced panels. Second, human labeling is costly as it requires human supervision and expertise. LLMs can generate precise sentences quickly and cheaply (albeit at a possibly large environmental cost). Third, human labeling subjective because it is based on how individual people interpret different sentences. Despite its advantages, training based on synthetic data also comes with one major weakness: the synthetic sentences are likely simple and may not fully represent the complexity of true journalistic reporting. This may bias our model estimates. We overcome this weakness in Stage 2, where we update our model using few-shot learning based on a smaller sample of sentences that were actually published in the news and for which we manually verified the link label. Thus, the model estimated in Stage 1 is only a starting point and not the final model we will 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 them to the corresponding embedded sentences. We randomly split the synthetic sentence universe into a training set (60% of the sample or 5,400 sentences), a cross-validation set (20% of the sample or 1,800 sentences), and a test set (20% of the sample or 1,800 sentences) that each contain a balanced amount of each link type. Given the large complexity of the model, we use advanced machine learning techniques to train our model with reasonable computational resources.[4] We fit the model by minimizing the categorical cross-entropy loss, and repeatedly evaluate its performance in the validation set to determine certain hyper-parameters. We use the RMSprop adaptive learning rate algorithm that scales how quickly the model learns, and process mini-batches of 32 sentences at a time, ordered randomly 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 halt the training via early stopping when the validation loss fails to improve for three consecutive epochs. Then, we restore the parameters from the best-performing epoch. The full technical specification of our training procedure is described in Online Appendix G. 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 one containing 300 actual news

[4]The training runs on a standard MacBook laptop.

sentences from the NOW corpus with human-verified link labels. We describe our approach to construct human-verified data set in Section 4.2.

Tables 5 presents the out-of-sample 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 not as 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 through a difficulty in labeling credit, peer, production, and other links. We observe that the Stage 1 model has a low recall but high precision rate for peer, production, and other links in actual news data. This means that the Stage 1 model has a hard time identifying sentences that actually report about these links between two firms. But, once it recognizes such a link, it is very likely correct. In 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 links. This can also be observed through the confusion matrix in Table 5, 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%. Next, Stage 2 aims to improve the performance of the labeling model.

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 pick sentences that were actually published in our news sample and verify their labels using human labor force. 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. Such an approach has shown success in boosting the performance of AI models trained on synthetic data; see Duan and Pelger (2026) and

[5]Precision (recall) measures the ratio of true positive over all predicted (actual) positives, while the F1 score gives the harmonic mean of precision and recall. These performance metrics are commonly used in the machine learning literature; see El Mrabet et al. (2021), Shobha and Rangaswamy (2018), and others.

Ludwig et al. (2025), among others. 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 both in the NOW and New York Times corpuses.

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 equally split this initial set of human-verified news sentences into a training set covering 300 sentences that will be used to run a few-shot update of the Stage 1 model, and a test set covering 300 sentences that will be used to evaluate the resulting Stage 2 model. We ensure that the training and test sets have balanced category memberships.

The sentences picked previously pose an average difficulty for our labeling methodology because they are picked at random. To ensure that the model is updated with sentences that pose a high difficulty, we complement the initial set with human-verified false positives that the Stage 1 model got wrong. To do this, we take the sentences as labeled by the Stage 1 model and sort them by the 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 only able to find 8 false positive sentences, though we successfully find 25 false positives per category for the remaining categories. This results in a false positive test set containing 133 human-verified sentences. These sentences are included in the set of human-verified training sample that will be used to update the Stage 1 model. All in all, the human-verified training sample contains 433 sentences with uneven memberships (fewer credit sentences) while the human-verified test sample contains 300 sentences with equal proportions.

[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.

We implement the model update in Stage 2 by running back propagation on the Stage 1 model parameters using the same adaptive learning rate algorithm with randomly-ordered mini-batches of 32 sentences. However, unlike Stage 1, in Stage 2 we apply an inverse weighting scheme to ensure that the credit category with 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 back propagation over many more epochs to enable sufficient learning on the human-verified sentences. We pick the number of epochs to maximize the out-of-sample F1 score that balances both precision and recall. This is evaluated in the test set of human verified sentences.[7] Table 6 reports the out-of-sample F1 score of the resulting Stage 2 model over different number of epochs. We find that 100 epochs delivers the highest out-of-sample F1 score for the Stage 2 model. We fix this number of epochs for Stage 2 going forward.

Table 7 reports the out-of-sample precision and recall rates, the out-of-sample R^2, and the confusion matrix for the fine-tuned model of Stage 2. We observe that the few-shot update with human-verified sentences provides a boost in 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 thanks to increases in the precision score when labeling credit sentences and as well as increases in the recall score 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 leads to near balanced recall and precision rates across categories, suggesting that the model does not systematically over-predict or under-predict any link type. Despite the improvements in accuracy, the Stage 2 model still exhibits the lowest precision when labeling credit links. We attribute this to the scarcity of credit sentences in the news sample.

[7] Ideally, we would select the number of epochs on a separate set of human-verified sentences used for cross validation. However, given the large costs associated with human verification and the limited cases of clearly reported sentences in the news data set, we re-use the test set for both epoch selection and final evaluation. Reassuringly, the out-of-sample F1 score in Table 6 stays within about two percentage points of its peak for any epoch count between 75 and 250, suggesting our results are not overly sensitive to this choice.

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 5 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 allow for the identification of additional link types. Suppose that a user wanted to also build a network of analyst coverage (the third example in Table 4 suggests that the "other" category may include such links). To extend our methodology to also identify this kind of link, a user can follow these steps:

(1) Use a large language model (LLM) to construct a synthetic set of sentences that describe analyst coverage links. To not confuse the learning, ask the LLM to include exactly two

fictional firms. Also ask to write the sentences in a way that would appear in common financial news outlets. Generate the same number of sentences as had previously been generated for all other link types in order to keep the panel balanced.

(2) Expand the deep learning classifier to include an additional category. Then, retrain the model in Stage 1 including the newly added synthetic sentences, and relabel all news sentences.

(3) Construct a new set of human-verified sentences that include analyst co-coverage. Then, re-run Stage 2 including these new human-verified sentences.

(4) Use the updated Stage 2 model to compute predicted probabilities for all news sentences.

(5) Recalibrate the round-up and round-down thresholds p^+ and p^- in Stage 3.

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 one single multi-class classifier of Figure 4. 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 as well as the implied labeled networks, both in aggregate over the whole sample as well as 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 folows. 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 8–9 shows 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 suggest that our model adequately identifies the different types of firm links. The model correctly identifies direct credit exposures, such as Delta providing a loan to PanAm. It also picks up liabilities that may arise during merger and acquisition (M&A) transactions (such as NLT executing a delayed payment for American General shares, or Texas Air assuming debts from People Express). In unreported analyses, we uncover that the credit category also captures indirect credit exposures (like when two companies invest in common credit assets) as well as alternative credit relationships between industrial firms (such as the leasing of assets). For production links, the model captures customer-supplier links between industrial firms (Nissan and Chrysler), as well as joint production ventures (Apple and Motorola, Hitachi and Nippon Electric). We observe that some sentences report about more than two firms even though the firm identification methodology of Schwenkler and Zheng (2026) we use only recognizes two firm mentions (like the sentence describing the production link between Apple and Motorola, which also refers to IBM). This shows that we do not discard all sentences that refer to more than two firms even though we only assign links between two firms at most from each sentence.

Peer links clearly capture competitive links, while equity links pick up direct M&A investments and equity. Leadership links capture instances in which companies have overlapping shareholders (Jos. A. Bank and Men's Wearhouse), or there may be close relationships between employees, directors, or shareholders of different companies (Merrill Lynch and Enron). They also capture instances in which one company assigns directors or manages employees at another company (such as between Chris-Craft and Warner, or Columbia and Sony). In unreported sentences, we uncover that leadership links also capture situations in which employees worked at different companies over time. Notably, our model allows for two firms to share multiple types of relationships. For example, for Chris-Craft and Warner Communications, our model assigns a 56.7% likelihood that the link is equity and a 30.2% likelihood that the link is leadership.

The representative sentences for "other" links suggest that this category may subsume known

relationships that cannot be cleanly identified from the text they are extracted from. All of the strongest other links are between well-known peers. However, unlike the sample sentences for identified peer links, none of the sentences for the identified other links use explicit comparative or competitive language.

Tables 8–9 document that certain types of links are more abundant than others. Judging by the number of strong sentences (i.e., those that score link probabilities above our round-up threshold p^+), production links are most abundant while credit links are most sparse. Naturally, many sentences are classified as describing "other" links. Figure 6 showcases the distribution of the news sources that published strong sentences, broken down by link category. We observe that we primarily identify strong links in the New York Times sample. CNBC and Forbes typically come in as the next most frequent sources. Almost all strong credit sentences (91%) are identified in the New York Times sample that restricts the usage of the data for the training of AI models. This confirms that the challenge we faced in Stage 2 when searching for false positive credit sentences not published in the New York Times is due to a structural feature of our data: credit sentences are rarely contained in the NOW Corpus.

5.2 Keyword analyses

We complement Tables 8–9 by running keyword analyses on the sample of strong sentences. Figures 7 and 8 display the most frequently mentioned verbs and nouns in these sentences across link categories. In the word clouds, we exclude 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 keywords further suggest that each network captures information that is related to the link type they represent. For example, nouns like *bank*, *bond*, *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*, *stake*, and *own* 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*, *develop*, *make*, *product*, and *service*. Reflecting their more person-oriented nature, sentences in which we identify leadership kinds of 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*, *rise*, and *stock*; analyst recommendations like *analyst*, *estimate*, *expect*, *recommend*, and *target*; and event reporting like *announce*, *case*, *court*, and *lawsuit*. Complementing the sample sentences of Tables 8–9, the keywords suggest that "other" links may mask structural relationships between firm that are not identified by our methodology.

Putting everything together, the results of this section suggest that our methodology adequately identifies the different types of firm links we consider. While the evidence in this section is only superficial, we provide more rigorous evidence validating our methodology in Section 5.4, in which we compare our labeled links to those available in existing databases.

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 take all sentences with two identified firm mentions published in our sample, compute their corresponding link probabilities, round up or down using the thresholds p^- and p^+, and then average the rounded link probabilities across all sentences that contain the firm pair to compute the strength of the link. We plot networks for each category including only firms links that were identified at least 100 times in our sample and for which the resulting link strength is not zero. Figure 9 shows the six resulting networks.

Compared to the unlabeled network in Figure 3 that is based purely on sentence co-mentions, the labeled networks of Figure 9 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 graphs suggest that Goldman and Apply most likely share a production link, while Goldman and Tesla most likely share 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 Babus and Hu (2017), 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 10 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 H. 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.[8] In contrast, the evolution of the production network suggests that the US production economy

[8] As Schwenkler and Zheng (2025) show, the intensity of news reporting may be an important signal of systematic risk exposure for individual stocks.

may have become more interconnected. The production network for the earlier period shows many disjoint production clusters. The graphs for the latter period show more connections across clusters. For example, GE connects Boeing (BA) to the technology cluster through Microsoft (MSFT) in the intermediate period. Goldman (GS) connects financial services companies such as American Express (AXP) to Tesla (TSLA) and Apple (AAPL) in the later period. The evolution of the production network aligns with Acemoglu and Azar (2020), Carvalho and Voigtländer (2014), and others, who empirically and theoretically show 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 exercise in Section 6.

5.4 Comparison to existing networks

We compare our labeled networks to alternative networks available in different databases. We consider the following firm-level networks, which serve as our benchmarks:[9]

(1) An equity network based on parent-subsidiary relationships as reported in the WRDS company subsidiary data, spanning the years 1996 through 2018. We match a parent firm's GVKEY to PERMCOs in our sample. The subsidiary firm is only identified by its name. We thus use an analogous matching methodology as in Schwenkler and Zheng (2026) to match firm names to the CRSP sample, and obtain the corresponding subsidiary PERMCO. We restrict ourselves to US subsidiaries. Links are given by the indicators of whether two firms shared a parent-subsidiary relationship in the WRDS data during sample period.

(2) A firm-level leadership network based on common executive and director positions as reported

[9]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)), rendering the construction of a benchmark infeasible. The other network captures residual links that combine a range of potential alternative links. It is therefore not possible for us to construct a clean benchmark for this network.

in the BoardEx database. We match ISINs in the BoardEx sample to CUSIPs in our sample, and build links as the indicators that mark whether two companies shared at least one common executive or director over the sample period, which starts in the year 2000.

(3) A firm-level peer network based on the Text-based Network Industry Classifications (TNIC) of Hoberg and Phillips (2016). We match GVKEYs in the TNIC data to PERMCOs in our data, and restrict our data to the same sample period as the TNIC data (1988-2021). Links in the TNIC data are sampled annually and bounded between zero and one. They measure the similarity between business descriptions in the annual 10-Ks of two different firms. We construct aggregate links as averages of the annually sampled business similarity scores.

(4) A firm-level production network based on the Compustat customer segments data. Links in these data are sampled quarterly and represent sales between a customer and a supplier. We construct aggregate links as the logarithm of one plus the cross-quarter average of the sum of quarterly sales between a customer and a supplier, and vice versa. We use the same approach as in Schwenkler and Zheng (2026) to match firms in the Compustat sample to our firms, and also restrict our data to the same sample period as Compustat (starting in 1994-Q4).

Because the alternative networks are measured over differing sample periods, we construct our labeled networks independently for each alternative while matching its sample period. To enable a fair comparison, we standardize the universe of firm links across the two networks and set the strength of any link missing from either to zero. For TNIC and Segments, whose link strengths are continuous, the universe is the set of active links present in the alternative network. BoardEx and Subsidiary only contain active links with strength equal to one, so an unreported pair cannot be distinguished from a non-existent one. Thus, we fix the universe to the links identified in our news data, supplemented with the links reported by the alternative network. In all cases, we restrict ourselves to the 1,000 largest firms by average market cap over the relevant sample period. Online Appendix H shows that a larger sample of firms yields analogous results.

In addition to the firm networks above, we also consider an inter-sectoral production network based on the BEA industry-by-industry total requirement tables. These tables are sampled every

five years and capture the total inputs that one industry in the US economy requires from another. We aggregate industries at the two-digit NAICS code, measure link strengths as the sum of each sector's inputs from the other (and vice versa) in a given table, and aggregate across the 1997–2021 sample by taking the logarithm of one plus the average of these sums. We construct the corresponding labeled inter-sectoral network by averaging our firm-level links between the same two-digit NAICS sectors. The BEA network is full, meaning that every sector has at least one connection. Thus, we include all links contained in the BEA sample and set the link strength in our network to zero when missing.

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 across networks and is commonly used to compare networks; see Abbassi et al. (2017), 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 cluster, whichever is appropriate. We repeatedly resample firms (sectors) with replacement, reconstruct every pair among the resampled entities, re-estimate the regression, and take the standard deviation of the resulting coefficients as the standard error.[10]

Table 10 presents our results, where we mark in green (yellow) the strongest (second strongest) associations. The cosine similarities show that the links in the alternative networks correlate with links in our networks that match the corresponding link type. For example, the equity network

[10]Standard approaches to compute clustered standard errors require a fixed assignment of each observation to a cluster. However, a firm can appear on either side of a link depending on the (arbitrary) order in which the two firms are listed. No fixed firm assignment exists here, so we resort to bootstrap.

extracted from the WRDS subsidiary data shares the strongest cosine similarity with our equity network. We observe similar behavior for the peer network based on TNIC and the production network based on Compustat customer segments. For the production network based on BEA inputs (leadership network based on BoardEx), the matching network constructed with our methodology comes in second (third) in terms of cosine similarity. While the cosine similarities reveal correlations to some degree, they are generally low because there tends to be small overlap between links in the alternative and our networks, which results in a large number of zero entries. For example, there are 5,599 combined active links between our equity network and the one based on WRDS subsidiaries, but the latter network only includes 110 active links. The situation is similar for the production network based on Compustat segments. On the flip side, the peer network based on TNIC includes more than 10,000 active links while our peer network only includes 1,034 active links. This makes the cosine similarity scores noisy as measures of actual network correlations.

The link regressions reveal more precise insights about how the alternative networks correlate with our networks by controlling for coverage variations at the firm level through the log market cap regressor, by relying on clustered standard errors, and by accounting for the covariance structure between our different networks. The regression estimates of Table 10 show that the alternative networks generally correlate positively and in a statistically significant way with the matching networks constructed using our methodology. For the equity network based on WRDS subsidiaries, the peer network based on TNIC, and the production network based on BEA total requirements, our networks of the corresponding type achieve the strongest and most statistically significant positive association. The association between the BEA production network and our production network is statistically weak. However, we find that this is the result of the aggregation of links within a sector. Once we remove self-links in the BEA network (last column of Table 10), the relationship with our production network becomes statistically significant. For the leadership based on the BoardEx data, the association with our leadership network is not as statistically significant as the association with our peer network. However, the magnitude of the regression coefficient for our leadership network is nearly 74% larger while still statistically significant at the 1% confidence level. For the production network based on Compustat segments, the association with our production

network is positive but not statistically significant. We find that this is due to the sparse firm coverage in the Compustat data. Out of the nearly 4,900 combined production links contained in our data and the Compustat data, only 230 are contained in Compustat. Those only cover 224 firms. In Online Appendix H, we show that the association between the Compustat segments production network and our production network becomes statistically significant once we increase the universe of firms we consider (say, moving from the top 1,000 to the top 3,000 firms, which enlarges the universe of firms covered in the Compustat sample to 725).

We note that there are several statistically significant associations across network types. For example, we find that the equity network based on WRDS subsidiaries and the production network based on Compustat segments positively correlate with our credit network. Similarly, the leadership network based on BoardEx positions correlates with our equity network. We show in Online Appendix H that these kinds of off-diagonal associations tend to be spurious, driven by low frequency and low intensity links. The off-diagonal regression coefficients become statistically insignificant for the alternative equity, leadership, and production networks once we exclude noisy links that are identified less than five times over the corresponding sample periods by our methodology. For the peer network based on TNIC, the positive association with our other network becomes statistically weaker while the negative association with our production network becomes stronger when we remove infrequent links identified fewer than five times by our methodology while also excluding links that fall in the bottom quintile based on TNIC link strength. The strong (weak) positive association with our peer (equity) network remain.

All in all, the results of this section demonstrate that our labeled networks genuinely capture the types of firm links they target. Our results provide both external validation by demonstrating correlations with existing networks, as well as internal validation by analyzing the informational content of the associated news sentences. While some of our networks can also be constructed using existing databases, the results show that our approach has important advantages over existing solutions. First, our networks can be constructed at arbitrary frequencies and in real time based on the amount of available news. In contrast, many of the existing network databases are only infrequently updated (TNIC, Compustat segments, and BEA total requirements data) or have

cutoff dates that lie in the past (TNIC and WRDS subsidiaries). Second, our approach is able to identify links among a large universe of firms, covering most of the firms listed in the CRSP and Compustat databases. In contrast, some of the existing databases are sparse and only cover few firms (Compustat segments and WRDS subsidiary data). These insights highlight that our network data could facilitate new empirical research that studies different types of firm links, similar to our study in Section 6. To facilitate easy usage, we have shared our network data at the monthly frequency in our online data repository.

6 Aggregate outcomes

Firm links can facilitate risk diversification by diffusing shocks across firms. However, they can also facilitate the propagation of idiosyncratic shocks across economically connected firms, giving rise to aggregate fluctuations. A large literature studies the microeconomic origins of aggregate fluctuations and proposes competing mechanisms through which economic networks shape aggregate outcomes. Yet empirical evidence on these mechanisms remains mixed, in part because economic networks are difficult to observe at scale and over time. Our setting allows us to jointly examine several economically distinct networks among the same set firms and over the same sample period. We exploit this variation to distinguish among competing mechanisms by examining how different dimensions of network structure predict future aggregate outcomes.

6.1 Empirical framework

Table 1 highlights the competing mechanisms through which firm networks may affect aggregate outcomes. We now use the common measurement framework provided by our news-implied networks to examine how these channels are reflected in the data. We proceed as follows.

We first select a set of macroeconomic and financial variables that will proxy for aggregate outcomes. This set includes:

- BAA and AAA credit spreads as measured by the average Moody's Seasoned corporate bond yield minus the Federal Funds rate,

- The log-return of the CRSP value-weighted total market index,
- The average level of the VIX,
- The log-growth rate of the industrial production index of the Board of Governors of the Federal Reserve System.
- The log-growth rate of consumption as measured by the Real Personal Consumption Expenditures Index of the U.S. Bureau of Economic Analysis.

We measure these variables at the monthly frequency and collect their data from WRDS and FRED. Table 11 provides summary statistics. The credit spreads proxy for credit market outcomes, the VIX and stock return proxy for equity market conditions, while consumption and industrial production proxy for real economic outcomes.

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

- The density, is the average number of active (i.e., non-zero) links by a node in the network. The density captures how many links are present in a network.
- The concentration given by the coefficient of variation of the node degree distribution. Here, the degree of a node is the sum of all the link strengths that the node has. The concentration of a network is large whenever few nodes have many strong links, while the many nodes have only few and weak links.
- The *intermediation intensity* (or, simply, *intermediation*) given by the weighted degree covariation of two firms that are indirectly connected through a third firm. It is large whenever there are many firms acting as intermediaries connecting two strongly connected nodes.

Schwenkler and Zheng (2026) show that these measures applied for a production network built using an alternative approach associates with future output growth, providing a good starting point for our analysis. The metrics also closely relate to alternative network architecture metrics used by

Herskovic (2018).[11]

Figure 11 shows the time series of the connectivity metrics for the different networks, also showcasing the exact formulas for how the metrics are computed. On a rolling monthly basis, we collect all news links identified in the previous 12 months (not including the current month to not introduce forward-looking bias), and compute news-implied networks using the methodology of Section 3.1. We restrict ourselves to the 2,000 largest firms by average market cap over the previous 12 months to ensure a stable firm universe over time, and only consider links that we identified at least 10 times over that sample period to exclude noisy observations. Then, we compute the monthly network metrics for each network. We observe in Figure 11 that the credit network tends to showcase the highest concentration score, consistent with the strong core-periphery structure highlighted in Section 5.3. On average across the sample, the peer network is densest while the production network has the highest intermediation intensity. There appears to be commonality in the density metrics across networks, which we account for in our regressions.

Finally, we run regressions of the cumulative change in the aggregate variables over future periods on our network metrics constructed only using contemporaneously available data.[12] For Month t and Horizon h, we run regressions of the type:

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

Here, $D_{n,t}$, $C_{n,t}$, and $I_{n,t}$ are the monthly density, concentration, and intermediation metrics for Network n measured at the start of Month t using data only through the end of the previous month, X_{t-1} is a vector of controls measured in the previous month, and $Y_{t,h}$ is the cumulative change in the targeted aggregate variable from Month t to Month $t+h-1$. Because the market stock return, consumption, and industrial production are already measured as log-growth rate, their cumulative

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

[12] Although the underlying news articles contain only information publicly available at the time of publication, our classification procedure employs modern language models whose training corpora may include information from later periods (Kong et al. (2026), Levy (2026)). This creates a potential source of look-ahead bias in the econometrician's classification technology. Importantly, the networks themselves are constructed exclusively from contemporaneously published news, and the classifier assigns relationship types based on the textual description of firm links rather than forecasts of subsequent outcomes.

change is simply their cumulative sum. For the VIX and credit spreads, we measure the monthly change in the average monthly values from previous to current month, and cumulate these through the prediction period. We always include as controls the lagged target outcome variable, as well as the lagged number of news articles in our sample to control for the commonality in the link density highlighted in Figure 11. For the market return and VIX, we also include the reciprocal value in the regression to control for the leverage effect (i.e., lagged VIX in the market return regression and lagged market return in the VIX regression; see Black (1976)).

Our network metrics are naturally autocorrelated because they are computed using networks constructed with 12-months of rolling news data. The outcome variables are also autocorrelated because they are computed cumulatively over rolling windows of length h. We therefore adopt a Newey and West (1987) standard error correction with a lag of 12 months or $h-1$, whichever is higher. We run Regression (4) independently for each network type and then compare the estimates of the regression coefficients $\beta_{D,n,h}$, $\beta_{C,n,h}$, and $\beta_{I,n,h}$ across network types. To ensure comparability, we standardize all regressors and outcome variables with their time-series mean and standard deviation before running the regressions. Thus, the regression coefficients in Eq. (4) measure the standard-deviation multiple change in the outcome variable for a one-standard-deviation shock in one of the network metrics.

6.2 Network Structure and Future Aggregate Outcomes

We begin by testing the joint hypothesis that $\beta_{D,n,h} = \beta_{C,n,h} = \beta_{I,n,h} = 0$ for a given network type n and prediction horizon h using Wald tests. To address the multiple-hypothesis testing concern of Harvey et al. (2016), we adjust the resulting p-values across network types using the false-discovery-rate procedure of Benjamini and Hochberg (1995). Table 12 presents our results, focusing on 1-month, 1-year, and 3-year prediction horizons. We present the results for additional prediction horizons in the Online Appendix.

Table 12 reveals three main patterns. First, there is little evidence that network structure predicts aggregate outcomes at the one-month horizon. None of the joint hypotheses are rejected after adjusting for multiple testing. Predictive relationships instead emerge primarily at intermediate

and longer horizons, suggesting that the network measures contain information about relatively persistent changes in financial and economic conditions rather than high-frequency fluctuations.

Second, the predictive relationships exhibit a pronounced economic structure across network types and outcomes. At the one-year horizon, equity network connectivity strongly predicts both future stock-market volatility and returns. Leadership and peer networks also predict stock returns, while leadership, peer, and production networks contain significant information about future consumption growth. These patterns broadly correspond to the economic domains represented by the underlying networks: financing and information networks are particularly informative about future financial-market conditions, whereas information and production networks contain information about subsequent real activity.

This structure becomes particularly pronounced at the three-year horizon. Credit network connectivity predicts both AAA and BAA credit spreads and is strongly associated with future stock-market volatility. In contrast, the other network types exhibit little evidence of predicting future credit spreads. All five networks contain information about future industrial production, with particularly strong predictability associated with the credit, equity, and leadership networks. The network measures also jointly predict future consumption growth across all five network types, although the statistical evidence is weaker than for industrial production. Stock-market outcomes occupy an intermediate position: several networks predict future volatility and returns, but the strongest relationship is between the credit network and future VIX.

Taken together, these findings suggest that the predictive content of network structure is neither confined to a particular network nor common across all outcomes. Instead, different economic networks contain distinct information about subsequent aggregate conditions, and this information becomes substantially more pronounced at longer horizons. The absence of widespread short-horizon predictability also makes it less likely that the results simply reflect contemporaneous news (which we control through lagged news intensity) or high-frequency movements in financial conditions. Next, we exploit the separate coefficients on density, concentration, and intermediation to examine which dimensions of network architecture account for these predictive relationships.

6.3 The Topology of Predictive Networks

We examine which dimensions of network topology drive the predictive relationships documented above. This decomposition also allows us to return to the competing economic channels summarized in Table 1. In particular, greater connectivity may facilitate beneficial interactions such as information diffusion, resource sharing, and financial intermediation, but may also increase firms' exposure to common shocks and create additional channels for shock propagation. Density, concentration, and intermediation capture distinct dimensions of this architecture and therefore allow us to examine whether their predictive implications differ systematically. Figures 12–14 report the coefficients on density, concentration, and intermediation from Regression (4) for prediction horizons of one and three years. The figures report 95% confidence intervals based on Newey-West standard errors.

Several patterns emerge. First, network density is predominantly negatively associated with subsequent real activity and stock market performance. Across the leadership, peer, and production networks, greater density predicts lower future stock returns, industrial production, and consumption. These relationships are already economically sizable at the one-year horizon and generally remain large at the three-year horizon. For example, at the three-year horizon, a one-standard-deviation increase in leadership-network density predicts approximately a 0.45 standard-deviation decline in industrial production and a 0.5 standard-deviation decline in consumption. The corresponding coefficients for the peer and production networks are similarly negative. Thus, greater connectivity among firms tends to precede weaker aggregate real outcomes.

Second, network concentration exhibits a similar pattern. Greater concentration generally predicts lower future stock returns, industrial production, and consumption, with particularly pronounced effects in the leadership and peer networks. At the three-year horizon, a one-standard-deviation increase in leadership-network concentration predicts approximately a 0.7 standard-deviation decline in industrial production and a 0.55 standard-deviation decline in consumption. Peer-network concentration exhibits similarly large negative coefficients. Production-network concentration also predicts lower subsequent real activity and stock returns. These findings suggest that the distribution of connections across firms contains information beyond the overall number

of connections: aggregate outcomes tend to be weaker when economic relationships become more concentrated among a subset of firms.

These predominantly negative relationships are consistent with several of the adverse channels summarized in Table 1. Denser economic networks provide additional pathways through which firm-level disturbances can propagate, as emphasized for credit and production networks by Eisenberg and Noe (2001) and Barrot and Sauvagnat (2016). Concentration may further increase aggregate exposure when economically important relationships become concentrated around a smaller set of firms. The particularly strong negative coefficients for leadership and peer concentration are also consistent with the possibility that concentrated information networks facilitate correlated behavior or herding (Chiu et al. (2013), Ferris et al. (2017), Shue (2013)) and with evidence that greater peer concentration can reduce aggregate output (Pellegrino (2025)). Our evidence does not distinguish among these mechanisms causally, but it indicates that the adverse predictive content of interconnectedness is concentrated in precisely those topological dimensions that increase direct and concentrated exposure among firms.

The results for intermediation are markedly different. Within most networks, the estimated coefficient on intermediation lies substantially to the right of the corresponding coefficients on density and concentration. In several cases, the sign reverses altogether. The distinction is particularly apparent for the equity network. Greater equity-network intermediation predicts higher subsequent stock returns and consumption at both the one- and three-year horizons, as well as higher industrial production at the three-year horizon. At the three-year horizon, a one-standard-deviation increase in equity intermediation predicts increases of approximately 0.15 standard deviations in industrial production and 0.25 standard deviations in consumption. Hence, greater connectivity and greater intermediation need not have the same implications for subsequent aggregate activity.

This pattern is more consistent with the beneficial channels in Table 1. In the equity network, intermediation may facilitate the diffusion of information and governance practices across otherwise distinct groups of firms, consistent with Gao et al. (2017) and Edmans et al. (2018). More generally, intermediating firms may connect economically distinct parts of the network without requiring every firm to become directly exposed to every other firm. This distinction provides a natural

interpretation of why intermediation can predict stronger subsequent outcomes even when density and concentration predict weaker ones: the benefits of economic connectivity may depend on how links are organized rather than simply on how many links exist.

A similar distinction emerges for credit market conditions. In the credit network, greater intermediation predicts a pronounced decline in subsequent market volatility. At the three-year horizon, credit-network intermediation is associated with a decline of approximately 0.5 standard deviations in the VIX and with narrower subsequent AAA and BAA credit spreads. This pattern is consistent with the beneficial role of financial intermediation and monitoring emphasized by Greenwood and Jovanovic (1990) and Sufi (2007), and stands in contrast to the contagion and amplification channels associated with direct credit exposures (Eisenberg and Noe (2001), Bernanke et al. (1999), and others).

The figures also reveal systematic heterogeneity across network types. Leadership and peer networks are particularly informative about future real activity. Their density and concentration measures consistently predict lower subsequent industrial production and consumption, and also tend to predict lower stock market returns. Production-network density and concentration exhibit qualitatively similar relationships.[13] The equity network, by contrast, stands out for the positive predictive relationship between intermediation and subsequent aggregate outcomes. Finally, the topology of the credit network appears particularly informative about future financial conditions, especially at longer horizons.

Taken together, these results show that characterizing economic networks solely by their overall degree of connectivity can obscure economically important differences in network topology. Density and concentration generally predict weaker subsequent aggregate outcomes, whereas intermediation has substantially different, and in several cases opposite, predictive implications. Greater direct connectivity therefore need not have the same aggregate implications as an architecture in which firms intermediate relationships between otherwise distinct parts of the network.

This distinction helps reconcile the competing mechanisms summarized in Table 1. Much of

[13]This finding closely parallels Schwenkler and Zheng (2026), who document a particularly strong negative relationship between production-network connectivity and subsequent real activity using production networks constructed from a different data source and methodology. Recovering the same predictive relationship from our independently constructed news-implied production network provides useful external validation of our network measures.

the existing literature documents either beneficial effects of economic links through information diffusion, resource sharing, monitoring, and financial intermediation, or adverse effects through contagion, herding, competition, and shock propagation. Our results suggest that these mechanisms need not be contradictory. Their relative importance may depend on the topology through which economic relationships are organized. The predominantly negative predictive relationships associated with density and concentration are consistent with architectures that create greater direct or concentrated exposure to shocks, while the more favorable outcomes associated with intermediation are consistent with networks that connect otherwise distinct parts of the economy and thereby facilitate beneficial economic interactions. Although our predictive evidence does not identify these mechanisms causally, it provides a unified empirical result: the aggregate implications of economic interconnectedness depend not only on the nature of economic links, but critically on how those links are organized.

7 Conclusion

This paper develops a common empirical framework for measuring several economically distinct networks of firms from financial news. Using more than four decades of news coverage, we construct time-varying credit, equity, leadership, peer, and production networks. The resulting networks exhibit economically interpretable architectures and align with independently constructed network measures from existing databases. Our approach therefore provides a scalable way to observe multiple forms of firm connectivity within the same universe of firms and at high frequency.

We use these networks to study how firm connectivity relates to subsequent aggregate financial and economic conditions. Network topology contains substantial predictive information, particularly at longer horizons, but its implications depend on both the economic nature of the underlying links and the way those links are organized. Density and concentration predominantly predict weaker subsequent real activity and stock-market performance, whereas intermediation often has different and sometimes opposite implications. In particular, credit-network intermediation predicts narrower credit spreads and lower market volatility, while equity-network intermediation predicts stronger subsequent real activity.

These findings suggest that apparently competing views of economic interconnectedness need not be contradictory. Firm links may simultaneously facilitate information transmission, monitoring, specialization, and intermediation while creating channels for contagion and shock propagation. Which force dominates appears to depend critically on network topology. Beyond the application studied here, the news-implied networks and methodology developed in this paper provide a new empirical foundation for studying how different forms of firm connectivity shape financial markets and the aggregate economy.

References

Abbassi, Puriya, Christian Brownlees, Christina Hans and Natalia Podlich (2017), 'Credit risk interconnectedness: What does the market really know?', *Journal of Financial Stability* **29**, 1–12.

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.

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

Azizpour, S., K. Giesecke and G. Schwenkler (2018), 'Exploring the sources of default clustering', *Journal of Financial Economics* **129**(1), 154–183.

Babus, Ana and Tai-Wei Hu (2017), 'Endogenous intermediation in over-the-counter markets', *Journal of Financial Economics* **125**(1), 200–215.

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.

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.

Bernanke, Ben S., Mark Gertler and Simon Gilchrist (1999), The financial accelerator in a quantitative business cycle framework, *in* 'Handbook of Macroeconomics', Elsevier, Amsterdam, pp. 1341–1393.

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.

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

Carvalho, Vasco (2010), Aggregate fluctuations and the network structure of intersectoral trade, Technical Report 1206, Department of Economics and Business, Universitat Pompeu Fabra.

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.

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.

Duan, Jin-Chuan, Jie Sun and Tao Wang (2012), 'Multiperiod corporate default prediction—a forward intensity approach', *Journal of Econometrics* **170**(1), 191–209.

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.

Ederer, Florian and Bruno Pellegrino (2024), A tale of two networks: Common ownership and product market rivalry. Working Paper.

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

Eisenberg, Larry and Thomas H. Noe (2001), 'Systemic risk in financial systems', *Management Science* **47**(2), 236–249.

El Mrabet, Mohammed Amine, Khalid El Makkaoui and Ahmed Faize (2021), Supervised machine learning: A survey, *in* '2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet)', pp. 1–10.

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.

Ferris, Stephen P., David Javakhadze and Tijana Rajkovic (2017), 'Ceo social capital, risk-taking and corporate policies', *Journal of Corporate Finance* **47**, 46–71.

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

Gabaix, Xavier (2011), 'The granular origins of aggregate fluctuations', *Econometrica* **79**(3), 733–772.

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.

García, Diego (2018), The kinks of financial journalism. Working Paper, CU Boulder.

Ge, Shuyi, Shaoran Li and Oliver Linton (2023), 'News-implied linkages and local dependency in the equity market', *Journal of Econometrics* **235**(2), 779–815.

Giroud, Xavier and Holger M. Mueller (2019), 'Firms' internal networks and local economic shocks', *American Economic Review* **109**(10), 3617–49.

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

Greenwood, Jeremy and Boyan Jovanovic (1990), 'Financial development, growth, and the distribution of income', *Journal of Political Economy* **98**(5), 1076–1107.

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.

Hautsch, Nikolaus, Julia Schaumburg and Melanie Schienle (2014), 'Financial Network Systemic Risk Contributions*', *Review of Finance* **19**(2), 685–738.

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

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.

Lee, Charles M.C., Paul Ma and Charles C.Y. Wang (2015), 'Search-based peer firms: Aggregating investor perceptions through internet co-searches', *Journal of Financial Economics* **116**(2), 410 – 431.

Lee, Charles M.C., Stephen Teng Sun, Rongfei Wang and Ran Zhang (2019), 'Technological links and predictable returns', *Journal of Financial Economics* **132**(3), 76–96.

Levy, Bradford (2026), 'Caution ahead: Numerical reasoning and look-ahead bias in ai models', *Journal of Accounting Research* **64**(3), 1139–1188.

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.

Mcdonald, Michael L., Poonam Khanna and James D. Westphal (2008), 'Getting them to think outside the circle: Corporate governance, ceos' external advice networks, and firm performance', *Academy of Management Journal* **51**(3), 453–475.

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

Newey, Whitney K. and Kenneth D. West (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation', *Econometrica* **55**(3), 703–708.

Nimark, Kristoffer P. and Stefan Pitschner (2019), 'News media and delegated information choice', *Journal of Economic Theory* **181**, 160 – 196.

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

Petersen, Mitchell A. and Raghuram G. Rajan (1997), 'Trade credit: Theories and evidence', *The Review of Financial Studies* **10**(3), 661–691.

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), Why does news coverage predict returns? Working Paper.

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

Shobha, Gangadhar and Shanta Rangaswamy (2018), Chapter 8 - machine learning, *in* V. N.Gudivada and C.Rao, eds, 'Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications', Vol. 38 of *Handbook of Statistics*, Elsevier, pp. 197–228.

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.

Solomon, David H. and Eugene F. Soltes (2012), Managerial control of business press coverage. Working Paper.

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

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

Firm network	Beneficial channels	Adverse channels
Credit	Shared credit links incentivize stronger monitoring of risks (Sufi (2007)). Financial intermediation enables growth (Greenwood and Jovanovic (1990)).	Credit links enable contagion (Eisenberg and Noe (2001)). credit-market frictions can amplify shocks through the financial accelerator (Bernanke et al. (1999)).
Equity	Productivity gains through sharing of resources (Giroud and Mueller (2019)). Common ownership may facilitate information diffusion (Gao et al. (2017)) and strengthen governance (Edmans et al. (2018)).	Common ownership may weaken competition (Azar et al. (2018)) and result in welfare losses (Ederer and Pellegrino (2024)). It may also facilitate contagion through portfolio rebalancing (Manconi et al. (2012)).
Leadership	Information diffusion through managerial networks can improve corporate decision-making and firm performance (Fracassi (2017), Mcdonald et al. (2008), Sandvik et al. (2020)).	Common information in leadership networks can lead to herding (Chiu et al. (2013)), possibly exacerbating risks (Ferris et al. (2017)) and yielding 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 compress margins to withstand shocks (Valta (2012)). Information contagion may spread credit risk (Duffie et al. (2009)). Peer concentration may result in lower output (Pellegrino (2025)).
Production	Production links allow firms to access and outsource to specialized suppliers, generating productivity gains (Bernard et al. (2019)) and facilitate financing through trade credit (Petersen and Rajan (1997)).	Production links propagate idiosyncratic shocks (Barrot and Sauvagnat (2016)) and transmit information into asset prices (Cohen and Frazzini (2008)). Congestions in production networks may result in slower growth (Schwenkler and Zheng (2026)).

Table 1: *Economic channels linking firm-level networks to aggregate outcomes.* The list of channels in this table is not necessarily complete. For a larger survey, please see Gofman et al. (2026).

	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.

Outlet	Articles	Start	End
New York Times	466,253	1981-01	2025-12
Forbes	127,917	2010-08	2025-12
CNBC	76,657	2010-01	2025-12
Business Insider	75,085	2010-01	2025-12
MarketWatch	63,117	2010-01	2025-03
Wall Street Journal	21,272	2010-01	2020-10
Reuters	8,531	2010-01	2025-07
Bloomberg	5,140	2010-06	2025-11
Fortune	3,002	2011-08	2025-09
Barron's	2,105	2010-06	2020-10
Total	849,079	1981-01	2025-12

Table 3: *Article counts by news outlet.* We consider New York Times articles collected through the NYT Developer API and online news articles from different outlets available in the NOW Corpus at corpusdata.org.

"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 4: *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 5: *Out-of-sample accuracy of Stage 1 model.* 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.

Epochs:	25	50	75	100	125	150	175	200	225	250
Credit	0.6772	0.6935	0.7395	0.7748	0.7523	0.7523	0.7593	0.7593	0.7477	0.7238
Equity	0.9525	0.9615	0.9709	0.9804	0.9804	0.9804	0.9804	0.9804	0.9804	0.9615
Leadership	0.9703	0.9800	0.9800	0.9800	0.9703	0.9703	0.9703	0.9703	0.9703	0.9703
Peer	0.7660	0.7872	0.7957	0.8125	0.7917	0.8041	0.8163	0.8041	0.8163	0.8041
Production	0.7609	0.7872	0.7917	0.7755	0.7755	0.7755	0.7755	0.7755	0.7677	0.7677
Other	0.6914	0.6988	0.6742	0.7097	0.7021	0.7097	0.7097	0.7021	0.7097	0.7021
Overall	0.8030	0.8167	0.8253	0.8388	0.8287	0.8320	0.8352	0.8320	0.8320	0.8216

Table 6: *Out-of-sample F1 score of Stage 2 model over different epochs.* We update the model of Stage 1 using back propagation with a smaller learning rate over an increasing number of epochs to obtain the Stage 2 model. We then evaluate the out-of-sample F1 score of the resulting model on a test set containing 300 human-verified sentences describing relationships between two firms and published in the NOW corpus. See Table 5 for a definition of the F1 score.

		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 7: *Out-of-sample accuracy of Stage 2 model.* We evaluate the accuracy of the model of Stage 2 fine-tuned over 100 epochs using the same synthetic and human-verified news sentences as in Table 5. 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 8: *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	Barns & 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 9: *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	0.0764	0.1027	0.0641	0.2055	0.2276	0.2924
	Equity	0.1558	0.1375	0.0977	0.1947	0.2656	0.3040
	Leadership	0.0546	0.2196	0.0926	0.2444	0.3239	0.4021
	Peer	0.0700	0.2327	0.1537	0.3017	0.4881	0.4572
	Production	0.0704	0.2130	0.1075	0.4263	0.4489	0.5090
	Other	0.0739	0.2301	0.1507	0.3228	0.4423	0.4940
Link regressions:							
Our networks	Credit	* 1.9732	0.4468	0.2581	** 0.3893	0.9220	2.1109
		(2.1620)	(0.9250)	(0.9608)	(2.6666)	(0.4739)	(1.2632)
	Equity	*** 2.6791	0.2642	° 0.3968	-0.1723	-0.1323	0.2092
		(5.0641)	(1.1422)	(1.6561)	(-0.2764)	(-0.2732)	(0.4409)
	Leadership	0.8191	** 0.8792	0.2395	0.2008	-0.2044	0.4451
		(1.0970)	(2.8643)	(1.1845)	(0.3995)	(-0.5676)	(1.1872)
	Peer	° 1.0487	** 0.5043	*** 0.5849	-0.0071	1.4965	0.4920
		(1.7686)	(2.9327)	(6.8250)	(-0.0563)	(1.4030)	(0.5143)
	Production	0.9882	0.3270	-0.0468	0.0472	° 0.5588	* 1.0301
		(1.5876)	(1.2922)	(-0.4063)	(0.6059)	(2.0423)	(2.4881)
	Other	1.0188	° 0.3115	** 0.4528	0.0982	0.2194	0.5495
		(1.3885)	(1.9398)	(3.1466)	(0.4108)	(0.5190)	(1.7045)
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		5,599	5,134	13,411	4,889	276	253
Total linked nodes		801	802	926	817	23	23
Included active links		3,132	4,176	10,204	230	276	253
Included linked nodes		665	765	818	224	23	23
Active links (alt. net.)		110	855	10,204	230	276	253
Active links (our net.)		3,115	4,069	1,034	65	170	155

Table 10: *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 (or sector) level via a bootstrap: we resample nodes with replacement, reconstruct all sampled pairs (dropping self-pairs), and re-estimate the regression; the standard error is the standard deviation of the resulting coefficients across many 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.
Monthly stock market log-return (%)	1981-01	2025-12	0.77	1.18	4.52	-25.98	12.15
Average monthly VIX	1990-01	2025-12	19.46	17.67	7.42	10.13	62.67
Average monthly AAA credit spread	1981-01	2025-12	2.52	2.54	1.69	-6.27	5.73
Average monthly BAA credit spread	1986-01	2025-12	3.59	3.50	1.68	0.27	8.82
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 11: *Summary statistics of our aggregate variables.* All factors are sampled at the monthly frequency.

	Corporate credit market		Stock market		Output	
	AAA spread	BAA spread	VIX	Return	Ind. prod.	Consumption
Short prediction horizon (1 month):						
Credit	2.0876	1.8076	0.4742	0.5585	0.3345	1.3684
Equity	1.0161	1.0829	0.1342	0.2038	0.2364	0.760
Leadership	1.2947	1.6991	0.7190	0.9241	1.0108	1.5509
Peer	0.9688	0.7643	0.6603	0.6134	1.9946	2.5998
Production	1.3717	2.0697	0.5559	0.7897	0.8288	2.2452
Intermediate prediction horizon (1 year):						
Credit	1.0601	0.6840	2.3082	2.0625	0.6995	* 2.9529
Equity	1.3627	0.8364	*** 6.8545	*** 6.8649	0.9612	° 2.2283
Leadership	1.2304	0.5023	1.2544	* 3.5813	° 3.5664	** 5.4014
Peer	1.0079	0.2370	1.5687	** 5.2685	2.3099	* 3.8261
Production	1.8559	0.6382	1.3436	* 3.3910	1.8906	* 4.0926
Long prediction horizon (3 years):						
Credit	° 3.5035	* 4.2777	*** 10.7631	0.3937	*** 13.4521	* 3.4703
Equity	1.3165	1.3403	** 4.9191	° 2.8450	*** 16.0045	* 4.7528
Leadership	0.8603	0.7832	* 2.9555	° 3.1392	*** 11.5798	* 4.1642
Peer	1.6622	1.5128	* 2.8854	° 2.8605	*** 10.6035	* 2.9354
Production	0.9491	0.8613	° 2.3450	° 2.5545	*** 9.6533	* 3.2314

Table 12: *Wald tests of the joint hypothesis that network metrics do not predict aggregate outcomes.* For each network type and prediction horizon, we test the null hypothesis $\beta_{D,n,h} = \beta_{C,n,h} = \beta_{I,n,h} = 0$ in Regression (4). Reported values are Wald test statistics. We adjust the corresponding p-values across network types within each outcome and prediction horizon 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.

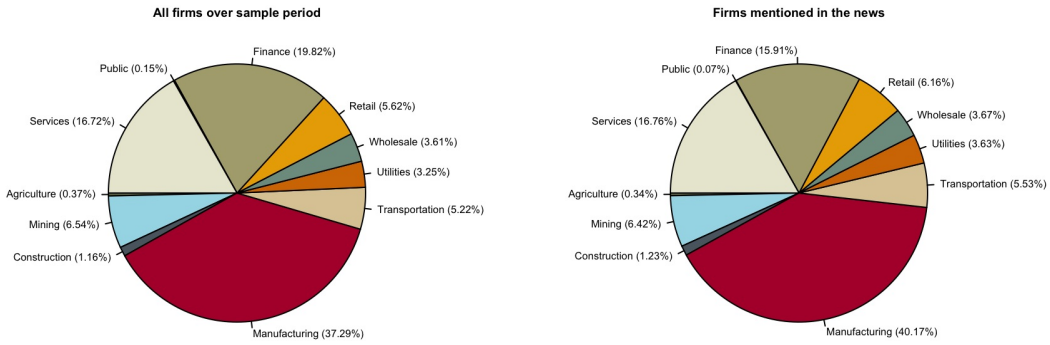


Figure 1: *Industry distribution in population and news-mentioned sample.* The population includes 27,103 distinct firms and the sample of firms mentioned in the news covers 16,387 unique firms. The data cover the time period between January 1, 1981, and December 31, 2025.

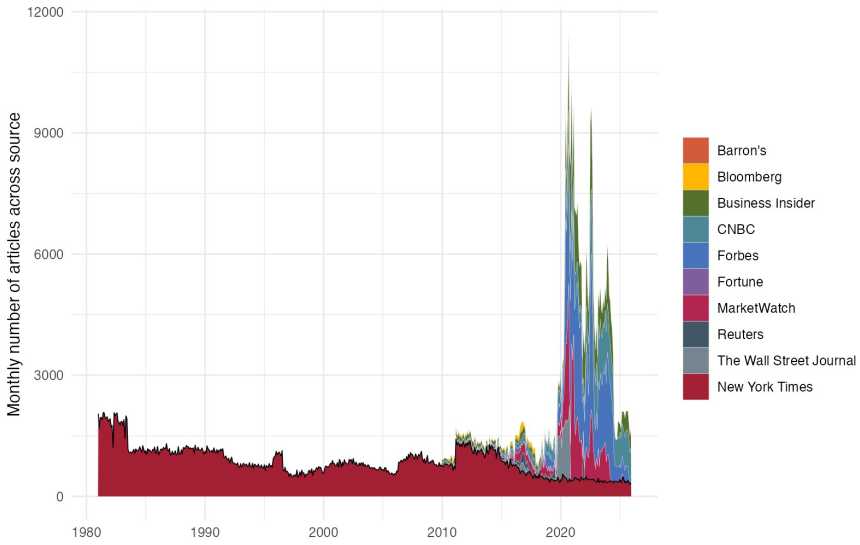


Figure 2: *Monthly number of articles.* The plot shows the monthly number of articles contained in our news sample across the different news outlet sources. The black line separates data collected from the New York Times Archive from those obtained from the NOW Corpus.

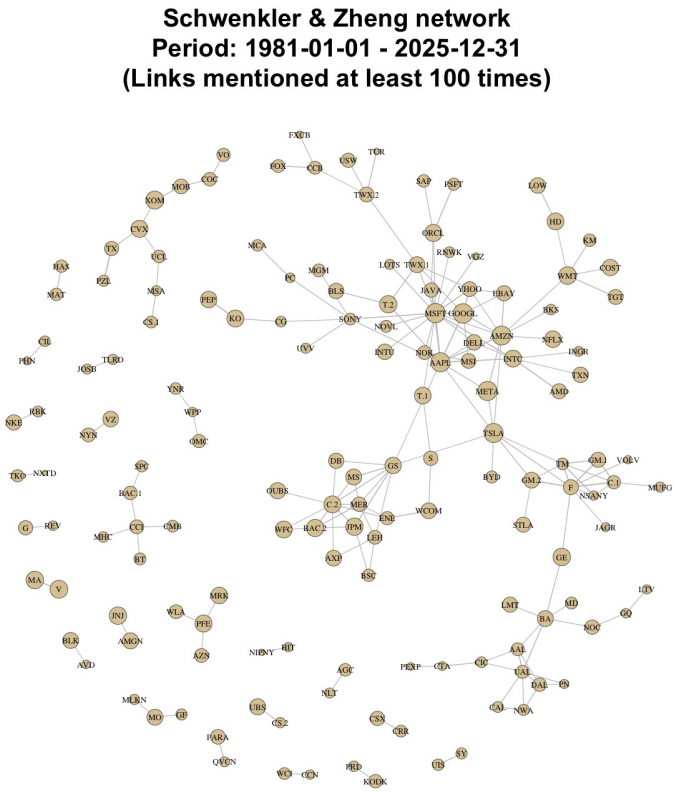


Figure 3: *News-implied network in our sample.* This graph shows the network of sentence-based firm co-mentions as proposed by Schwenkler and Zheng (2026). We establish a link whenever the two firms are mentioned in the same sentence of an article. We only consider sentences in which we identify only two distinct firm mentions. The size of a node is proportional to the logarithm of one plus the average market cap of the firm over its lifetime that overlaps with our sample. The width of a link is proportional to the logarithm of one plus the number of link mentions in the news sample. We use the Fruchterman-Reingold algorithm to plot the network.

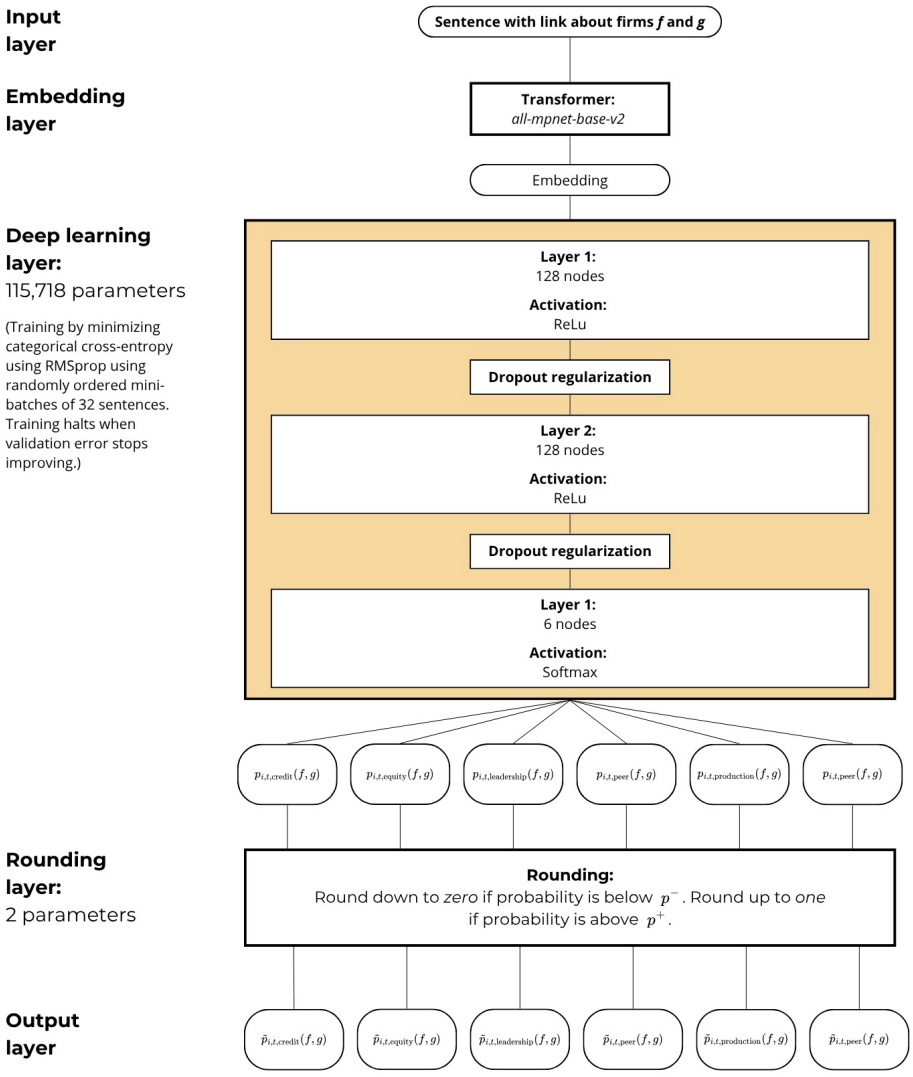


Figure 4: *Model architecture.* This figure shows the architecture of the model underlying our link-labeling methodology.

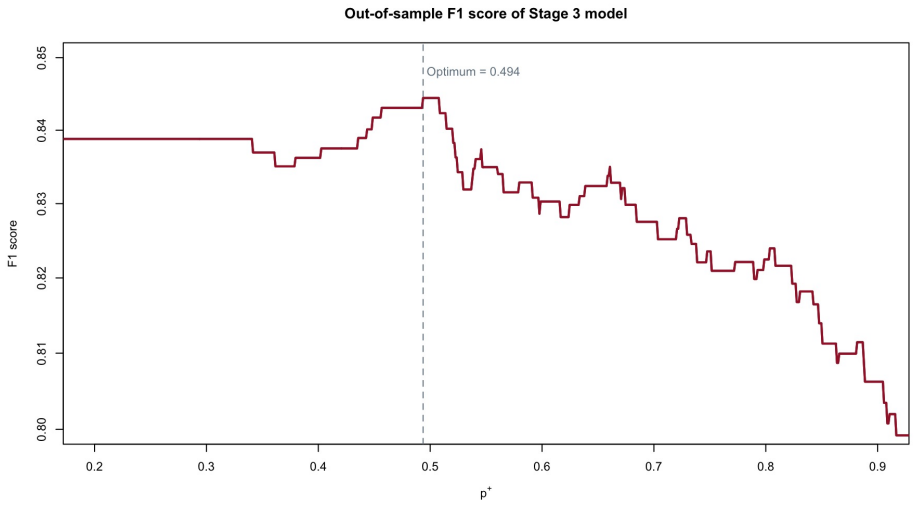
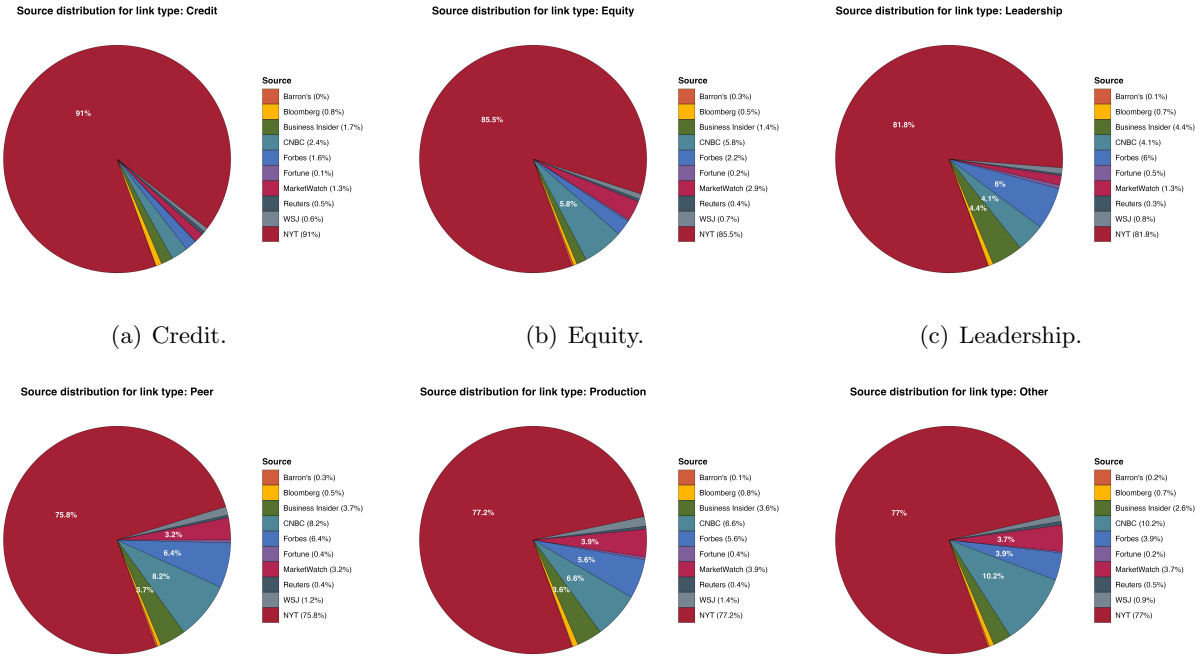


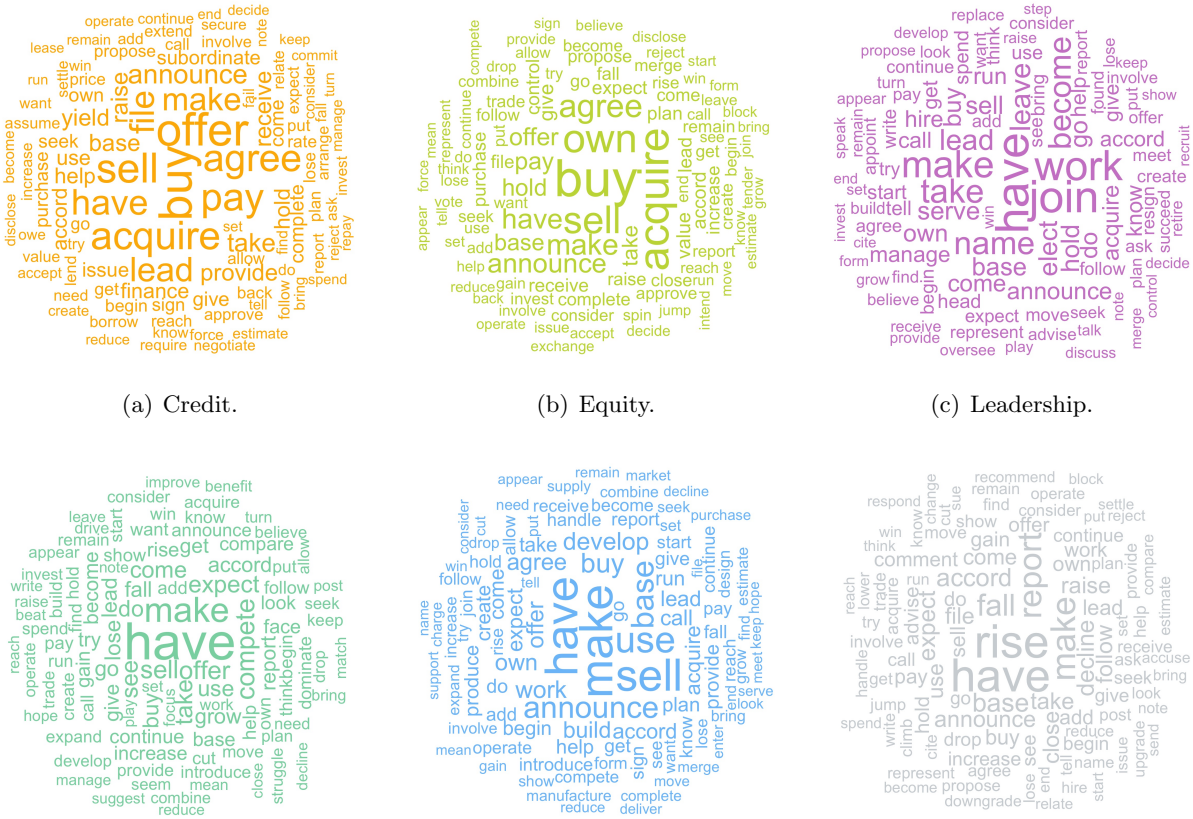
Figure 5: *F1 score against round-up threshold* p^+. This figure plots the out-of-sample F1 score of the Stage 3 model with rounding as in Eq. (2). We fix $p^- = 1/6$ and evaluate the F1 score on the human-verified test set of Stage 2 using different values of p^+.



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

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

Figure 6: *News sources of strong sentences.* These graphs show the distribution of news sources that published sentences for which our methodology assigned link probabilities above the round-up threshold p^+ (i.e., the *strong sentences*), broken down by link type. The pie charts show the proportion of sentences associated with a news source whenever the proportion is at least 3%. Otherwise, the proportions are displayed in the legends of the graphs.



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

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

Figure 7: *Most frequent verbs.* These word clouds display the 100 most frequent verbs 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 verbs prior to displaying them in a word cloud. We exclude the verbs "be," "include," "say," and "share."



(a) Credit.



(b) Equity.



(c) Leadership.



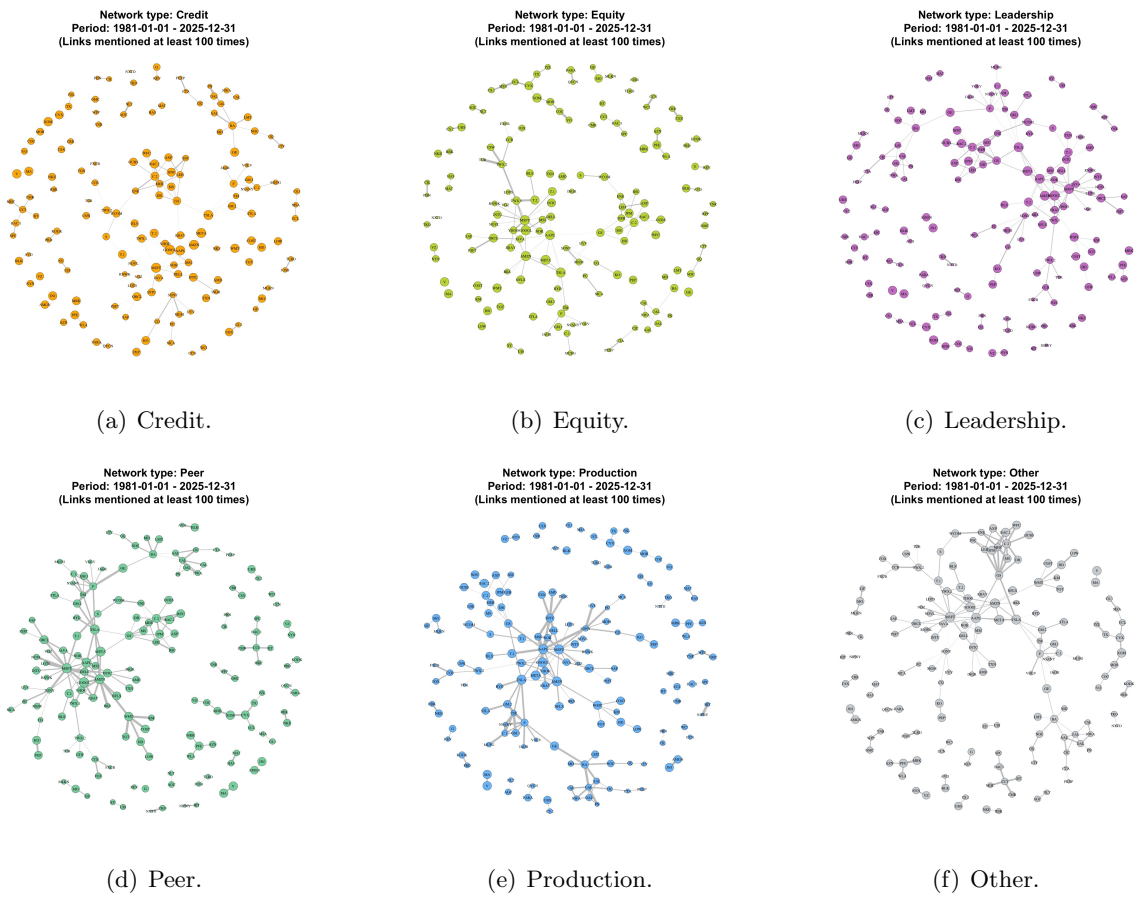
(d) Peer.



(e) Production.

(f) Other.

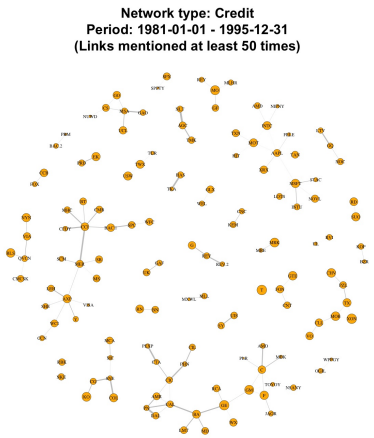
Figure 8: *Most frequent nouns.* These word clouds display the 100 most frequent nouns in strong sentences (i.e., those for which our methodology assigns a link probability above the round-up threshold p^+). We use the universal point-of-speech tagger from the cleanNLP package available in R. We restrict 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."



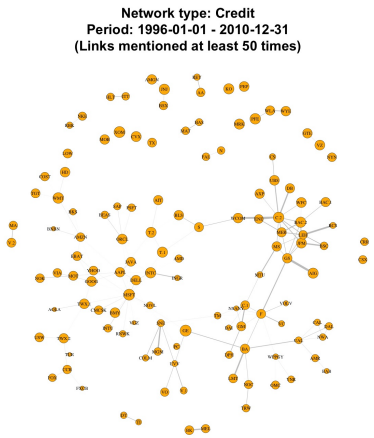
(a) Credit. (b) Equity. (c) Leadership.

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

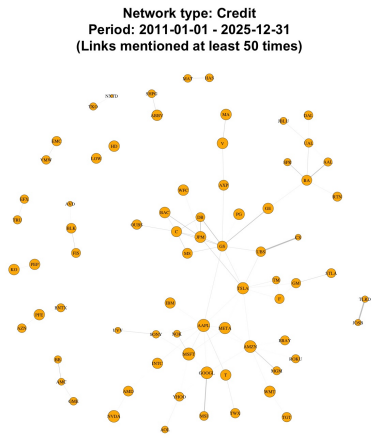
Figure 9: *Aggregate news-implied networks.* These graphs show the different news-implied networks constructed using our methodology, aggregated over the sample period. We only include links that were identified at least 100 times over the full sample. The size of a node is proportional to the logarithm of 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.



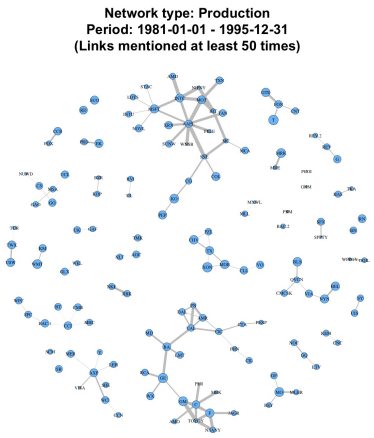
(a) Credit (1981-1995).



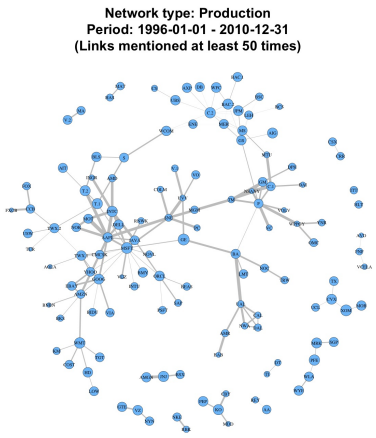
(b) Credit (1996-2010).



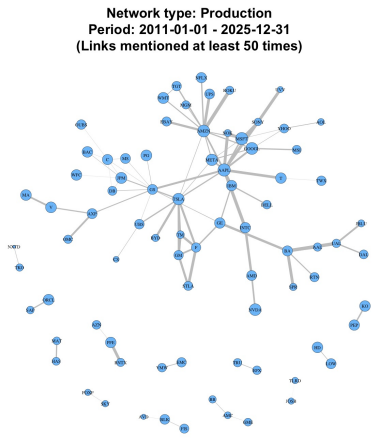
(c) Credit (2011-2025).



(d) Production (1981-1995).

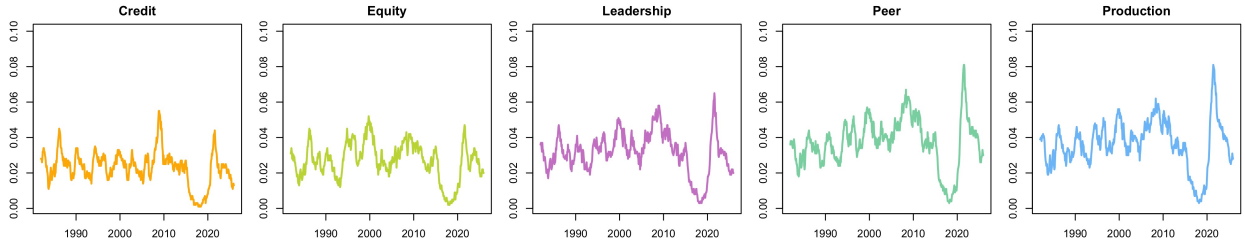


(e) Production (1996-2010).

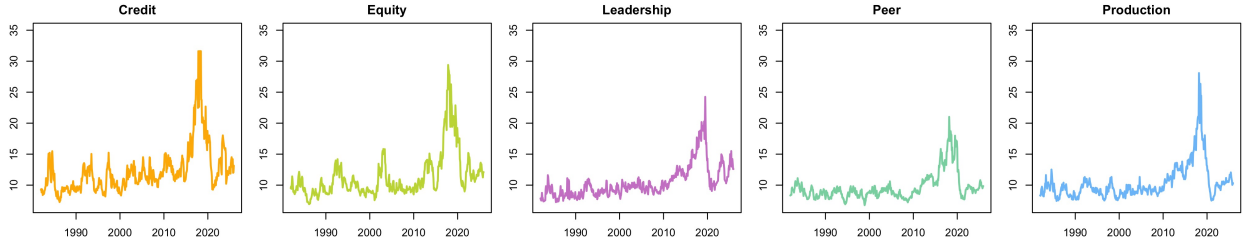


(f) Production (2011-2025).

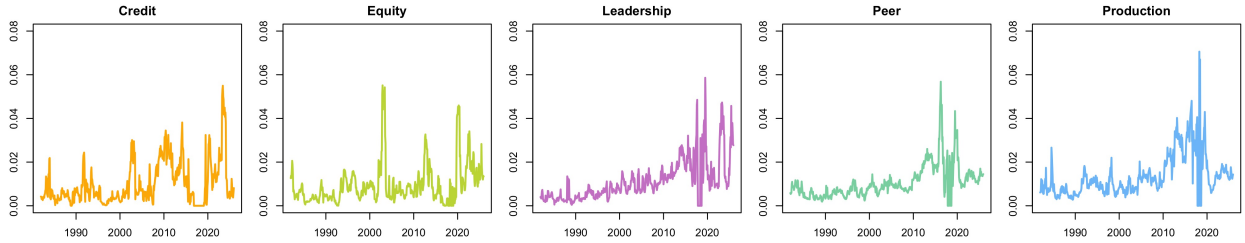
Figure 10: *Time series of news-implied networks.* These graphs show the different news-implied networks constructed using our methodology, aggregated over subsequent 15-year periods. We only include links that were identified at least 50 times over the respective sub-samples. The size of a node is proportional 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.

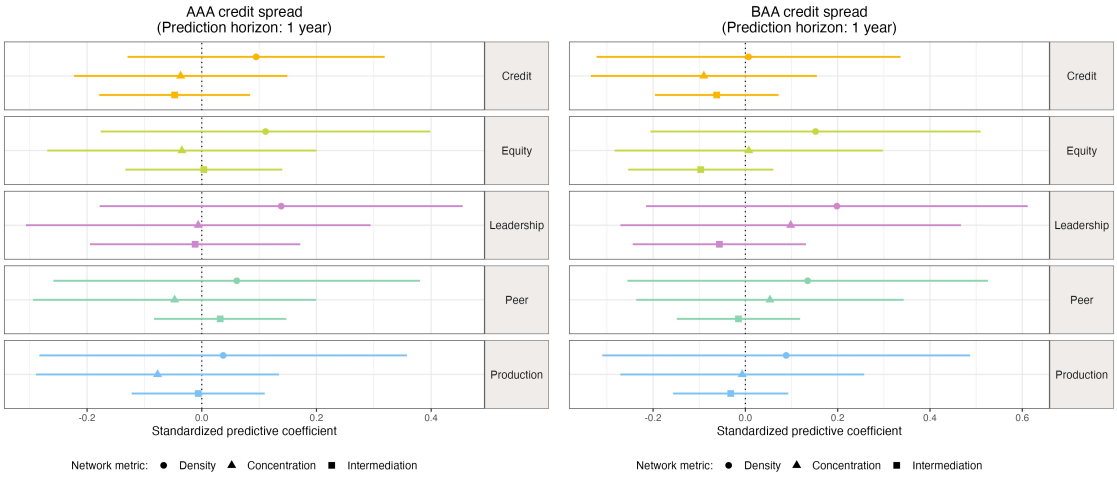


(b) Concentration.

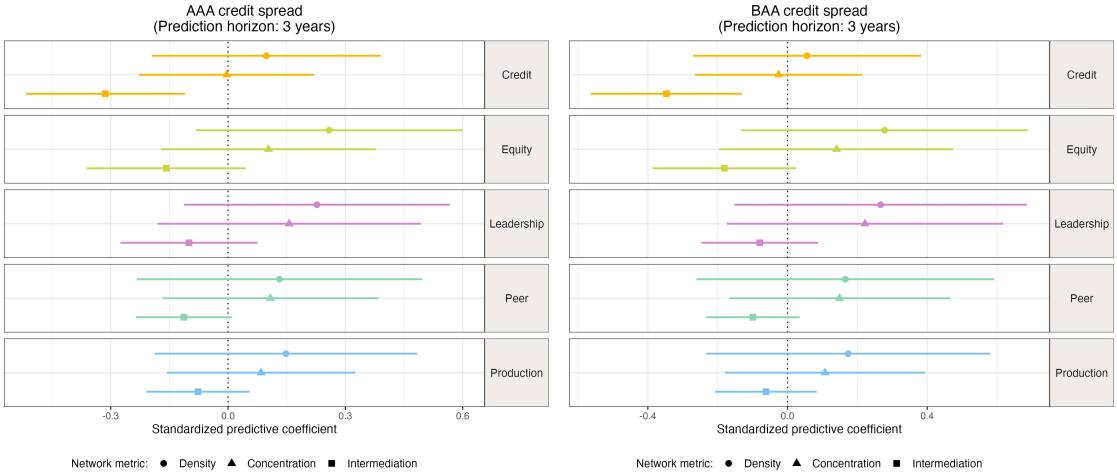


(c) Intermediation.

Figure 11: *Time series of connectivity metrics.* Given a monthly network, we follow Schwenkler and Zheng (2026) and evaluate the 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 Month t. Here, $w_t^{j,n}$ is the number of identified links that connect Nodes j and n in 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).

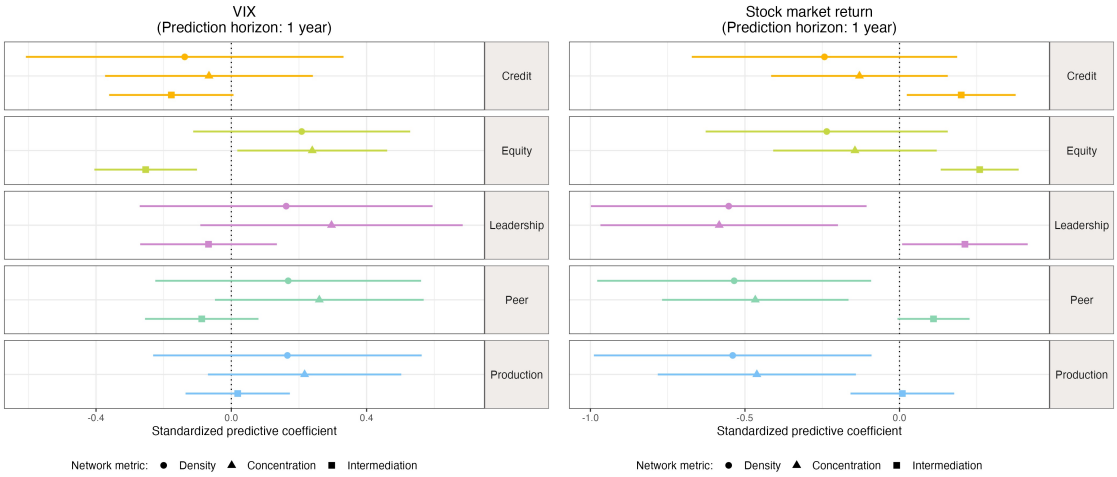


(a) Intermediate prediction horizon (1 year).

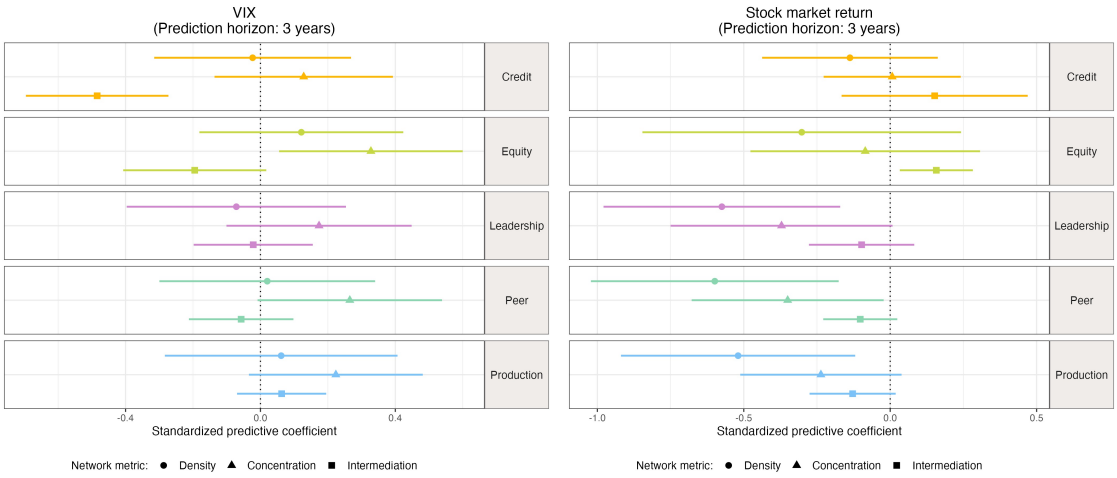


(b) Long prediction horizon (3 years).

Figure 12: *Standardized regression coefficients for credit market variables.* These figures show the regression point estimates of Eq. (4) through shapes, with 95% asymptotic confidence bands based on Newey-West-adjusted standard errors through horizontal lines. Because both the dependent variables and network metrics are standardized, the coefficients measure the standard-deviation change in the future aggregate outcome associated with a one-standard-deviation increase in the corresponding network metric, conditional on the other network metrics and controls.

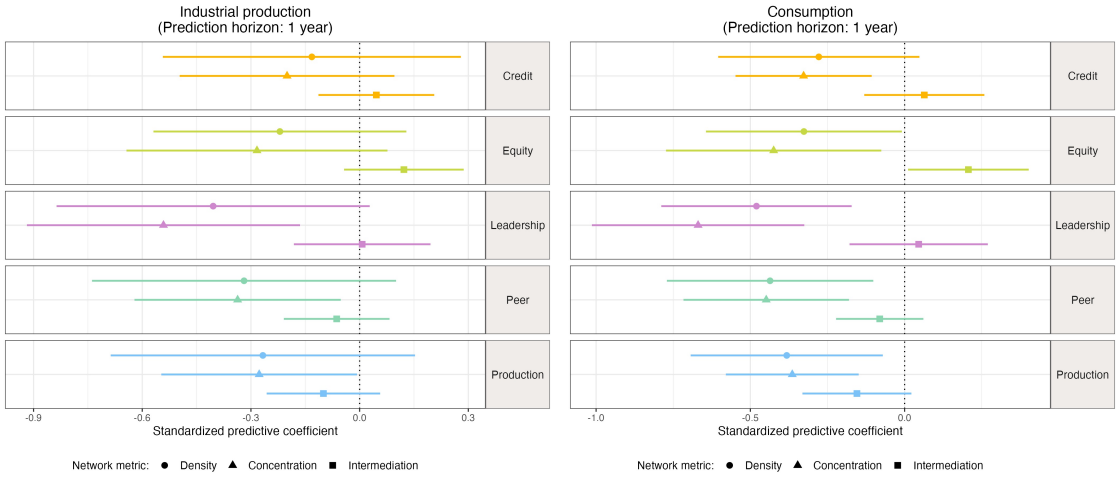


(a) Intermediate prediction horizon (1 year).

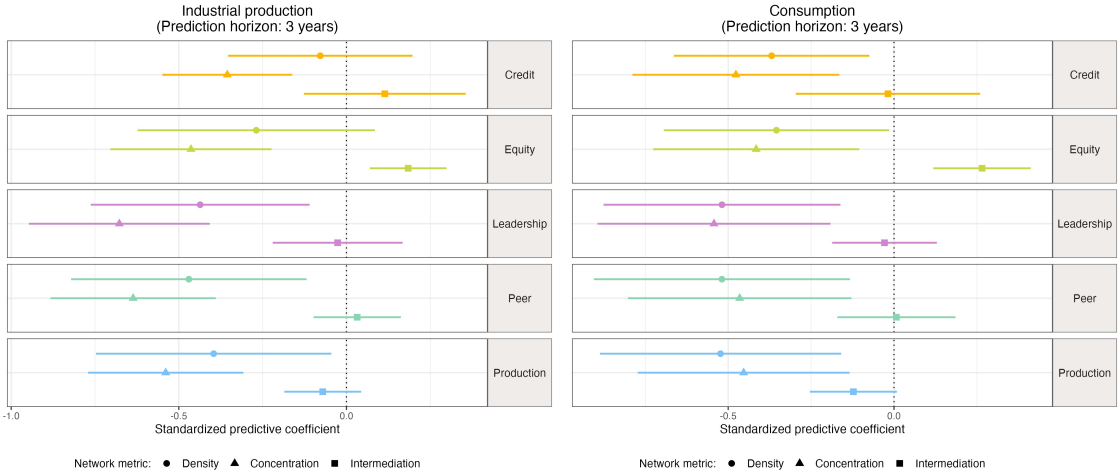


(b) Long prediction horizon (3 years).

Figure 13: *Standardized regression coefficients for stock market variables.* These figures show the regression point estimates of Eq. (4) through shapes, with 95% asymptotic confidence bands based on Newey-West-adjusted standard errors through horizontal lines. Because both the dependent variables and network metrics are standardized, the coefficients measure the standard-deviation change in the future aggregate outcome associated with a one-standard-deviation increase in the corresponding network metric, conditional on the other network metrics and controls.



(a) Intermediate prediction horizon (1 year).



(b) Long prediction horizon (3 years).

Figure 14: *Standardized regression coefficients for output variables.* These figures show the regression point estimates of Eq. (4) through shapes, with 95% asymptotic confidence bands based on Newey-West-adjusted standard errors through horizontal lines. Because both the dependent variables and network metrics are standardized, the coefficients measure the standard-deviation change in the future aggregate outcome associated with a one-standard-deviation increase in the corresponding network metric, conditional on the other network metrics and controls.