The Network of Firms Implied by the News

G. Schwenkler and H. Zheng*

October 29, 2019†

Abstract

We show that the news is a rich source of data on distressed firm links that drive firm-level and aggregate risks. The news tends to report about links in which a less popular firm is distressed and may contaminate a more popular firm. This constitutes a contagion channel that yields predictable returns and downgrades. Shocks to the degree of news-implied firm connectivity predict increases in aggregate volatilities, credit spreads, and default rates, and declines in output. To obtain our results, we propose a machine learning methodology that takes text data as input and outputs a data-implied firm network.

Keywords: Networks, contagion, predictability, risk measurement, machine learning, natural language processing. JEL codes: E32, E44, L11, G10, C82.

*Schwenkler & Zheng are at the Department of Finance, Boston University Questrom School of Business. Schwenkler is corresponding author: Phone (617) 358-6266, email: gas@bu.edu, web: http://www.gustavo-schwenkler.com.

†We are grateful to Federico Bandi, Christian Brownlees (discussant), Jie Cao (discussant), Francis Diebold, Diego García, Kay Giesecke, Tarek Hassan, Christian Hayerdahl-Larsen (discussant), Bernard Herskovic, Andreas Neuhierl (discussant), Emre Ozdenoren, Norman Schürhoff, Alireza Tahbaz-Salehi, Scott Weisbenner (discussant), seminar participants at Boston University and the Office of the Comptroller of the Currency, and participants at the 2019 Future of Financial Information Conference, Front Range Finance Seminar, Annual Conference of the Society for Financial Econometrics, European Summer Symposium in Financial Markets at Gerzensee, Annual NUS Risk Management Conference, WRDS Advanced Research Scholar Program, Econometric Society European Meeting, Northern Finance Association Annual Meeting, International Workshop in Financial Econometrics, and INFORMS Annual Meeting.

1 Introduction

Recent research shows that the network of business linkages across firms is a key determinant of firm-level risks and aggregate outcomes. Azizpour et al. (2018), Cohen and Frazzini (2008), Jorion and Zhang (2009), Herskovic et al. (2019) and others show that firm links facilitate the contagion of risks across firms, affecting asset prices, volatilities, and default probabilities. Acemoglu et al. (2012), Carvalho (2010), Gabaix (2011), and Herskovic (2018) show that the architecture of the network of firm links determines whether idiosyncratic shocks are amplified to aggregate shocks in the broader economy. In spite of the demonstrated importance of the network of firm links for risk measurement, data access is notoriously limited. Often, only incomplete and lagged data are available. The unavailability of extensive and timely firm network data hinders the precise measurement of risks that drive economic outcomes.

We show that the news is a rich source of information about distressed firm links that drive firm-level and aggregate risks. We develop a machine learning methodology that takes news data as an input and outputs a network of firm connections implied by the news. Our news-implied networks include a vast majority of the links recorded in currently available data sets. In contrast to the currently available networks, however, news-implied networks capture a wider range of firms and links, and are available in high frequencies. Consistent with a reader demand consideration mechanism, we find that the news tends to report about links in which a less popular firm is distressed and may contaminate a more popular firm. These links enable contagion effects that yield predictable stock returns and credit downgrades. On an aggregate level, we show that news-implied firm networks capture information about contagion and uncertainty effects that drive aggregate outcomes. We find that measures of connectivity in the news-implied firm network predict short-term increases in aggregate volatilities and bond spreads, as well as persistent increases in default activity and declines in output. Our methodology and data are freely available for download, facilitating the use of news-implied firm networks for empirical work. All in one, the results of this paper enable the estimation of accurate measures of firm-level and aggregate risks.

We analyze an extensive data set containing over 100,000 financial news articles published by Reuters between 2006 and 2013. In order to understand the informational content of the news, we develop a machine learning methodology that takes text data as an input and outputs a network of firm links implied by the data.[1] We exploit novel *natural language processing* (NLP) tools to identify the names of corporations in text data.[2] NLP is commonly used to estimate the sentiment of media content – that is, whether the media expresses negative or positive opinions – and how sentiment affects asset prices and macroeconomic factors; see Baker et al. (2012), Beber et al. (2015), Chen et al. (2014), Da et al. (2015), Das and Chen (2007), Engelberg et al. (2012), García (2013), Jegadeesh and Wu (2013), Ke et al. (2019), Tetlock (2007), and Shen et al. (2017), among others. The application of NLP for sentiment analysis is a univariate exercise: It extracts from a large dimensional text data set an aggregate measure of sentiment. In contrast, we extract bivariate signals from text data. Our methodology identifies two firms that are connected to each other and assesses how strong this relationship is. Our identifying assumption is that if two firms share a business connection, then the news should report about this link in an article by mentioning the two firms in the same sentence.[3] The stronger the relationship is, the more often should the news report about this relationship in different articles.

Our NLP methodology identifies about 3,000 firms from the CRSP / Compustat universe in our news data, together with over 20,000 distinct firm links. Our approach is highly accurate, correctly identifying more than 70% of all firms mentioned in the text data. We capture a majority of the links implied by the Compustat segments data (customer-supplier links), the 10-K similarity scores of Hoberg and Phillips (2016) (peer links), the EDGAR co-search measures of Lee et al. (2015) (peer links), and the covariance structure of firms' stock returns (correlation links). We also capture strategic partnerships, intra and inter-sectoral competitive links, as well as credit, financing, banking, and subsidiary relations. The network of firms implied by our news data showcases a core consisting of large banks that are strongly interconnected and several smaller banks that are connected to the larger banks, making up a core-periphery structure for the financial sector. Core-periphery structures are often identified in empirical and theoretical studies of interbank networks; see Babus and Hu (2017), Farboodi (2017), and Gofman (2017). There are several clusters of non-financial firms surrounding the financial firms, delivering a star architecture for the broader network of firms as in Acemoglu et al. (2012). Turning to the dynamic evolution of the network over time, we find that some sectors become more or less

[1] All codes have been written in R and are available for download at `http://www.gustavo-schwenkler.com`.

[2] NLP has become increasingly popular in economics research; see Engle et al. (2019) and Jelveh et al. (2018) for recent applications of NLP for the analysis of climate change and the influence of political partisanship.

[3] We show in an online appendix that the majority of the economically relevant information about firm links is communicated in individual sentences rather than across sentences of a news article, validating our approach.

prominent over time but the financial sector remains central and strongly connected.[4] These observations are consistent with the centrality of the financial sector highlighted by Bernanke et al. (1999) and Carvalho and Gabaix (2013).

Our first set of results shows that demand-side considerations incite the news to report about firm links that actively transmit risks across firms and lead to contagion. These results highlight the news as a primary data source to identify distressed firm links. Logit regressions reveal that the news is more likely to report about a firm link when one of the linked firms experiences negative stock returns, high volatility, credit downgrades, negative net income, or downward revisions by earnings analysts. We find that the likelihood of observing a firm link in the news is primarily driven by whether the less popular linked firm experiences financial distress in the form of negative stock returns, credit downgrades, or downward earnings estimate revisions. Our results suggest that the link likelihood is higher if the less popular linked firm experiences distress, but it is not higher if the more popular linked firm does. We establish these results controlling for firm characteristics and market conditions with time, firm, or link fixed effects, regardless of whether we proxy popularity by the market capitalization or the number of institutional investors of a firm. These findings are highly robust. They hold when we consider the 3,000 most frequently identified links in the data, suggesting that this is a pervasive phenomenon. They also hold when we only consider the links among the 500 largest firms in our data, suggesting that our results are not driven by the fact that there are more small firms in the economy. Finally, our results also hold when we exclude data recorded during the financial crisis, indicating that our results are not driven by this unique period in our sample.

Our data imply that the smaller linked firm is in distress at the time the link is reported and its health continues to deteriorate in the 6 months after. Even though the larger linked firm does not clearly experience distress at the time the link is reported, it accrues significantly negative stock returns in the months post link. Its credit rating also significantly drops after being linked with a smaller distressed firm in the news. The effects for the larger firm are transient and dissipate after four months. These results are consistent with a contagion channel through which investors slowly learn about the larger firm's exposure to the smaller distressed firm and adjust their trading behavior, which results in predictability in the asset prices of the larger firm.[5] A counterfactual simulation study shows that our results are not driven by the

[4] The monthly series of news-implied networks can be downloaded at `http://www.gustavo-schwenkler.com`.

[5] That predictability can arise as a result of slow information diffusion across economically linked firms was established by Cohen and Frazzini (2008). Our results do not contradict alternative findings about lead-lag

fact that the news tends to report about firms that experience negative shocks, as was recently established by Niessner and So (2018). Instead, we find that the news choses to report about links that actively transmit risks across firms. Our findings are consistent with a mechanism advocated by Mullainathan and Shleifer (2005) and García (2018) that posits that demand-side considerations incite the news to report about adverse shocks that affect the health of popular firms. They extend the results of Scherbina and Schlusche (2015) by showing that the news is not an unbiased source of information about firm links.[6]

Our next set of results shows that the information contained in news-implied firm networks is highly predictive of aggregate outcomes. Following Acemoglu et al. (2012) and Herskovic (2018), who theoretically show that the degree of connectivity in economic networks is a key driver of aggregate risks, we compute several measures of connectivity in the news-implied firm network: An average degree measure that is proportional to the number of links reported in the news in a given month, a first-order interconnectivity measure that states whether the monthly network is more centralized or more dispersed, and a second-order interconnectivity measure that captures whether clusters of firms are strongly or weakly connected to each other through intermediate firms. All three connectivity measures spike during recessions and are mostly unrelated to the sentiment of the news articles from which we extract our networks. However, their influence on aggregate risk measures is quite different. We consider a monthly vector autoregressive model of the three connectivity measures together with the return of the S&P 500 index, the VIX, the level and slope of the Treasury yield curve, the AAA and BAA corporate credit spreads, the GDP growth rate, and the aggregate default rate among U.S. corporations. Impulse response functions show that orthogonal shocks to our average degree measure trigger short-run increases in the VIX and the corporate credit spreads that remain significant for up to three months.[7] In contrast, orthogonal shocks to the second-order interconnectivity measure trigger significant increases in the aggregate default rate and significant declines in the GDP growth rate that can persist for 12 or more months.

The different impacts of shocks to average degree and second-order interconnectivity are

relationships in equity markets that show that information tends to flow from large to small firms (see Lo and MacKinlay (1990) and Hou (2007)). Instead, we show that the news reports about links in which information flows from a smaller to a larger firm.

[6] Nimark and Pitschner (2019) show that selective news reporting is an equilibrium outcome when agents have attention constraints and delegate the collection of information to news outlets.

[7] Our identification strategy is based on a Cholesky decomposition of the residual variance-covariance matrix.

due to the different informational content of these measures. We find that the average degree of the news-implied network is closely related to the financial uncertainty measures of Baker et al. (2019), Carriero et al. (2018), and Jurado et al. (2015). This suggests that the average degree captures information about financial uncertainty that drives short-term fluctuations in risk premia. On the other hand, the second-order interconnectivity measure is closely related to a measure of credit risk contagion introduced by Azizpour et al. (2018). This observation suggests that the second-order interconnectivity measure captures information about contagion effects that drive aggregate credit risk in the economy. Consistent with the theoretical models of Acemoglu et al. (2012) and Herskovic (2018), our results show that connectivity in the news-implied network is related to measures of aggregate risks and is predictive of adverse aggregate outcomes. Our findings support a mechanism proposed by Chahrour et al. (2019) that posits that information disseminated in the news can trigger aggregate shocks when firms are constrained and outsource the monitoring of their production networks to news publishers. They align with the results of Manela and Moreira (2017) and Liu and Matthies (2018), who in different settings also show that information contained in the news can be used to forecasts aggregate risks.

Finally, we show that news-implied firm networks capture information that is not contained in alternative networks. We consider a customer-supplier network extracted from Compustat's segments data, a firm similarity network extracted from the the 10-K textual similarity scores of Hoberg and Phillips (2016), a peer network proposed by Lee et al. (2015) that is implied by the frequency with which users look up two firms within a short period of time on the Securities and Exchange Commission's Electronic Data-Gathering, Analysis, and Retrieval (EDGAR) website, and a network implied by a variance decomposition for firms' stock returns as proposed by Demirer et al. (2018). Regressions suggest that the connectivity measures of our news-implied network are negatively related to the connectivity measures of the firm similarity network implied by Hoberg and Phillips (2016). Still, the R^2 are low. Compared to the connectivity measures of the alternative networks, we find that the connectivity measures of our news-implied network are better predictors of the levels of the VIX, the corporate credit spreads, as well as the growth rates of the S&P 500 Index and GDP at the monthly horizon. These findings highlight the prominent nature of news data to identify firm networks that are predictive of aggregate outcomes.

The rest of this paper is organized as follows. Section 2 introduces our data and methodology and Section 3 summarizes the methodology's output. Section 4 describes the estimated networks. Sections 5 through 7 present our empirical results. Section 8 concludes. There is an online appendix that describes details of our methodology and also contains robustness tests. The Online Appendix is available at http://www.gustavo-schwenkler.com.

2 Data & methodology

We obtain an extensive full-text news dataset from Ding et al. (2015). The data contains Reuters financial news articles published between October 20, 2006, and November 20, 2013. There are 106,521 articles in total. Table 1 provides summary statistics of the news articles and Panel (a) of Figure 1 provides a sample news article in the data. We see that an average article is fairly large, including about 600 words and 21 sentences. There is also significant variability across articles: One article contains over 6,000 words while others only contain a few sentences sentence. Panel (b) of Figure 1 shows that the number of articles published each year is fairly constant, although we have a much shorter sample for the year 2006.

2.1 Identification

We analyze each news article in our data to identify whether an article reports about a relationship between two firms. Intuitively, if an article reports about two firms that share some sort of business relation, then these two firms should be mentioned within close proximity from each other. Based on this insight, we identify a link whenever two firms are mentioned in the same sentence of an article.

We show in the Online Appendix that the majority of the information about economically relevant firm links is contained in individual sentences rather than across sentences of an article. As a result, our approach provides a robust alternative to identifying firm links when two firms are co-mentioned in an article regardless of where in the article they are mentioned.

2.2 Methodology

We require a methodology that can identify firms mentioned in each sentence of a news article such as the one in Figure 1. This is not a trivial task. One could use a static list of firm names but, given the dynamic nature of firm birth and failure, a static firm name list may miss some firms. Furthermore, firm names are often abbreviated or replaced with pseudonyms in the news. For example, General Electric Company is often just called GE, Ford Motor Company is often just referred to as Ford, and JPMorgan Chase often goes by JPMorgan, J. P. Morgan, or J. P.

Morgan Chase. Keeping track of all possible abbreviations or pseudonyms is computationally costly. Finally, the use of alternative firm identifiers, such as tickers, also presents a series of challenges. Tickers are not always mentioned in news articles. Even when they are, tickers change periodically and this restricts the usefulness of a static list of tickers.

We develop a three-step machine learning methodology to address these challenges. We summarize the methodology here and provide details in the Online Appendix. The methodology we develop can be applied for any sort of text data, although we focus here on news data.

The first step consists of using a natural language processing (NLP) toolkit to identify all nouns mentioned in a news article that could potentially be firm names.[8] We use the Stanford *coreNLP* toolkit available in R for this step (see Manning et al. (2005)). The coreNLP toolkit is a popular natural language processing software that identifies in text data nouns that refer to entities and classifies these into different categories: named entities ("PERSON", "LOCATION", "ORGANIZATION", "MISC"), numerical entities ("MONEY", "NUMBER", "ORDINAL", "PERCENT"), and temporal entities ("DATE", "TIME", "DURATION", "SET"). Consider as an example the first sentence of the article in Figure 1: *"Several aspects of the tentative contract between General Motors Corp (GM.N) and the United Auto Workers union will be hard for Ford Motor Co. (F.N) and Chrysler LLC to match in labor talks expected to heat up in coming days, people familiar with the negotiations said."* Figure 1 shows the output of the coreNLP algorithm applied to this sentence. The coreNLP algorithms recognize the following entities in the sentence: (GM, ORGANIZATION), (Ford, ORGANIZATION), (Chrysler, ORGANIZATION), and (Tuesday, DATE). Even though coreNLP does not recognize United Auto Workers union as an entity, it performs well at recognizing all three corporations mentioned in the sentence. The coreNLP toolkit has been demonstrated to be highly accurate in identifying named entities (see Abdallah et al. (2017), Atdag and Labatut (2013), and Costa et al. (2017)).

In the second step, we take all organizations identified by the coreNLP toolkit and run an algorithm developed by us to determine which of these organizations are corporations (details can be found in the Online Appendix). We first remove all organizations whose names contain words that signal government agencies or nonprofit institutions, such as "agency", "cooperation", "federal", "foundation", or "university." For the remaining organizations, we remove from their names all special symbols, unreasonable postfixes, and words that indicate business types (like

[8] Natural language processing (NLP) is a branch of machine learning that focuses on processing and analyzing text data. Gentzkow et al. (2019) provide an overview of how NLP is used for financial economic research.

"Co.," "Inc.," and "Ltd"). We assume that every organization that survives these steps is a firm. Still, there may be instances in which one firm goes by several names. We run additional steps to determine a unique name for each firm. We begin by creating clusters of firms with common words in their names and consider the most frequently mentioned name in a cluster as the name stem. Consider the following example. Suppose there is a cluster consisting of 6 firms that go by the names "Toyota," "Toyota USA," "Toyota Motor," "Toyota Motor Credit," "Toyota Motor," and "Toyota Motor". In this cluster, the most frequent name is "Toyota Motor" so we designate "Toyota Motor" as the name stem for the cluster. Then, for each one of the firms in the cluster we check whether the name of the firm is fully contained in the stem or vice versa. If so, we update the stem to be either the name of the firm or the prevalent stem, whichever is shorter. If not, we remove the firm from the original cluster. We proceed iteratively until no more improvements of the name stem can be made. All firms that remain in the cluster are considered to be the same firm and we assign the name stem as the name of this firm. In our example, we would iterate through the firms named "Toyota," "Toyota USA," and "Toyota Motor Credit." Given that "Toyota" is the shortest name fully contained in the original stem, we would update "Toyota" to be the new firm name stem. Then, because "Toyota" is contained in all other firm names in this cluster, we would update all other names to "Toyota" and terminate the iteration.

In a final step, we match the firms identified in the previous steps with firms in the merged CRSP / Compustat database. We follow a similar procedure as in Step 2. Details can be found in the Online Appendix.

Steps 1 (coreNLP), 2 (firm identification), and 3 (firm matching) introduced above deliver a list of firms mentioned in our news data. When running these steps, we keep track of the article in which a firm is mentioned, the sentence within an article where the firm was identified, and the publishing date of the article. We establish that two firms share a connection whenever the firms are identified in the same sentence of an article. All codes used to run our algorithms have been written in R and are available for download at http://www.gustavo-schwenkler.com.

3 Output of methodology

Our methodology finds 656,167 firm mentions in the data. Figure 2 shows the number of recognized firm mentions in each year. Except for the year 2006, for which we have a shorter data sample, we see that our algorithm recognizes around 90,000 firm mentions in any given year.

We also see that the number of firm mentions in any given year is fairly constant. Of course, not every mention corresponds to a different firm: Some firms are mentioned repeatedly. Our algorithm identifies 2,961 different firms during the time span covered by the data. On any given year, our data covers about 1,300 distinct firms. The five most frequently identified firms are General Motors, Chrysler, Citigroup, Apple, and Goldman Sachs.

Table 2 and Figure 2 provide descriptive statistics for the firms in our sample. We see that the majority of the firms are publicly listed in U.S. exchanges. The median firm in our sample is an investment-grade small cap firm with a market capitalization of about $1.6 billion. There is significant dispersion in the distribution of firm sizes, covering the whole spectrum between small and large caps. The OLS estimator for the exponent of a power law that approximates the market capitalization distribution of the largest 1000 firms in our sample is 0.952 with a standard error of 0.043, which is similar to the power law exponents estimated by Gabaix and Landier (2008) and Luttmer (2007) for the size distribution of the largest U.S. firms. We see that firms in our sample cover all 11 GICS sectors. The distribution of sectors in our data is comparable to the sector distribution in the whole CRSP / Compustat universe. Minor differences can be observed in the telecommunication services and consumer staples and discretionary sectors, which are slightly overrepresented in our sample. The IT and health care sectors are slightly underrepresented. We also see that our data contains firms that are domiciled in the U.S. and abroad, in countries such as Canada, China, Great Britain, Bermuda, and many others.

Firms in our sample hold an average of $28 billion in assets, $7 billion in debt, and $990 million in cash. The average firm is profitable, with a quarterly reported net income of $128 million. It pays dividends on an annual basis of around $0.13 per share. The average firm has a 24% leverage ratio and an annualized realized volatility of 67%. It is followed by 7 analysts and has an institutional ownership of about 50%.

Turning to our sample of firm links, Table 1 shows our methodology identifies 1.31 firm connections per article with a standard deviation of 3.68 connections per article. Over the whole data sample, there are 177,300 instances in which two firms are mentioned in the same sentence. This corresponds to 20,504 unique links between 2,406 firms.[9] Table 3 shows the results of a logit regression for the likelihood that a firm was linked at some point in our sample based on characteristics of the firm. We see that the firms that are linked in our sample tend to be larger firms with large analyst coverage.

[9] Considering that there are 2,961distinct firms in our sample, this suggests that not all firms are connected.

Table 4 reports the most frequently identified links in our sample together with samples of the sentences in which these links are identified. We see that several of the sentences point to competitive relations. These competitive relationships can be strategic (such as when Google cooperated with Apple) or destructive (as in the case of Microsoft challenging Google). Some sentences point to joint investment banking solutions provided by big banks (such as Goldman and Morgan Stanley sponsoring the Alibaba IPO) while other sentences point to interbank relationships (like when Citigroup, Credit Suisse, Deutsche Bank, Goldman Sachs, Merrill Lynch, Morgan Stanley and UBS formed a new company to develop a trading platform). We identify some credit relationships, as when it was reported that Daimler is covering liabilities at Chrysler. We also identify parent-subsidiary relationships (Vodafone owns a majority stake in Verizon) and M&A links (Bank of America buys Merrill Lynch).

Table 5 provides descriptive statistics of the firm links in our data, where the firms in a link are sorted by market capitalization. We see that the larger firm in a link tends to be higher ranked and have more assets, debt, cash, net income, sales, and expenses than the smaller counterpart. The larger firm also tends to be less volatile and pay out more dividends than the smaller firm. Only about 20% (40%) of the links are composed of firms in the same industry (sector).

Table 6 lists the 20 most frequently identified cross-sectoral links, which cover about 78% of all cross-sectoral connections in our sample. It also provides sample sentences for each cross-sectoral link. We see that there are several links between the financial and non-financial sectors, pointing to rating, banking, credit, and other financing solutions provided by the financial sector to the wider economy. We also identify links across non-financial sectors, such as customer-supplier links (Boeing selling airplanes to UPS), strategic partnerships (collaboration between GlaxoSmithKline Plc and IBM), and several mergers and acquisitions.

3.1 Validation

Reuters news articles have a helpful feature that facilitates the validation of our algorithm. As showcased in the sample sentence of Figure 1, the names of publicly traded firms in Reuters financial news are often followed by an identifier known as the *Reuters Instrument Code* (RIC). The RIC characterizes the ticker of the firm and the exchange where its stock is traded. We exploit the availability of RIC in our text data to validate our methodology.

We collect all RIC in our text data and match the resulting tickers with tickers in the CRSP / Compustat database. We are able to identify 77,080 RIC mentions corresponding to

1,858 distinct firms that are also matched in CRSP / Compustat. For the set of RIC mentions, we ask the following questions: How many of the RIC mentions are identified in Step 1 as organizations by the Stanford coreNLP algorithm? How many of the RIC mentions identified as organizations in Step 1 are classified as firms in Step 2 of our algorithm? How many of the RIC mentions identified as firms by Step 2 are matched to the correct firm in the CRSP / Compustat database by Step 3 of our algorithm? Out of the firms that are not properly matched by Step 3 of our algorithm, is the failure due to a mismatch (i.e., we match with a different firm than suggested by the RIC) or due to a non-match (i.e., we are unable to find a firm in the CRSP / Compustat database that matches with the firm name assigned by our algorithm)?

Table 7 provides answers. We see that out of the 77,080 RIC mentions, Step 1 of our algorithm correctly identifies around 84% as organizations. This rate of accuracy for the coreNLP toolkit is in line with similar accuracy estimates by Abdallah et al. (2017), Atdag and Labatut (2013), Costa et al. (2017), and Pinto et al. (2016) based on alternative text data. Once an RIC mention is identified as an organization by Step 1 of our algorithm, Step 2 also correctly labels that RIC mention as a firm. This suggests that our firm identification algorithm is highly accurate. Finally, out of the set of RIC mentions that we correctly identify as firms, Step 3 of our algorithm matches around 87% to the correct firm in CRSP / Compustat. Putting everything together, these results show that our approach correctly identifies and matches 73% of all RIC mentions in the data. When our methodology fails, it is primarily due to an inability to identify the firm as an organization or to match with a firm in the CRSP / Compustat data. These types of errors only reduce our sample size; they should not bias our results. We match a firm in our data with a wrong firm in the RIC data only in 2.75% of all RIC mentions. Table 7 shows that there are no firms with RIC that are never correctly identified and matched by our algorithm. While mismatches are serious, the facts that mismatch is an extremely rare phenomenon and that no firm is consistently mismatched mitigate the concerns.

The above analysis highlights an advantage of our approach compared to relying on the availability of Reuters Instrument Codes. Our methodology is able to identify 2,961 firms that are cross-matched with firms in the CRSP / Compustat database. In contrast, there are only 1,858 firms with RIC mentions that are also matched in CRSP / Compustat. We find that 78% of all firms with RIC mentions are also included in the firm sample identified by our algorithm. Table 8 provides summary statistics of the firms with RIC mentions. We see that the RIC firms tend to be mid cap or larger firms with less volatility and more assets, cash, debt, sales, expenses, and net income than the firms in our sample. The maximum column of Table 8 is almost the same as the maximum column of Table 2, suggesting that our sample and the RIC sample cover similar sets of large firms. However, our approach is able to identify small cap firms and firms with stocks that are traded over-the-counter more frequently than an alternative approach based on RIC. These results indicate that our methodology can identify a more extensive range of firms than an alternative approach based on RIC. They provide further validation for our approach.

4 Estimated news-implied networks

4.1 Full data sample

We plot in Figure 3 the network of firms implied by all news articles in our data sample. Each node represents a firm. The size of a node is proportional to the number of times that firm is found in the data while the width of a link is proportional to the number of times that link is identified in the data. For clarity, in Figure 3 we only show the largest 50 nodes, which correspond to the most frequently identified firms in the sample. We label firms with their tickers.

We observe several interesting features. We first see that the big banks – Citigroup, Goldman Sachs, JPMorgan Chase, Bank of America, and Morgan Stanley – represent some of the largest and most central nodes in our network, suggesting that the news reported very frequently about relationship between these major banks and other firms. The large banks are also highly interconnected, indicating that the news often reported about the relation between big banks. There are several smaller banks that lie on the periphery: Deutsche Bank, Lehman Brothers, Credit Suisse, UBS, Barclays, Merrill Lynch, Wells Fargo, RBS, ABN Amro, and HSBC. Banks in the network of Figure 3 have a core-periphery structure with large banks being highly central and highly interconnected and smaller banks being connected to the larger banks on the outskirts. Such a core-periphery network is often observed in interbank data; see Craig and von Peter (2014), in 't Veld and van Lelyveld (2014), and Gofman (2017), among others. Core-periphery networks have also been demonstrated to arise naturally in interbank network formation models; see Babus and Hu (2017) and Farboodi (2017).

The network in Figure 3 also highlights the central position of the banking sector in the general economy, as advocated by Bernanke et al. (1999) and Carvalho and Gabaix (2013). We see that most non-financial firms are located in the outskirts of the network, surrounding the large banks in the center. Several firms are only indirectly connected because they share a common link with one of the banks. For example, Chrysler and Apple are indirectly connected in Figure 3 because they share a link with JPMorgan.

Figure 3 exhibits several sector-based clusters. On the top right corner, we find a cluster of firms associated with the IT and telecommunication services sectors. Below it we find an automobile cluster. The bottom left part of Figure 3 is dominated by financial firms. These clusters arise because the news often report about connections between firms in the same sector in addition to intersectoral relationships (see Table 6). The general architecture of the news-implied network resembles the star network of intersectoral connections estimated by Acemoglu et al. (2012) from input-output linkage data for the United States. For a full comparison, we aggregate firms in our sample by two-digit NAICS codes and display in Figure 4 the resulting news-implied intersectoral network. We also display in Figure 4 the intersectoral networks implied by the 2012 BEA industry-by-industry total requirement tables.

We see that the news-implied intersectoral network in Panel (a) of Figure 4 exhibits a similar star structure as highlighted in Acemoglu et al. (2012) and also showcased in the BEA input and output networks in Panels (b) and (c). Similar as in the BEA input network, the most prominent sector in our intersectoral network is the manufacturing sector (NAICS code "33"). This sector includes computer, electrical, furniture, machinery, metal, and transportation manufacturing firms which heavily dominate the production of final goods. Because the BEA data mostly measures the use and production of commodities, the BEA input-output networks diminish the importance of the insurance and financial sectors (NAICS codes "51" and "52", respectively). In contrast, those sectors are highly central and prominent in our news-implied network, consistent with the theoretical models of Bernanke et al. (1999) and Carvalho and Gabaix (2013) that put the financial industry at the center of the U.S. economy.

4.2 Time series of networks

We plot yearly time series of networks implied by news articles in our data sample in Figures 5 and 6. For each year between 2006 and 2013, we use the methodology of Section 2 to extract all firm links implied by news articles published in that year. For clarity, we only plot the connections between the largest 50 firms in every year together with their tickers. Similar plots can be constructed for arbitrary frequencies – as frequently as daily or hourly and as infrequently as quarterly or annually. A monthly time series of the estimated news-implied firm network is available for download at `http://www.gustavo-schwenkler.com`.

The time series of news-implied networks yields several additional insights. We see that the architecture of the news-implied network can change drastically from year to year, according to how the news report about the relationships between firms. The news-implied networks in 2006 and 2007 were relatively dispersed with a central cluster associated with the financial sector and some non-financial clusters dispersed in the periphery. Entering the financial crisis in 2008, the news-implied network became more centralized, showing a strongly connected core of banks. The automobile sector became dominant in the network for the year 2009, consistent with the prevailing crisis in that sector. After 2010 when the great recession ended, the news-implied networks again showcase a more common star structure as in Acemoglu et al. (2012), with banks located in the center and other sectors positioned around the financial sector.

We summarize the information contained in the time series of news-implied networks. For this, we consider the time series of three measures of connectivity:

(1) The average degree, which measures the number of connections of an average node.[10] The average degree is inversely related to the network sparsity measure of Herskovic (2018).

(2) The first-order interconnectivity measure of Acemoglu et al. (2012), which is given by the coefficient of variation of the degree distribution in the network and measures how dispersed the degree distribution is. A more dispersed distribution implies that there are few large nodes that have connections with many small nodes. As a result, high first-order interconnectivity is characteristic of a network that showcases few very large nodes in the center and many smaller nodes in the periphery. First-order interconnectivity is closely related to the network concentration measure of Herskovic (2018).

(3) The second-order interconnectivity measure of Acemoglu et al. (2012), which is the weighted covariance of the degree of two nodes that are indirectly connected through a third node. Second-order interconnectivity highlights how strongly two clusters of nodes are indirectly connected through intermediate nodes.

Figure 7 plots the monthly time series of these connectivity measures for our news-implied network. For this, we generate analogous networks as those in Figures 5–6 but on a monthly basis and then compute the implied connectivity measures. We see that there is significant time variation in the interconnectivity measures. The time series of connectivity measures appear to be persistent. The regression results in Table 9 confirm these visual insights by showing that the

[10] The degree of a node is the number of links that a node shares with other nodes in the network.

monthly AR(1)-coefficients of 0.4940 for the average degree, 0.533 for first-order interconnectivity, and 0.479 for second-order interconnectivity are significantly large.

We evaluate the relationship between connectivity and sentiment (see Figure 7 for the time series of the average article sentiment in our sample). One may be concerned that the connectivity measures capture negative sentiment in the news articles. We check whether this is the case by running monthly regressions of our connectivity measures on the average article sentiment. Table 9 summarizes the results. We find that our measures of news-implied connectivity are indeed negatively related to sentiment. However, the R^2 of the regressions are low. After controlling for the autoregressive nature of connectivity and the influence of sentiment, about 70% of the time series variation of the connectivity measures remains unexplained. We therefore reject the notion that spikes in our interconnectivity measures are only driven by sentiment. These results suggest that news-implied connectivity conveys information that is complementary to the sentiment of the news articles from which we extract our networks.

5 Link level results

We evaluate whether the news is an unbiased source of information about firms links. For this, we study the drivers of the likelihood of observing a firm link in our news data. We consider the 3,000 most frequently identified links, which cover 904 distinct firms. For the firm pairs that were observed in at least one month of this subset of our data, we create monthly time series of link-level dummy variables that indicate whether in a given month we identified a link between the two firms. We then run logit regressions of the link dummies on several measures of financial performance for the linked firms as well as characteristics of the linked firms that were found to be link predictive in Table 3, and market and macroeconomic controls. Table 10 summarizes our findings, where we randomly assign the labels "Firm 1" and "Firm 2" between the two linked firms in a given month but keep the link identifier fixed throughout the sample. Table 11 provides summary statistics of our control factors.

We find that it is more likely to observe a firm link in the news when one of the two firms experiences high volatility, credit downgrades, downward revisions by earnings analysts, or negative net income. We also find some evidence that the likelihood of observing a link is higher when the one of the two firms experiences negative monthly stock returns or negative earnings surprises. When controlling for all firm characteristics and the state of the economy, we do not find that the link likelihood is higher when positive stock returns, credit upgrades, upward earnings estimate revisions, or positive net income occur. These result hold with clustered standard errors in the presence of time and link fixed effects, which control for cross-sectional differences in link frequencies and time series fluctuations of the network architecture.

We run several robustness tests in the Online Appendix. We obtain similar results if we include firm rather than link fixed effects. This suggests that our results are not driven by uncontrolled differences in how the news reports about different firms. We also obtain similar results when we exclude links observed during the financial crisis, suggesting that our results are not driven by the severe nature of this period of time in our sample.

All in one, our results show that the news tends to reports about firm links in which one of the two linked firms is distressed. Our findings extend the results of Scherbina and Schlusche (2015) by showing that the news is not an unbiased source of information about firm links.

5.1 Mechanism

Why does the news report about distressed firm links? Recent research suggests that this may be due to demand-side considerations. Mullainathan and Shleifer (2005) show that the news attracts readers by fine-tuning their reporting to match their readership's interests. It is well known that investors are more concerned about downside risk than upside potential; see Kahneman and Tversky (1979), Kuhnen (2015), and others. Because of this, news outlets have incentives to publish articles about negative events that represent risks for investors. Indeed, García (2018) shows that a negative market return triggers more negative news reporting than a positive market return of equivalent magnitude and that this is primarily driven by reader demand considerations. Niessner and So (2018) demonstrate the the news is more likely to report about firms with declining financial health. Nimark and Pitschner (2019) show that selective reporting about newsworthy shocks is an equilibrium outcome in a model in which agents have attention constraints and delegate the collection of information to news outlets.

We evaluate whether demand-side considerations drive the news to report about distressed firm links. For this, we consider the relative popularity of the linked firms. If the news reports about distressed links to attract concerned investors as readers, then there should be stronger incentives to report about links that affect the health of popular firms. We therefore conjecture that the news is more likely to report about links in which a less popular firm is distressed and may contaminate a more popular firm. We test this hypothesis by repeating the analysis of

Table 10 when controlling for how popular the two linked firms are. We consider two proxies for how popular firms are among investors: The market capitalization of a firm and the number of 13-F institutional investors in a firm. Table 12 shows our findings.

We find that the news is more likely to report about a firm link when the less popular of the two linked firms experiences negative stock returns, credit downgrades, or downward revisions of earnings estimates. We do not find that the same applies for the more popular firm: Negative returns, credit downgrades, or downward estimate revisions for the more popular firm do not significantly increase the likelihood of observing the link in the news. Extending the results of Table 10, we also find that the news is more likely to report about links between firms that experience high volatility or negative net income, regardless of how popular these firms are. The Online Appendix shows that similar results hold if we control for firm fixed effects rather than link fixed effects, suggesting that the results are not driven by differences in how the news reports about different firms. We also obtain similar results when we consider the links shared among the largest 500 firms in the data or if we exclude links observed during the financial crisis. These findings suggest that our results are not driven by the fact that there are more small firms in the economy or by the severity of the financial crisis in our data. Putting everything together, our results show that whether or not a firm link shows up in the news is primarily driven by whether the less popular linked firm experiences financial distress.

We test whether, when reporting about firm links, the news delivers information about a potential contagion from the less popular firm to the more popular firm. For this, we consider the subsequent performance of the two linked firms in the six months after a link between the two firms is observed in the news. Figure 8 summarizes our findings. We find that on average, in the month a link is observed in the data, the smaller counterparty of a link experiences a statistically significant credit downgrade and statistically significant negative monthly stock return. In the subsequent months, the health of the smaller counterparty continues to deteriorate: its net income falls, it accrues negative returns, and its credit worthiness declines. The larger counterparty, on the other hand, generally does not experience negative stock returns when a link with a smaller counterparty is reported in the news. In the months after being linked with a smaller firm in the news, however, the financial health of the larger firm deteriorates on average without a corresponding deterioration in fundamentals. It accrues significantly negative stock returns while experiencing significant declines in its credit worthiness, but its net income does not decrease. The negative effects on cumulative stock returns and credit worthiness are transient and dissipate after 4 months. These findings suggests that investors learn about the larger firm's exposure to the smaller distressed firm and they slowly incorporate this information in asset prices over the course of a few months.[11]

In summary, our results show that the news is more likely to report about links between firms in which a less popular firm experiences distress and may contaminate a more popular firm it is connected with. They highlight that the news is selective about what firm links it includes in its reporting. Consistent with Mullainathan and Shleifer (2005), our findings suggest that demand-side considerations drive the news to report about distressed firm links. Extending García (2018) and Niessner and So (2018), we show that the demand-side considerations incite the news to inform about the transmission of shocks from a less popular distressed firm to a more popular firm.

5.2 Counterfactual

One may be tempted to believe that our results are driven by the fact that the news tends to report about firms that experience a deterioration in financial health (Niessner and So (2018)). This would be the case, for example, if the news only reported about firms that experience financial distress and these firms ended up being mentioned in the same sentence of an article out of pure coincidence. To evaluate whether this is the case, we repeat the analysis of Figure 8 under the assumption that firms are linked at random if, in the month in which the link appears in the news, one of the two linked firms experiences any one of the following distress cases: Its monthly stock return lies in the bottom quintile of the stock return distribution, its monthly realized volatility lies in the top quintile of the realized volatility distribution, its net income lies in the bottom quintile of the net incomedistribution, or it experiences a credit downgrade. Figure 9 shows the post-link performance of firms that are linked this way.

In contrast to Figure 8, Figure 9 shows that the larger linked firm would not experience transient financial distress after a link is reported in the news if the link were drawn randomly when one of the two linked firms is distressed. Similar as in Figure 8, Figure 9 also shows that the smaller linked firm would experience a persistent deterioration of financial health in the 6 months post link if the reason the news were reporting about a link of this firm is because it is distressed. The analysis of Figure 9 confirms a key takeaway: The news reports about links

[11] The slow incorporation of information into the asset prices of the larger linked firm is consistent with a limited attention channel highlighted by Cohen and Frazzini (2008).

between a less popular firm that experiences financial distress and a more popular firm that is seemingly healthy. The information reported in the news is digested slowly by investors, resulting in asset price predictability.

6 Aggregate level results

Several recent papers argue that the architecture of the network of firm links is a key driver of aggregate risks. Carvalho (2010) and Gabaix (2011) show that idiosyncratic shocks can amplify to large aggregate fluctuations when business links spread shocks across firms. Acemoglu et al. (2012) show that aggregate risks are high in a disaggregated economy that is highly interconnected.[12] Herskovic (2018) and Herskovic et al. (2019) show that the degrees of concentration and sparsity of a production network are systematic risk factors that drive the cross-section of asset returns and idiosyncratic volatilities. These theoretical results motivate us to investigate the relationship between news-implied firm networks and aggregate risks.

We estimate a one-lag vector autoregressive (VAR) model for the joint dynamics of GDP growth, S&P 500 returns, the VIX, the level and slope of the yield curve, the AAA and BAA corporate credit spreads, the aggregate default rate, and news-implied connectivity. Figures 10 through 12 show the cumulative impulse response functions of the VAR model for one-standard-deviation orthogonal shocks to the average degree, the first-order interconnectivity, and the second-order interconnectivity of a monthly news-implied network. Our identification strategy is based on a Cholesky decomposition of the residual variance-covariance matrix, where we assume that GDP growth is the most exogenous variable and the news-implied connectivity measures are the most endogenous variables. The captions of Figures 10–12 provide details.

We see that shocks to the average degree of a news-implied network cause short-lived increases in the VIX and the corporate bond spreads. Shocks to the first-order interconnectivity measure are not associated with any significant short-term or long-term increases in any of the risk measures. In contrast, shocks to the second-order interconnectivity measure cause persistent and significant increases in the VIX, the BAA credit spread, and the aggregate default rate, as well as persistent and significant GDP declines. Shock to second-order interconnectivity are also associated with short-lived declines in aggregate stock returns.

Validating the theoretical results of Acemoglu et al. (2012), Carvalho (2010), Gabaix (2011),

[12] Here, *"disaggregated"* means that there are several clusters or sectors of firms in the economy while *"interconnected"* means that different clusters share links with each other.

Herskovic (2018), and Herskovic et al. (2019), our empirical results confirm that the architecture of the network of firms links that actively transmit risks across firms is a key determinant of aggregate measures of risks. Our results show that changes in the architecture of the news-implied firm network predict large aggregate fluctuations, in particular when the network becomes denser or more interconnected. These findings are consistent with a channel posited by Chahrour et al. (2019) that states that information disseminated in the news can trigger aggregate shocks when firms are resource constrained and rely on the news to identify risks in their production networks.

6.1 Informational content

We analyze the information contained in the different connectivity measures. We begin with the average degree of the news-implied network. A large value of the average degree tells us that the news reports about many links between firms. Links in the news-implied firm network are typically distressed and facilitate the transmission of risks across firms (see Section 5). As a result, periods of times with elevated average degree correspond to episodes in which firms face elevated risks. Such episodes are typically associated with high financial uncertainty. Based on this intuition, we evaluate whether the average degree of the news-implied firm network is associated with measures of financial uncertainty.

We consider the level of the VIX, which is a forward looking measure of stock market volatility, as well as the level of the VXO index, which is the implied volatility of short-term S&P 100 Index options and has been employed as a measure of financial uncertainty by Basu and Bundick (2017). We also consider the financial uncertainty measures of Jurado et al. (2015) and Carriero et al. (2018), which are derived from econometric models for the volatilities of macroeconomic and financial data. Finally, we consider a measure of stock market volatility by Baker et al. (2019) that is extracted from news data. Table 13 reports the estimates of the regressions of the financial uncertainty measures on lagged values of themselves and contemporaneous values of the connectivity measures for the news-implied network. The regressions also control for the aggregate sentiment of our news articles. We find that the average degree is positively correlated with all five measures of financial uncertainty. This holds when controlling for their strongly autoregressive nature and for the other connectivity measures. Our results show that the average degree of the news-implied firm network captures information about financial uncertainty. Given that financial uncertainty is a key driver of risk premia in financial markets, the association between the average degree of the news-implied firm network and the measures of financial uncertainty explains why shocks to the average degree measure trigger short-term increases in the VIX and the corporate credit spreads.

Second-order interconnectivity captures how strongly two largely connected firms are connected to each other through intermediary firms. Given that the links reported in the news facilitate the contagion of risks across firms (see Section 5), second-order interconnectivity tells us something about how long the contagion chain can be. If the second-order interconnectivity measure is large, then contagion can spread far in the economy across clusters of firms. This suggests that second-order interconnectivity contains information about the potential for contagion in the economy. Now, firms face elevated risk of default whenever they are hit with contagion (see Azizpour et al. (2018), Jorion and Zhang (2009), and others). As a result, we evaluate whether periods of elevated second-order interconnectivity are associated with high default activity due to credit risk contagion.

We estimate a monthly one-lag vector-autoregressive (VAR) model for the joint dynamics of the GDP growth rate, the aggregate default rate, a credit risk contagion factor estimated by Azizpour et al. (2018) that measures the component of the conditional arrival rate of defaults in the U.S. economy that is due to contagion of credit risk acrosseconomically linked firms, as well as our three connectivity measures.[13] Figure 13 shows the implied impulse response functions for the aggregate default rate and the contagion factor, together with 95% confidence intervals, after one-standard deviation orthogonal shocks to each of the three connectivity measures of the news-implied network. Our identification is based on a Cholesky decomposition of the residual variance-covariance matrix, where we assume that GDP growth is the most exogenous variable and the news-implied connectivity measures are the most endogenous variables.

Figure 13 shows that shocks to second-order interconnectivity predict significant and persistent increases in the aggregate default rate and the contagion factor of Azizpour et al. (2018) that materialize after 6 months and persist for at least 12 months. We further find that neither shocks to the average degree nor shocks to first-order interconnectivity trigger increases in default activity. These results confirm that second-order interconnectivity in the news-implied network gives a measure of the contagion potential in the U.S. economy.

[13]The VAR model can be viewed as a discrete-time version of the reduced-form portfolio credit risk model of Azizpour et al. (2018).

7 Relationship with alternative networks

We assess the informational content of news-implied networks compared to networks derived from alternative data sets. Aside of the BEA input-output networks discussed in Section 4, we consider four additional alternative networks. A network derived from customer segments data provided by Compustat, where the width of a link is proportional to the total sales associated with a firm pair. A network derived from the 10-K textual similarity scores of Hoberg and Phillips (2016) in which the size of a link is proportional to the similarity score between the 10-K's of two firms. A network proposed by Lee et al. (2015) in which the width of a link is proportional to the frequency with which investors look up information in consecutive order about the two linked firms on the SEC's EDGAR website. And, finally, a variance decomposition network for cross-sectional stock returns obtained using the methodology of Demirer et al. (2018). For the largest 50 firms by market capitalization in our data sample, Figures 14 and 15 plots the alternative networks and compares them to the network implied by our approach. We mark in green any link that is included in an alternative network and our news-implied network. Red links are links that are included in the alternative network but not in our news-implied network.

It is visible that our news-implied network differs from the competing networks. Our approach is able to extract many more firm links than would be extracted from the Compustat segments data or from the 10-K similarity scores of Hoberg and Phillips (2016). We capture all of the customer-supplier implied by the Compustat segments data. We also capture 83.87% of the links implied by the 10-K similarity scores of Hoberg and Phillips (2016), 77.47% of the links in the EDGAR co-search network of Lee et al. (2015), and 55.94% of the links in the variance decomposition network. We see that the network implied by a stock return variance decomposition is very dense. It shows a mesh architecture rather than a star architecture. All the links in the news-implied network are also included in the variance decomposition network. This suggests that the news reports about stock market shocks that are transmitted across firms.

We proceed to analyze whether the differences highlighted in Figure 14 are significant. For each of the alternative networks, we compute monthly time series of the connectivity measures described in Section 4.[14] We then regress each connectivity measure for our news-implied network on the analogous connectivity measures of the alternative networks. Table 14 summarizes our results. We find that the connectivity measures of our news-implied network are negatively

[14] We interpolate with the most recently available observation whenever the data is available in lower frequencies. We exclude the EDGAR co-search network from this analysis because it only overlaps with our data for 3 years.

related to the connectivity measures of the Hoberg and Phillips (2016) network. We also find a negative relationship between the second-order interconnectivity of our news-implied network and that of the BEA input network. The R^2 are low, however, maxing out at 16% for the first-order interconnectivity measure. These findings show that the connectivity measures of the news-implied network contain information that is not fully captured by alternative networks.

7.1 Horse race

In a final step, we evaluate the predictive power of news-implied networks in conjunction with alternative networks. We run one-step ahead predictive regressions for the S&P 500 returns, the VIX, the AAA and BAA corporate credit spreads, and the growth rate of GDP at the monthly frequency. We found in Section 6 that the connectivity measures of the news-implied network are predictive of these variables. We proceed to verify whether news-implied connectivity is a better predictor of these variables than the connectivity measures of alternative networks.

Table 15 summarizes the results of the predictive regressions. We find that measures of connectivity of the news-implied network predict the S&P 500 returns, the VIX, the AAA and BAA corporate credit spreads, and the GDP growth rates one month ahead even when controlling for the connectivity measures of the alternative networks. In contrast, we generally do not find that the connectivity measures of the alternative networks predict these proxies of aggregate distress. The out-of-sample R^2 suggest significantly higher predictive power when a model includes the connectivity measures of the news-implied network. F-tests reject the null hypothesis that the connectivity measures of the news-implied network are not predictive when also controlling for the connectivity of the alternative networks. The signs of the statistically significant estimates suggest that a more connected news-implied network predicts aggregate distress in the next month, consistent with the theoretical results of Acemoglu et al. (2012).

All in one, the results of this section show that news-implied networks are better predictors of measures of aggregate risks than networks derived from currently available data sets.

8 Conclusion

We show that the news is a primary source of data on distressed firm links that drive firm-level and aggregate risks. Our results demonstrate that the news tends to report about links that facilitate contagion and induce asset price predictability. Measures of news-implied firm connectivity are correlated with measures of financial uncertainty and credit risk contagion, and predict aggregate outcomes out-of-sample. The information contained in news-implied networks is complementary to the information contained in alternative networks. To obtain our results, we develop a machine learning methodology that takes news articles as input and extracts a network of firm connections implied by the news. The results of this paper enable the precise measurement of risks.

References

Abdallah, Zahraa S., Mark Carman and Gholamreza Haffari (2017), 'Multi-domain evaluation framework for named entity recognition tools', *Computer Speech & Language* **43**, 34 – 55.

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

Atdag, Samet and Vincent Labatut (2013), A comparison of named entity recognition tools applied to biographical texts, *in* 'Proceedings of the 2nd International Conference on Systems and Computer Science', IEEE, pp. 228–233.

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.

Baker, Malcolm, Jeffrey Wurgler and Yu Yuan (2012), 'Global, local, and contagious investor sentiment', *Journal of Financial Economics* **104**(2), 272–287.

Baker, Scott R., Nicholas Bloom, Steven J. Davis and Kyle J. Kost (2019), Policy news and stock market volatility. NBER Working Paper No. 25720.

Basu, Susanto and Brent Bundick (2017), 'Uncertainty shocks in a model of effective demand', *Econometrica* **85**(3), 937–958.

Beber, Alessandro, Michael W. Brandt and Maurizio Luisi (2015), 'Distilling the macroeconomic news flow', *Journal of Financial Economics* **117**(3), 489 – 507.

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

Carriero, Andrea, Todd E. Clark and Massimiliano Marcellino (2018), 'Measuring uncertainty and its impact on the economy', *The Review of Economics and Statistics* **100**(5), 799–815.

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 and Xavier Gabaix (2013), 'The great diversification and its undoing', *American Economic Review* **103**(5), 1697–1727.

Chahrour, Ryan, Kristoffer Nimark and Stefan Pitschner (2019), Sectoral media focus and aggregate fluctuations. Working Paper.

Chen, Hailiang, Prabuddha De, Yu (Jeffrey) Hu and Byoung-Hyoun Hwang (2014), 'Wisdom of crowds: The value of stock opinions transmitted through social media', *The Review of Financial Studies* **27**(5), 1367–1403.

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

Costa, C. M., G. Veiga, A. Sousa and S. Nunes (2017), Evaluation of stanford ner for extraction of assembly information from instruction manuals, *in* '2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)', pp. 302–309.

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

Da, Zhi, Joseph Engelberg and Pengjie Gao (2015), 'The sum of all fears investor sentiment and asset prices', *The Review of Financial Studies* **28**(1), 1–32.

Das, Sanjiv R. and Mike Y. Chen (2007), 'Yahoo! for amazon: Sentiment extraction from small talk on the web', *Management Science* **53**(9), 1375–1388.

Demirer, Mert, Francis X. Diebold, Laura Liu and Kamil Yılmaz (2018), 'Estimating global bank network connectedness', *Journal of Applied Econometrics* **33**(1), 1–15.

Ding, Xiao, Yue Zhang, Ting Liu and Junwen Duan (2015), Deep learning for event-driven stock prediction, *in* 'Proceedings of the 24th International Conference on Artificial Intelligence', AAAI Press, pp. 2327–2333.

Engelberg, Joseph E., Adam V. Reed and Matthew C. Ringgenberg (2012), 'How are shorts informed?: Short sellers, news, and information processing', *Journal of Financial Economics* **105**(2), 260–278.

Engle, Robert, Stefano Giglio, Heebum Lee, Bryan Kelly and Johannes Stroebel (2019), 'Hedging climate change news', *Review of Financial Studies* . forthcoming.

Farboodi, Maryam (2017), Intermediation and voluntary exposure to counterparty risk. Working Paper.

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

Gabaix, Xavier and Augustin Landier (2008), 'Why has CEO Pay Increased So Much?*', *The Quarterly Journal of Economics* **123**(1), 49–100.

García, Diego (2013), 'Sentiment during recessions', *The Journal of Finance* **68**(3), 1267–1300.

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

Gentzkow, Matthew, Bryan T. Kelly and Matt Taddy (2019), 'Text as data', *Journal of Economic Literature* . forthcoming.

Gofman, Michael (2017), 'Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions', *Journal of Financial Economics* **124**(1), 113–146.

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

Herskovic, Bernard, Bryan Kelly, Hanno Lustig and Stijn Van Nieuwerburgh (2019), Firm volatility in granular networks. Working Paper.

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

Hou, Kewei (2007), 'Industry Information Diffusion and the Lead-lag Effect in Stock Returns', *The Review of Financial Studies* **20**(4), 1113–1138.

in 't Veld, Daan and Iman van Lelyveld (2014), 'Finding the core: Network structure in interbank markets', *Journal of Banking & Finance* **49**, 27–40.

Jegadeesh, Narasimhan and Di Wu (2013), 'Word power: A new approach for content analysis', *Journal of Financial Economics* **110**(3), 712 – 729.

Jelveh, Zubin, Bruce Kogut and Suresh Naidu (2018), Political language in economics. Working Paper, Columbia Business School.

Jorion, Philippe and Gaiyan Zhang (2009), 'Credit contagion from counterparty risk', *The Journal of Finance* **64**(5), 2053–2087.

Jurado, Kyle, Sydney C. Ludvigson and Serena Ng (2015), 'Measuring uncertainty', *American Economic Review* **105**(3), 1177–1216.

Kahneman, Daniel and Amos Tversky (1979), 'Prospect theory: An analysis of decision under risk', *Econometrica* **47**(2), 263–291.

Ke, Zheng Tracy, Bryan T. Kelly and Dacheng Xiu (2019), Predicting returns with text data. Working Paper.

Kuhnen, Camelia M. (2015), 'Asymmetric learning from financial information', *The Journal of Finance* **70**(5), 2029–2062.

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.

Liu, Yukun and Ben Matthies (2018), Long run risk: Is it there? Working Paper.

Lo, Andrew W. and A. Craig MacKinlay (1990), 'When Are Contrarian Profits Due to Stock Market Overreaction?', *The Review of Financial Studies* **3**(2), 175–205.

Luttmer, Erzo G. J. (2007), 'Selection, Growth, and the Size Distribution of Firms*', *The Quarterly Journal of Economics* **122**(3), 1103–1144.

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

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard and David Mc-Closky (2005), The Stanford CoreNLP natural language processing toolkit, *in* 'Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations', ACL, pp. 363–370.

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

Niessner, Marina and Eric C. So (2018), Bad news bearers: The negative tilt of financial press. Working Paper.

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

Pinto, Alexandre, Hugo Gonçalo Oliveira and Ana Oliveira Alves (2016), Comparing the Performance of Different NLP Toolkits in Formal and Social Media Text, *in* 'Proceedings of the 5th Symposium on Languages, Applications and Technologies (SLATE '16)', Vol. 51, pp. 3:1–3:16.

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

Shen, Junyan, Jianfeng Yu and Shen Zhao (2017), 'Investor sentiment and economic forces', *Journal of Monetary Economics* **86**, 1 – 21.

Tetlock, Paul C. (2007), 'Giving content to investor sentiment: The role of media in the stock market', *The Journal of Finance* **62**(3), 1139–1168.

Variable	Mean	Std dev.	Min.	Max.
Number of words per article	583	359	19	6658
Number of sentences per article	20.61	13.34	1	253
Number of firms per article	2.36	2.56	0	26
Number of connections identified in an article	1.31	3.68	0	94

Table 1: Summary statistics of news articles in our data set. We consider 106,521 news articles from Reuters financial news published between October 20, 2006, and November 20, 2013.

	Mean	Median	Std dev.	Min.	Max.
Market capitalization (million USD)	8511.26	1593.28	23490.77	5.87	392547.03
Rating		BBB-		D	AAA
Total assets (million USD)	27801.30	1957.49	159333.86	7.42	2554300.77
Total debt (million USD)	7183.25	407.00	45780.22	0.00	998531.73
Book leverage	24.44%	19.62%	23.76%	0.00%	432.01%
Cash holdings (million USD)	990.55	145.35	3795.57	0.00	62439.79
Net income (million USD)	128.18	10.46	532.44	−5056.76	9199.68
Sales (million USD)	1990.50	281.52	6633.45	0.00	116084.71
Cost of goods sold (million USD)	1362.51	157.76	5168.50	0.00	106142.24
Annualized realized volatility	30.91%	28.03%	15.58%	0.06%	143.47%
Quarterly dividends per share (USD)	0.13	0.04	0.22	0.00	2.33
Number of 13-F institutional owners	164.62	109.88	207.10	0.00	1662.33
Institutional ownership ratio	0.52	0.61	0.37	0.00	4.42
Ownership concentration	0.11	0.05	0.16	0.00	1.00
Analyst coverage	6.91	5.16	6.84	0.00	37.81
Number of firms whose stocks are traded in U.S. exchanges			2567 (86.69%)		
Number of firms whose stocks are traded over-the-counter			338 (11.42%)		
Number of private firms			35 (1.18%)		
Number of firms with missing fundamentals data			75 (2.53%)		

Table 2: Summary statistics of firms in our data set. Our sample includes 2,961 distinct firms and spans the time period between October 20, 2006, and November 20, 2013. The above statistics are time series moments over firm lifetimes that overlapped with our sample. Total debt is the sum of current and long-term debt. Book leverage is the ratio of total debt over total assets. We compute the annualized realized volatility of a firm as the standard deviation of monthly log-returns over the sample (adjusted for dividends and stock splits) times the square-root of 12. We take the rating class to be the median S&P Domestic Long Term Issuer Credit Rating over a firm's lifetime. Data on institutional ownership is obtained from Thomson Reuters. We measure the institutional ownership ratio as the fraction of the number of stocks held by institutions over the number of shares outstanding at the end of the quarter. Ownership concentration is measured via the Herfindahl-Hirschman index. Analyst coverage counts the number of analysts following a firm. These data are obtained from I/B/E/S.

	(1)	(2)	(3)
Intercept	*** 0.826	· 0.322	0.413
	(6.118)	(1.919)	(0.560)
Market capitalization	*** 0.119	*** 0.065	* 0.036
(x 10^{-3})	(6.978)	(3.854)	(1.983)
Total assets	0.000	0.003	0.005
(x 10^{-3})	(0.150)	(0.858)	(0.983)
Leverage	· 0.004	* 0.006	0.002
	(1.944)	(2.495)	(0.381)
Volatility	1.375	* 2.630	· 4.110
	(1.202)	(2.196)	(1.778)
Number of 13-F institutional owners		0.001	0.002
		(1.437)	(1.465)
Institutional ownership ratio		−0.192	−0.508
		(−0.865)	(−1.469)
Ownership concentration		−0.313	−0.951
		(−1.047)	(−1.629)
Analyst coverage		*** 0.083	** 0.073
		(5.397)	(3.183)
Rating			0.030
			(0.742)
Data points	2836	2836	1415

Table 3: Estimates of a logit regression of the indicator that a firm was ever linked in our sample based on characteristics of the firm. All characteristics are time-series averages over our sample as summarized in Table 2. The values in parentheses give z-statistics. ***, **, *, and · denote significance on the 99.9%, 99%, 95%, and 90% confidence levels, respectively.

Link	Representative sentence
(Chrysler, GM)	*"Larger rival General Motors Corp. reported 6 percent sales growth, while Chrysler Group posted a 3 percent rise, breaking a nine-month losing streak."*
(Ford, GM)	*"GM, like Ford Motor Co and privately held Chrysler LLC, reached an agreement with the UAW that allows it to hire new workers for some jobs starting at $14 per hour, or about half the current average hourly wage."*
(Goldman, Morgan Stanley)	*"Goldman and Morgan Stanley are also locked in discussions with the Federal Reserve over their right to keep owning and operating physical commodity assets."*
(B of A, Citigroup)	*"According to Peabody's estimates, Bank of America has a higher percentage of residential mortgage-backed securities in its available-for-sale portfolio than JPMorgan and Citigroup."*
(B of A, Merrill Lynch)	*"Bank of America Corp on Sunday is in advanced talks to acquire Merrill Lynch & Co Inc, a deal that would give the bank the world's largest brokerage and a sizable investment bank."*
(Ford, Chrysler)	*"Having sold fewer cars and light trucks in November than Chrysler overall, Ford slashed its planned production."*
(Microsoft, Google)	*"Microsoft, the world's largest software company has stepped up its efforts within its online services division – which lost $2.3 billion last fiscal year – to challenge the dominance of Google, the world's largest search engine."*
(Apple, Google)	*"Chief Executive Eric Schmidt signaled that Google, the world's dominant Web search provider, is working more closely with Apple Inc., the pace setter in the consumer electronics world."*
(JPMorgan, Citigroup)	*"On Monday, Citigroup, JPMorgan Chase and Wells Fargo Corp said they plan to sell covered bonds to help kickstart the new U.S. market."*
(JPMorgan, B of A)	*"IntraLase's lead financial adviser is Bank of America and JPMorgan is co-financial adviser."*
(JPMorgan, Goldman)	*"Goldman Sachs and JPMorgan & Chase & Co., for instance trade many hedged strategies through their giant proprietary trading desk and hedge funds, which few outsiders see."*
(Citigroup, Morgan Stanley)	*"Citigroup, Credit Suisse, Deutsche Bank, Goldman Sachs, Merrill Lynch, Morgan Stanley and UBS said in a statement they will form a new company with an independent management team to develop the trading platform."*
(Vodafone, Verizon)	*"Verizon Wireless, a 55 percent owned joint venture with Vodafone Group Plc, added 2.3 million net customers in the fourth quarter."*
(JPMorgan, Bear Stearns)	*"All eyes will again be on Bear Stearns after news on Sunday that the fifth-largest U.S. investment bank was close to selling itself to JPMorgan Chase & Co."*
(AT&T, Verizon)	*"Even as competitors AT&T Inc. and Verizon Communications Inc. have started their own TV offerings to compete with cable, Comcast reported record growth in digital video, phone and Internet subscribers."*
(Toyota, GM)	*"GM argued that it needed to cut the nearly $5 billion per year it spends on health care to compete against Japanese rivals led by Toyota Motor Corp."*
(Citigroup, Goldman Sachs)	*'The equity investment for the transaction will be contributed by the investors, including Becker, and debt financing will be provided by Citigroup and Goldman Sachs."*
(Daimler, Chrysler)	*"The German company Daimler AG will contribute another 650 million euros to cover long-term liabilities at Chrysler."*
(Sprint, Clearwire)	*"Sprint and Clearwire also said last month they were continuing to talk even after they announced late last year that they ditched an agreement to let customers roam between both companies' WiMax networks."*
(BHP, Potash)	*"Canada's Potash Corp spurned an unsolicited $38.6 billion takeover offer by BHP Billiton, raising the prospect that the Anglo-Australian miner could launch a hostile bid to take the top position in the global fertilizer industry."*

Table 4: Sample sentences in which our methodology recognizes links between two firms. The links are sorted according to the frequency with which they are identified in the text data, with the most frequently identified link at the top.

		Mean	Median	Std dev.	Min.	Max.
Market capitalization (million USD)	Larger Firm	15570.82	4424.87	31798.58	15.14	392547.03
	Smaller Firm	10101.64	1986.06	24773.00	6.29	258664.37
Rating	Larger Firm		BBB		D	AAA
	Smaller Firm		BBB-		D	AAA
Total assets (million USD)	Larger Firm	52942.84	5627.93	224069.07	18.35	2554300.77
	Smaller Firm	34368.44	2467.77	179692.54	9.61	2554300.77
Total debt (million USD)	Larger Firm	13611.07	1357.11	64442.71	0.00	998531.73
	Smaller Firm	8931.73	549.49	51719.64	0.00	998531.73
Book leverage	Larger Firm	26.02%	22.23%	22.62%	0.00%	321.73%
	Smaller Firm	24.83%	19.79%	23.73%	0.00%	432.01%
Cash holdings (million USD)	Larger Firm	1790.06	388.25	5182.17	0.00	62439.79
	Smaller Firm	1194.19	189.24	4226.21	0.00	62439.79
Net income (million USD)	Larger Firm	240.37	40.09	737.23	−5056.76	9199.68
	Smaller Firm	151.67	14.56	561.83	−5056.76	6992.88
Sales (million USD)	Larger Firm	3635.99	815.52	9107.33	0.00	116084.71
	Smaller Firm	2382.79	366.37	7165.25	0.00	116084.71
Cost of goods sold (million USD)	Larger Firm	2456.88	448.33	7104.52	0.00	106142.24
	Smaller Firm	1628.23	210.28	5593.35	0.00	106142.24
Annualized realized volatility	Larger Firm	28.01%	24.89%	13.56%	0.08%	99.41%
	Smaller Firm	30.62%	27.62%	15.34%	0.08%	143.47%
Quarterly dividends per share (USD)	Larger Firm	0.17	0.08	0.24	0.00	1.77
	Smaller Firm	0.14	0.04	0.22	0.00	2.23
Distinct linkswith both firms in same industry				4093 (19.96%)		
Distinct linkswith both firms in same sector				8177 (39.88%)		

Table 5: Summary statistics of linked firms in our data set. Our sample includes 20,504 distinct links between 2,406 firms and spans the time period between October 20, 2006, and November 20, 2013. The above statistics are time series moments over firm lifetimes that overlapped with our sample period. Total debt is the sum of current and long-term debt. Book leverage is the ratio of total debt over total assets. We compute the annualized realized volatility of a firm as the standard deviation of monthly log-returns over the sample times the square-root of 12. The rating class of a firm is the median S&P Domestic Long Term Issuer Credit Rating over a firm's lifetime.

Sector-Sector pair	# unique links	Sample sentence
(Cons. Discretionary, Financials)	1138	*“U.S. investment banks are helping Cerberus Capital Management LP finance its buyout of Chrysler.”*
(IT, Financials)	975	*“IBM, which had owned about 15 percent of the Chinese company, sold the shares at a discount of 7 percent to the stock’s Monday close, in a deal handled by Citigroup Inc.”*
(Industrials, Financials)	867	*“GE has told private-equity firms that they face restrictions on their ability to team up with other private-equity bidders, citing people familiar with the sale effort, and that Goldman Sachs is acting on GE’s behalf for the sale.”*
(Health Care, Financials)	623	*“Express Scripts said it had received commitments from Citigroup and Credit Suisse to fully finance the proposed transaction.”*
(Communication Serv., Financials)	600	*‘Most employees will be able to sell vested options granted after Google went public in 2004 to qualified financial institutions via a private auction managed by Morgan Stanley and Citigroup.”*
(Cons. Staples, Financials)	563	*“Procter & Gamble gained 1.4 percent to $65.60 after Goldman Sachs raised its rating on the stock to a ”buy.””*
(Cons. Discretionary, IT)	558	*“But IBM faces an unlikely challenger in Amazon.com Inc, the e-commerce retail giant that is becoming a force in the booming business of cloud computing, even winning backing from America’s top spy agency.”*
(Communication Serv., IT)	516	*“An Apple’s lawsuit against Google’s Motorola Mobility unit over alleged patent abuse was thrown out on Monday just hours before trial, a setback for the iPhone maker in its efforts to gain leverage in the smartphone patent wars.”*
(Cons. Discretionary, Cons. Staples)	432	*“Kraft Foods Inc said on Tuesday it struck deals for Starbucks coffee to be sold for its Tassimo hot beverage machine.’*
(Energy, Financials)	432	*“JPMorgan Chase & Co was ordered to restore to the trust 220,122 shares of Exxon Mobil lost under the VPF investment strategy.”*
(Industrials, Cons. Discretionary)	429	*“U.S. aircraft maker Boeing Co. confirmed an order for 27 Boeing 767-300 Freighters from package delivery company United Parcel Service Inc.”*
(Cons. Discretionary, Communication Serv.)	425	*“Top PC makers including Dell Inc. and Apple Inc. are recalling up to 9.6 million Sony batteries, which on rare occasions could overheat and catch fire.”*
(Materials, Financials)	423	*“BHP has already dispatched advisers Citigroup and Goldman Sachs to arrange $70 billion in debt refinancing with a small number of banks to help pay for a takeover.”*
(Industrials, IT)	362	*“Nokia and Siemens merged their equipment units in 2007, in search of critical mass, as did Alcatel and Lucent in 2006.”*
(Real Estate, Financials)	217	*“It wasn’t clear why Mack-Cali pulled out of the deal two days after it announced that it would join Rome Acquisition L.P., the entity billionaire Icahn and New York real estate magnate Macklowe formed to takeover Reckson, a real estate investment trust (REIT) that owns office buildings in New York City and its suburbs.”*
(Health Care, IT)	212	*“British drug maker GlaxoSmithKline Plc has said it is working with IBM on the radio frequency identification devices (RFID) project.”*
(Industrials, Communication Serv.)	203	*“Ericsson, Alcatel-Lucent and Nokia Siemens supply AT&T with wireless network equipment.”*
(Industrials, Cons. Staples)	192	*“Wal-Mart already offers a number of financial services in its stores, like a credit card in partnership with General Electric and money transfers through MoneyGram.”*
(Industrials, Health Care)	181	*“Bayer has also sold its diagnostics business to engineering conglomerate Siemens for 4.2 billion euros to help pay for Schering.”*
(Cons. Staples, Health Care)	178	*“P&G, whose lineup includes Gillette razors and Olay skin creams, also announced a new joint venture with Teva Pharmaceutical to sell over-the-counter medicines.”*

Table 6: List of the 20 most frequently identified cross-sectoral links together with sample sentences in which our methodology recognizes the links. We aggregate each firm to sectors according to the GICS sector code.

RIC mentions	77080
RIC mentions identified as organizations by Step 1	64525
RIC mentions also identified as firms by Step 2	64525
RIC mentions correctly matched in CRSP / Compustat by Step 3	56458
RIC mentions incorrectly matched in CRSP / Compustat by Step 3	2121
RIC mentions not matched in CRSP / Compustat by Step 3	5946
Number of RIC that are always incorrectly matched	0
Number of RIC that are never matched in CRSP / Compustat	0

Table 7: Accuracy analysis for our methodology. We collect all mentions of Reuters Instrument Codes (RIC) in our news data. We then match the tickers implied by the RIC to tickers in the CRSP / Compustat database. The data sample includes 1,858 distinct firms.

	Mean	Median	Std dev.	Min.	Max.
Market capitalization (million USD)	10325.16	2249.50	26611.88	5.85	391914.93
Rating		BBB-		CCC	AAA
Total assets (million USD)	32320.85	2820.11	172817.73	5.84	2650692.69
Total debt (million USD)	9188.13	646.26	60128.42	0.00	1125964.27
Book leverage	25.07%	20.33%	24.26%	0.00%	355.29%
Cash holdings (million USD)	1074.28	191.93	3915.14	0.00	61895.07
Net income (million USD)	150.92	20.00	650.47	−9658.00	9142.50
Sales (million USD)	2352.71	444.26	7249.88	0.00	118561.95
Cost of goods sold (million USD)	1593.71	246.12	5548.17	0.00	106142.24
Annualized realized volatility	28.30%	26.44%	12.16%	0.52%	101.24%
Quarterly dividends per share (USD)	0.15	0.06	0.22	0.00	1.84
Number of firms that only operate domestically			1446 (77.83%)		
Number of firms that operate internationally			404 (21.74%)		
Number of firms that do not operate domestically			0 (0.00%)		
Number of firms whose stocks are traded in U.S. exchanges			1812 (97.52%)		
Number of firms whose stocks are traded over-the-counter			28 (1.51%)		
Number of private firms			8 (0.43%)		
Number of firms with missing fundamentals data			6 (0.32%)		

Table 8: Summary statistics of firms with Reuters Instrument Codes (RIC). We identify all RIC in Reuters financial news published between between October 20, 2006, and November 20, 2013. We match firms with RIC with firms in the merged CRSP / Compustat database, from where we obtain fundamentals data. The sample includes 1,858 distinct firms. The above statistics are time series moments over firm lifetimes that overlapped with our sample period. Total debt is the sum of current and long-term debt. Book leverage is the ratio of total debt over total assets. We compute the annualized realized volatility of a firm as the standard deviation of monthly log-returns over the sample times the square-root of 12. The rating class of a firm is the median S&P Domestic Long Term Issuer Credit Rating over a firm's lifetime.

	Average degree			First-order interconnectivity			Second-order interconnectivity		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	*** 9.674	*** 10.992	*** 13.080	*** 1.619	*** 1.982	*** 2.258	*** 0.051	*** 0.055	*** 0.062
	(5.058)	(5.503)	(6.873)	(4.958)	(5.861)	(6.663)	(4.878)	(5.267)	(6.237)
Lagged variable	*** 0.494	*** 0.421	* 0.234	** 0.533	*** 0.425	** 0.310	*** 0.479	*** 0.429	* 0.245
	(5.160)	(4.169)	(2.258)	(5.773)	(4.416)	(3.061)	(4.963)	(4.384)	(2.351)
Average article sentiment		· −11.948	2.096		** −2.054	−0.877	−0.093	−0.093	0.021
		(−1.958)	(0.318)		(−2.851)	(−1.086)		(−1.981)	(0.384)
Recession indicator			*** 7.254			** 0.605	−0.093		*** 0.054
			(4.040)			(2.822)			(3.638)
Number of observations	84	84	84	84	84	84	84	84	84
Adjusted R^2	0.236	0.261	0.379	0.280	0.338	0.390	0.222	0.248	0.347

Table 9: Regressions of the average degree, the first-order interconnectivity, and the second-order interconnectivity on lagged values of themselves, the average article sentiment, and the NBER recession indicator. The time series are monthly. We construct our measure of sentiment article by article. For each article, we use the sentiment annotator in the coreNLP toolkit to evaluate the sentiment of each sentence and then take the average across the sentiment of all sentences in an article. For each month, we compute an average article sentiment measure as the average sentiment across all articles in that month. We standardize the monthly average measures using the full same mean and standard deviation. We construct a recession indicator from the NBER recession dates. The values in parentheses give t-statistics. ***, **, *, and · denote significance on the 99.9%, 99%, 95%, and 90% confidence levels, respectively.

		(1)	(2)	(3)	(4)	(5)	(6)
Firm 1	$(\text{Return})_+$	0.141				0.079	0.091
		(1.273)				(0.577)	(0.473)
	$(\text{Return})_-$	* −0.209				* −0.218	−0.173
		(−2.471)				(−2.111)	(−1.158)
	Volatility	*** 1.509				*** 1.382	*** 1.353
		(12.965)				(9.460)	(6.936)
	Upgrade dummy		0.113			−0.055	0.106
			(1.425)			(−0.485)	(0.725)
	Downgrade dummy		*** 0.492			*** 0.219	· 0.179
			(10.941)			(3.319)	(1.956)
	Revisions up			** 0.006			0.004
				(3.074)			(1.127)
	Revisions down			*** 0.010			*** 0.013
				(5.607)			(3.878)
	Earnings surprise			· −0.023			0.017
				(−1.773)			(0.780)
	$(\text{Net income})_+$				· −13.616	−16.497	21.525
	(x 10^{-6})				(−1.788)	(−1.460)	(1.305)
	$(\text{Net income})_-$				*** −28.365	*** −20.542	** −32.026
	(x 10^{-6})				(−7.917)	(−4.478)	(−2.889)
	Market cap	*** 3.149	*** 2.368	*** 2.610	*** 2.849	*** 3.527	*** 2.843
	(x 10^{-6})	(12.325)	(9.781)	(10.927)	(11.739)	(9.951)	(6.018)
	Analyst coverage	0.002	· −0.003	−0.001	0.001	0.000	−0.003
		(0.880)	(−1.839)	(−0.591)	(0.476)	(−0.167)	(−0.937)
Firm 2	$(\text{Return})_+$	0.036				0.010	−0.013
		(0.326)				(0.071)	(−0.065)
	$(\text{Return})_-$	· −0.141				· −0.181	−0.003
		(−1.646)				(−1.727)	(−0.017)
	Volatility	*** 1.388				*** 1.244	*** 1.336
		(12.038)				(8.202)	(6.174)
	Upgrade dummy		* 0.168			0.070	0.169
			(2.247)			(0.681)	(1.239)
	Downgrade dummy		*** 0.453			** 0.198	0.143
			(10.050)			(2.962)	(1.551)
	Revisions up			*** 0.007			0.004
				(3.722)			(1.272)
	Revisions down			*** 0.009			*** 0.011
				(5.499)			(3.792)
	Earnings surprise			* −0.029			0.012
				(−1.975)			(0.445)
	$(\text{Net income})_+$				−8.624	−0.454	9.358
	(x 10^{-6})				(−1.074)	(−0.040)	(0.555)
	$(\text{Net income})_-$				*** −26.412	*** −18.691	*** −42.598
	(x 10^{-6})				(−7.579)	(−4.871)	(−3.821)
	Market cap	*** 3.067	*** 2.396	*** 2.616	*** 2.767	*** 3.037	*** 2.987
	(x 10^{-6})	(12.239)	(9.879)	(10.963)	(11.110)	(8.367)	(6.313)
	Analyst coverage	· 0.003	−0.003	−0.001	0.001	0.001	−0.003
		(1.745)	(−1.601)	(−0.491)	(0.613)	(0.574)	(−0.873)
Link-month obs.		118002	148638	135209	249067	81416	48479
Positive obs.		18506	25311	20527	37233	13734	8133
Unique links		2593	2211	2211	2949	1927	1156
Unique firms		721	587	587	874	525	361

Table 10: Logit regressions of link dummies, which indicate whether in a given month we identified at least one link between two firms. We consider the 3,000 most frequently observed links in the data. For every link-month observation, we randomly assign the labels “Firm 1” and “Firm 2” among the two linked firms but keep the link name fixed. We compute upgrades and downgrades relative to the S&P Domestic Long Term Issuer Credit Rating of a firm in a given month. Revisions down and up count the number of analysts that altered their estimates down or up in a given month. Earnings surprise is the difference between the realized EPS and the average EPS estimate across all forecasters in the current quarter. Descriptions and summary statistics of the aggregate control factors are given in Table 11. Albeit unreported, each regression includes an intercept as well dummies for whether the two linked firms are in the same industry or sector. We also control for the monthly return of the S&P 500, the level of the VIX, the monthly industrial production growth rate, as well as a control for the level of the yield curve. The estimates for these controls are available upon request. All regressions include time and link fixed effects. The values in parentheses give z-statistics. Standard errors are based on sandwich estimators clustered by link and month. ***, **, *, and · denote significance on the 99.9%, 99%, 95%, and 90% confidence levels, respectively.

Variable	Mean	Std dev.	Min.	Max.
S&P 500 return	0.417	4.398	-20.395	12.022
VIX	22.736	10.324	10.820	62.640
VXO	22.552	10.995	10.484	65.447
AAA credit spread	3.670	1.616	0.040	5.730
BAA credit spread	4.949	1.911	0.950	8.820
Level of yield curve	0.990	1.645	0.010	5.030
Slope of yield curve	1.942	1.023	-0.480	3.430
GDP growth	0.707	0.781	-1.900	1.400
Industrial production growth	0.014	0.865	-4.300	1.400
Aggregate default rate	0.178	0.195	0.000	0.935
Contagion factor	43.931	24.608	14.805	113.516
Aggregate default probability	-8.010	0.910	-9.511	-5.541
Aggregate distance-to-default	4.037	1.084	1.754	5.805
Financial uncertainty (Jurado et al. (2015))	0.961	0.224	0.634	1.546
Financial uncertainty (Carriero et al. (2018))	0.961	0.224	0.634	1.546
Equity market volatility index	0.961	0.224	0.634	1.546

Table 11: Summary statistics of our financial and macroeconomic factors. All factors are sampled at the monthly frequency. We measure S&P 500 returns from open at the start of the month to close at the end of the month. The VIX (VXO) in a given month is evaluated as the average VIX (VXO) observed during that month. The BAA and AAA credit spreads correspond to the Moody's Seasoned corporate bond yields minus the Federal Funds Rate. The level of the yield curve is measured via the 3-month Treasury Bill secondary rate, while the slope is constructed as the spread between the fixed-maturity yields of the 10-year and the 1-year Treasury Bills. Both level and slope are evaluated at the end of a month. GDP growth is measured from quarter to quarter. We obtain a monthly GDP growth rate time series by interpolating with the most recent quarterly observation. Industrial production growth is given by the month-to-month growth rate in the industrial production index of the Board of Governors of the Federal Reserve System. Data on the above factors are obtained from the St. Louis Fed's FRED database. All macroeconomic time series are seasonally adjusted annualized rates. We compute a nonparametric measure of the aggregate default rate in the U.S. economy as the ratio of the number of observed defaults in a month over the number of days in that month. We use the same historical default timing data as in Azizpour et al. (2018), which is obtained from Moody's Default Risk Service and covers the years 1970 through 2012. The contagion factor is due to Azizpour et al. (2018) and measures the component of the conditional arrival rate of defaults that is due to contagion of credit risk across linked counterparties. It is estimated semiparametrically from default timing data via an autoregressive model and covers our sample through the year 2012. Data on the one-month aggregate default probability as well as the distance-to-default of U.S. firms is obtained from the Credit Research Initiative at the National University of Singapore (`https://www.rmicri.org/en/`). We assume that default probabilities and distance-to-default are estimated at the start of a month with the concurrently available data from the end of the previous month. We consider a logistic transformation of the default probability. We obtain data on the financial uncertainty measure of Jurado et al. (2015) from Sydney Ludvigson's website (`https://www.sydneyludvigson.com`), the financial uncertain measure of Carriero et al. (2018) from the Review of Economics and Statistics data replication website (`https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ENTXDD`), and the equity market volatility (EMV) index of Baker et al. (2019) from the Economic Policy Uncertainty website (`https://www.policyuncertainty.com/EMV_monthly.html`).

		Popularity proxy			
		Market capitalization		Institutional investors	
		(1)	(2)	(1)	(2)
More popular firm	$(\text{Return})_{+}$	0.206	0.052	0.239	0.264
		(1.272)	(0.231)	(1.251)	(1.227)
	$(\text{Return})_{-}$	0.005	0.055	−0.024	0.098
		(0.040)	(0.291)	(−0.147)	(0.521)
	Volatility	*** 1.402	*** 1.623	*** 1.524	*** 1.554
		(6.741)	(5.888)	(6.112)	(5.647)
	Upgrade dummy	−0.024	0.109	−0.141	−0.026
		(−0.212)	(0.711)	(−0.982)	(−0.168)
	Downgrade dummy	0.003	−0.042	0.056	0.054
		(0.035)	(−0.409)	(0.649)	(0.563)
	Revisions up		0.004		0.002
			(1.052)		(0.737)
	Revisions down		· 0.006		0.005
			(1.814)		(1.553)
	Earnings surprise		−0.008		−0.035
			(−0.264)		(−0.858)
	$(\text{Net income})_{+}$	** −37.178	· −27.005	** −23.868	−9.099
	(x 10^{-6})	(−2.935)	(−1.650)	(−1.522)	(−0.514)
	$(\text{Net income})_{-}$	*** −25.427	* −29.053	*** −33.241	* −26.949
	(x 10^{-6})	(−4.568)	(−2.285)	(−3.040)	(−2.298)
	Market cap	*** 1.726	** 1.794	*** 2.329	*** 2.199
	(x 10^{-6})	(3.669)	(3.016)	(4.530)	(3.905)
	Analyst coverage	0.002	0.004	0.000	0.002
		(0.506)	(0.782)	(0.026)	(0.462)
Less popular firm	$(\text{Return})_{+}$	−0.011	0.106	−0.026	−0.082
		(−0.091)	(0.597)	(−0.152)	(−0.443)
	$(\text{Return})_{-}$	*** −0.408	· −0.272	* −0.320	· −0.295
		(−3.939)	(−1.768)	(−2.129)	(−1.836)
	Volatility	*** 1.291	*** 1.206	*** 1.049	*** 1.191
		(7.783)	(5.261)	(4.442)	(4.891)
	Upgrade dummy	0.053	0.181	* 0.226	* 0.265
		(0.525)	(1.431)	((1.995)	(2.120)
	Downgrade dummy	*** 0.353	*** 0.297	*** 0.300	** 0.236
		(5.662)	(3.573)	(3.720)	(2.714)
	Revisions up		0.006		0.007
			(1.475)		(1.492)
	Revisions down		*** 0.019		*** 0.020
			(4.957)		(4.977)
	Earnings surprise		0.025		0.045
			(0.829)		(1.569)
	$(\text{Net income})_{+}$	* 35.021	*** 119.493	23.667	* 58.738
	(x 10^{-6})	(2.115)	(4.954)	(1.278)	(2.391)
	$(\text{Net income})_{-}$	*** −16.995	*** −48.336	** −24.610	*** −48.544
	(x 10^{-6})	(−4.051)	(−3.834)	(−3.205)	(−4.014)
	Market cap	*** 7.515	*** 5.857	*** 4.776	*** 4.679
	(x 10^{-6})	(9.066)	(5.417)	(5.851)	(4.805)
	Analyst coverage	0.000	· −0.009	* 0.009	−0.007
		(0.134)	(−1.927)	(2.102)	(−1.301)
Link-month obs.		81416	48479	59465	48185
Positive obs.		13734	8133	9682	8086
Unique links		1927	1156	1415	1149
Unique firms		525	361	411	356

Table 12: Logit regressions of link dummies, which indicate whether in a given month we identified at least one link between two given firms. We consider the 3,000 most frequently observed links in the data. For every link-month observation, we call the more (less) popular of the two linked firms "Firm 1" ("Firm 2"), where popularity is measured either by the market capitalization or the number of 13-F institutional investors of a firm in that month. Otherwise, we keep the link name fixed. Descriptions and summary statistics of all factors are given in Tables 11 and 10. Albeit unreported, each regression includes an intercept as well dummies for whether the two linked firms are in the same industry or sector. We also control for the monthly return of the S&P 500, the level of the VIX, the monthly industrial production growth rate, as well as a control for the level of the yield curve. The estimates for these controls are available upon request. All regressions include time and link fixed effects. The values in parentheses give z-statistics. Standard errors are based on sandwich estimators clustered by link and month. ***, **, *, and · denote significance on the 99.9%, 99%, 95%, and 90% confidence levels, respectively.

	VIX	VXO	Financial uncertainty (Jurado et al. (2015))	Financial uncertainty (Carriero et al. (2018))	EMV index (Baker et al. (2019))
Intercept	* −3.835	* −4.407	0.022	−0.034	−0.237
	(−2.030)	(−2.270)	(1.002)	(−0.393)	(−0.084)
Lagged value	*** 0.735	*** 0.721	*** 0.935	*** 0.759	*** 0.379
	(11.165)	(10.735)	(36.236)	(11.014)	(4.031)
News sentiment	3.843	3.191	* −0.117	−0.149	−4.563
	(0.672)	(0.522)	(−2.378)	(−0.525)	(−0.526)
Average degree	*** 0.521	*** 0.563	* 0.002	*** 0.016	*** 0.740
	(5.599)	(5.637)	(2.595)	(3.618)	(5.100)
First-order interconnectivity	0.834	0.934	−0.013	0.022	−0.179
	(0.792)	(0.839)	(−1.568)	(0.510)	(−0.126)
Second-order interconnectivity	−37.142	−36.772	0.194	* −1.884	58.903
	(−1.583)	(−1.481)	(0.972)	(−1.766)	(1.578)
Number of observations	84	84	84	84	84
Adjusted R^2	0.819	0.821	0.972	0.825	0.497

Table 13: Regressions of several measures of financial uncertainty on their lagged values and contemporaneous values of the connectivity measures. The time series are monthly. Table 11 provides summary statistics for all factors. The values in parentheses give t-statistics. ***, **, *, and $^{\cdot}$ denote significance on the 99.9%, 99%, 95%, and 90% confidence levels, respectively.

	Average degree	First-order IC	Second-order IC
Intercept	** 84.617	*** 10.218	0.236
	(2.992)	(3.797)	(0.957)
Segments network	−0.001	−0.151	−0.146
	(−0.266)	(−0.492)	(−0.638)
Hoberg and Phillips (2016) network	* −3.263	** −0.345	* −1070.403
	(−2.316)	(−2.725)	(−2.041)
Variance decomposition network	−0.032	0.004	−2.628
	(−0.996)	(0.257)	(−0.236)
BEA input network	−0.620	−0.293	* −0.028
	(−0.311)	(−1.489)	(−2.241)
Number of observations	71	71	71
Adjusted R^2	0.085	0.164	0.078

Table 14: Regressions of our connectivity measures on the analogous connectivity measures of the alternative networks. The time series are monthly. "IC" stands for interconnectivity. The values in parentheses give t-statistics. ***, **, *, and · denote significance on the 99.9%, 99%, 95%, and 90% confidence levels, respectively.

		AAA credit spread		BAA credit spread		S&P 500		VIX		GDP growth	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
	Intercept	0.300	0.423	−0.499	−0.278	−0.068	15.535	32.520	35.259	−1.158	−1.411
		(0.105)	(0.173)	(−0.122)	(−0.083)	(−0.002)	(0.412)	(0.750)	(0.993)	(−0.325)	(−0.469)
	Lagged value	*** 0.481	*** 0.477	*** 0.668	*** 0.610	0.241	0.179	*** 0.742	*** 0.710	*** 0.814	*** 0.842
		(5.324)	(6.041)	(7.067)	(7.596)	(1.562)	(1.202)	(6.298)	(6.680)	(9.622)	(8.965)
	Lagged GDP growth	· −0.107	−0.076	−0.144	· −0.142	0.480	0.479	−1.151	0.486		
		(−1.741)	(−1.324)	(−1.567)	(−1.750)	(0.409)	(0.414)	(−0.799)	(−0.928)		
	Lagged S&P 500 return	−0.011	−0.007	* −0.025	−0.016			−0.038	0.014		
		(−1.393)	(−0.934)	(−2.096)	(−1.508)			(−0.922)	(−0.240)		
	Lagged VIX	−0.003	−0.005	0.004	0.007	0.042	0.034				
		(−0.521)	(−0.804)	(0.402)	(0.735)	(0.383)	(0.315)				
	Lagged yield curve level	−0.205	−0.138	−0.233	−0.158	1.040	−0.747				
		(−1.529)	(−1.188)	(−1.280)	(−1.056)	(0.571)	(−0.433)				
	Lagged yield curve slope	*** 0.475	*** 0.409	· 0.282	· 0.218						
		(4.264)	(4.125)	(1.945)	(1.756)						
	Lagged sentiment	−0.387	−0.310	−0.208	−0.114	−0.686	−2.812	4.180	8.272	−0.274	−0.594
		(−0.773)	(−0.716)	(−0.283)	(−0.186)	(−0.071)	(−0.321)	(0.364)	(0.877)	(−0.282)	(−0.721)
News	Lagged average degree		*** 0.021		** 0.027		−0.187		* 0.327		−0.010
			(3.725)		(3.351)		(−1.597)		(2.629)		(−0.947)
	Lagged 1st-order IC		−0.013		−0.056		−0.167		−0.635		0.080
			(−0.238)		(−0.723)		(−0.149)		(−0.522)		(0.826)
	Lagged 2nd-order IC		* 2.608		*** 5.812		** −82.610		*** 113.676		*** −11.546
			(2.212)		(3.543)		(−3.375)		(4.323)		(−4.996)
Segments	Lagged average degree	0.000	0.000	0.000	0.000	0.001	−0.001	0.001	0.000	0.000	0.000
		(0.104)	(1.221)	(−0.561)	(0.434)	(0.293)	(−0.410)	(0.693)	(0.198)	(−1.018)	(−0.509)
	Lagged 1st-order IC	−0.058	0.296	−0.331	0.193	1.880	−2.479	2.183	2.846	−0.102	−0.132
		(−0.251)	(1.403)	(−0.990)	(0.664)	(0.422)	(−0.586)	(0.975)	(1.550)	(−0.589)	(−0.889)
	Lagged 2nd-order IC	· 6.315	· 4.408	8.755	7.549	−4.293	2.132	−23.009	−25.632	2.912	0.875
		(1.756)	(1.404)	(1.621)	(1.674)	(−0.062)	(0.033)	(−0.277)	(−0.370)	(0.411)	(0.144)
V-D	Lagged average degree	* 0.001	0.000	0.001	0.000	−0.014	0.005	0.022	0.007	−0.003	−0.001
		(0.592)	(−0.020)	(0.667)	(−0.093)	(−0.579)	(0.215)	(0.802)	(0.303)	(−1.199)	(−0.631)
	Lagged 1st-order IC	* 0.010	* 0.008	· 0.012	· 0.011	0.032	0.013	−0.032	−0.040	−0.011	−0.012
		(2.213)	(2.005)	(1.926)	(1.938)	(0.361)	(0.158)	(−0.308)	(−0.449)	(−1.265)	(−1.598)
	Lagged 2nd-order IC	0.010	−0.021	0.069	0.031	−2.175	−2.279	1.233	1.618	−0.072	−0.098
	(x 10^3)	(−0.259)	(0.115)	(0.526)	(0.269)	(−1.231)	(0.603)	(0.506)	(0.931)	(−0.415)	(−0.642)
H-P	Lagged average degree	0.051	−0.007	0.129	0.060	−0.228	−0.408	−1.890	−2.015	0.129	0.121
		(0.381)	(−0.065)	(0.660)	(0.371)	(−0.108)	(−0.214)	(−0.758)	(−0.985)	(0.617)	(0.683)
	Lagged 1st-order IC	0.011	−0.058	−0.067	−0.183	7.812	7.648	−10.673	· −12.966	1.092	* 1.321
		(0.027)	(−0.156)	(−0.106)	(−0.346)	(1.035)	(1.106)	(−1.205)	(−1.767)	(1.448)	(2.048)
	Lagged 2nd-order IC	−4.310	−5.651	−6.800	−8.536	102.288	87.301	−120.211	· −142.381	8.638	9.632
	(x 10^3)	(−1.300)	(−0.855)	(−0.928)	(−1.400)	(1.311)	(1.236)	(−1.379)	(−1.996)	(1.165)	(1.541)
	Number of observations	83	83	83	83	83	83	83	83	83	83
	Out-of-sample R^2	0.982	0.988	0.974	0.983	0.155	0.341	0.770	0.853	0.703	0.799
	F-statistic		*** 9.019		*** 11.555		** 6.105		*** 12.468		*** 10.759

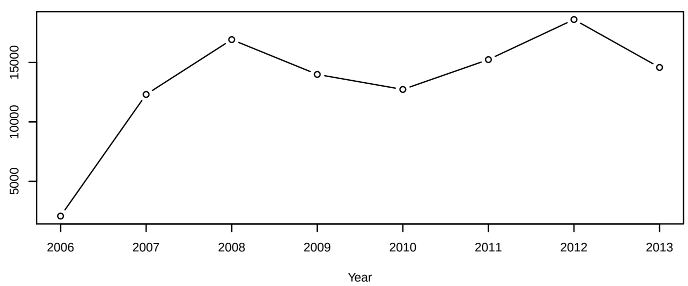
Table 15: Predictive regressions for corporate credit spreads, S&P 500 returns, VIX, and GDP growth rates on lagged values of the connectivity measures of competing firm networks and several controls. "IC" stands for interconnectivity, "V-D" for variance decomposition, and "H-P" for Hoberg and Phillips (2016). For each network, we orthogonalize the vector of average degree, first-order, and second-order interconnectivity. The sampling horizon is monthly. The values in parentheses give t-statistics. The F-statistic is for a test of the null hypothesis that a model that includes connectivity measures of the news-implied network predicts equally well as a model that does not control for news-implied connectivity. ***, **, *, and · denote significance on the 99.9%, 99%, 95%, and 90% confidence levels, respectively.

```
-- GM-UAW contract seen hard to match fully by rivals
-- By  David Bailey
-- Tue Oct 2, 2007 3:47pm EDT
-- http://www.reuters.com/article/2007/10/02/us-gm-uaw-idUSN0242907020071002


 DETROIT  (Reuters) - Several aspects of the tentative contract between General Motors Corp ( GM.N ) and the United Auto Workers union will be hard for Ford Motor Co.
( F.N ) and Chrysler LLC to match in labor talks expected to heat up in coming days, people familiar with the negotiations said.

 The adoption of second-tier wages for new hires at GM represents an attractive concession for Ford and Chrysler, but the structure of a retiree health-care trust
could prove difficult to transfer, sources familiar with the matters said on Tuesday. The establishment of a Voluntary Employees Beneficiary Association trust, or
VEBA, was a centerpiece of the UAW's agreement with GM, allowing the automaker to take some $50 billion of liabilities off its books. Privately held Chrysler has been
focused on cash flow since Cerberus acquired the automaker over the summer, to the point that it has been taking daily cash flow reports. The GM-UAW health-care trust
would not provide savings until 2010, when the new trust is expected to take over some $3 billion in annual retiree health care payments from the top U.S. automaker.
Ford and Chrysler would be hard-pressed to match the bump-up in pension payments to their retirees that GM has agreed to give to its UAW retirees under the tentative
contract, people familiar with the talks said. UAW President Ron Gettelfinger said on Friday he expected to assess the state of talks with both Ford and privately held
Chrysler after local UAW leaders unanimously recommended that workers approve the GM contract. Gettelfinger wants the agreement with GM to serve as a basic pattern for
talks with Ford and Chrysler in keeping with a long-held tradition that has kept all three Detroit-based automakers on a similar labor-cost footing. The union's deal
with GM includes a second-tier wage for new hires outside the production line, a health-care trust for retirees and some job security. UAW VOTES ON GM CONTRACT
CONTINUE The UAW may not resume full negotiations with Ford or Chrysler until it completes the ratification, or has enough of a favorable indication from the voting at
GM locals first, one person close to the talks said. Subcommittees for the UAW and Chrysler had been meeting this week, but there was no indication when full talks
would resume, said the person, who asked not to be named because of the private nature of the talks. In the meantime, negotiators at Ford and Chrysler have been poring
over the details in the UAW contract with GM. The UAW and GM reached a tentative four-year contract last week to end a two-day national strike -- the first full-scale
walkout by the UAW against GM since 1970. The union wants to wrap up the ratification voting by October 10. A majority of the UAW members at GM must approve of the
contract for the agreement to be ratified. The more than 73,000 GM hourly workers represented by dozens of UAW locals across the United States have begun voting on the
contract. Members of UAW Local 174 near Detroit voted Monday in favor of the contract after a heavy turnout among the 250 to 300 members in one of the first tests of
the new contract. A local in Lansing, Michigan, was voting on Tuesday, while other major locals had scheduled informational meetings and votes for later in the week
and running into next week. Local 174 members said job security promises, the hiring of temporary workers as permanent employees and a better understanding of the
impact of the health-care trust on retirees may have tipped the scale toward ratification. GM gave the UAW job guarantees, made 3,000 temporary workers permanent and
promised to insource some jobs done by contractors in addition to the health-care trust. The automaker gave binding commitments to its 16 U.S. assembly plants through
the four-year contract, but three do not have binding commitments beyond that and GM expressly excluded two powertrain plants and a service parts operation from a
moratorium on plant closings and sales. A local representing workers at an assembly plant in Orion, Michigan, that has no GM commitments beyond 2013 is scheduled to
vote on Wednesday on the contract. (Additional reporting by  Kevin Krolicki  and  Poornima Gupta )
```

(a) Example of a news article in our data.

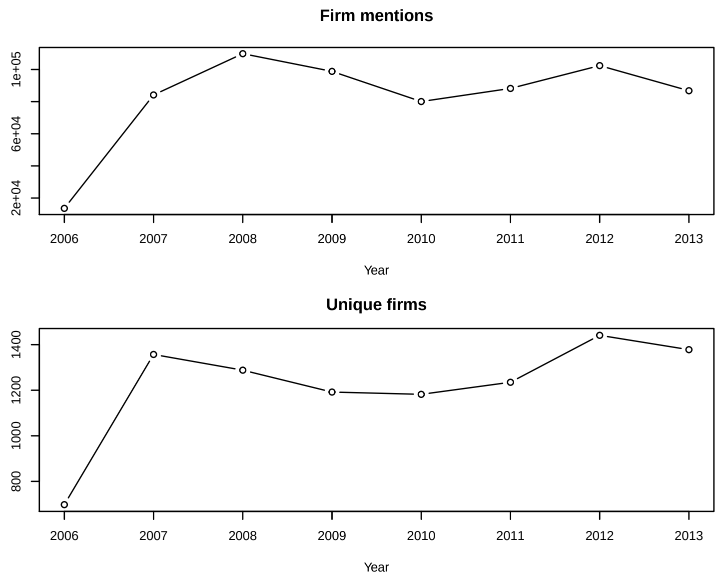


(b) Number of articles per year.

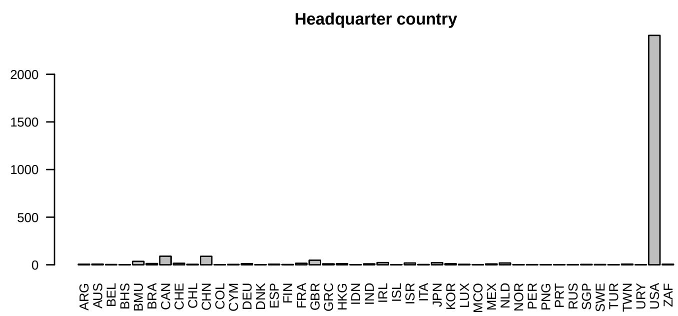
```
# A tibble: 4 x 7
  id      sid   tid tid_end entity_type  entity   entity_normalized
  <chr> <int> <int>   <int> <chr>        <chr>    <chr>
1 doc1      1    10      10 ORGANIZATION GM       ""
2 doc1      1    16      16 ORGANIZATION Ford     ""
3 doc1      1    18      18 ORGANIZATION Chrysler ""
4 doc1      1    41      41 DATE         Tuesday  XXXX-WXX-2
```

(c) Sample output of the coreNLP toolkit.

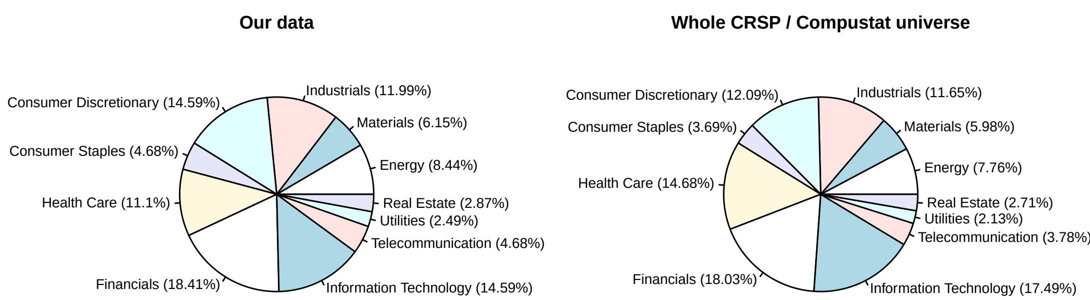
Figure 1: Sample of a news article in our data together with the time series of the number of articles per year and a sample output of the coreNLP software.



(a) Time series of the number of firms recognized per year.



(b) Distribution of headquarter countries.



(c) GICS sector distributions in our sample and in the CRSP / Compustat universe.

Figure 2: Graphical summary of the sample of firms in our data. The sample includes 2,961 distinct firms and covers the time period between October 20, 2006, and November 20, 2013. In Figure (c), we obtain data for the whole universe of firms in the CRSP / Compustat database and display the distribution of sectors in those data.

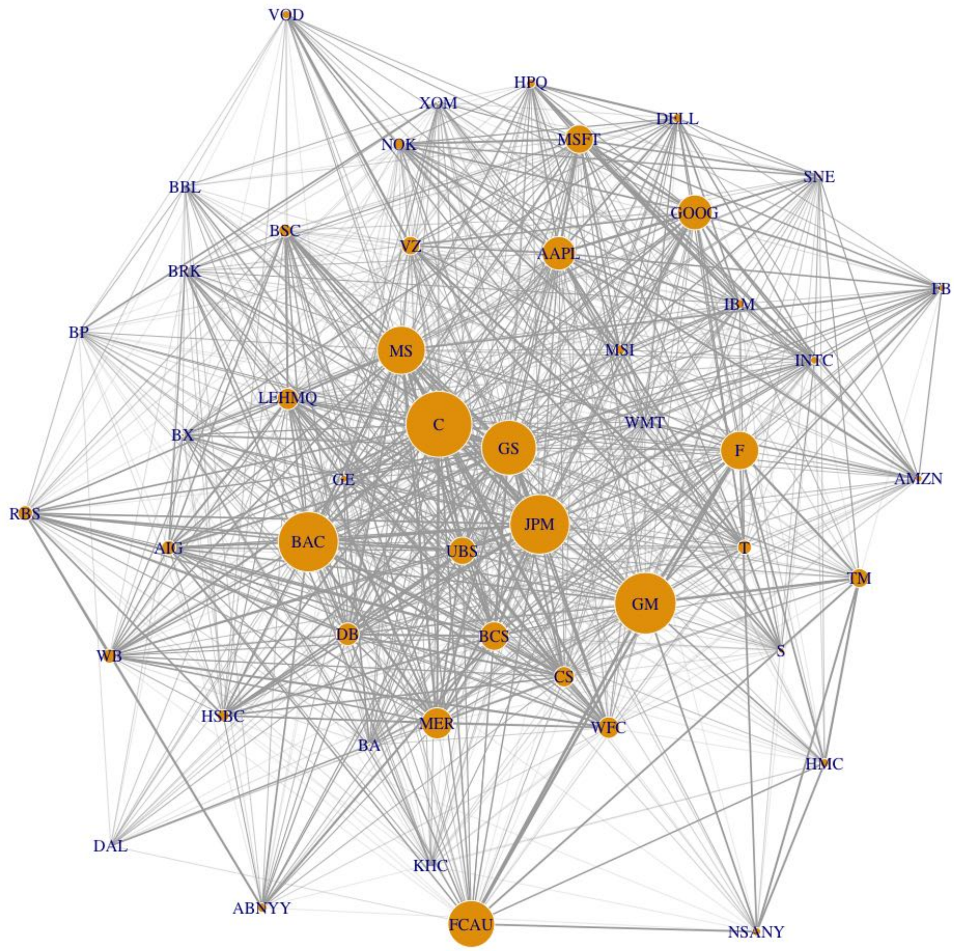
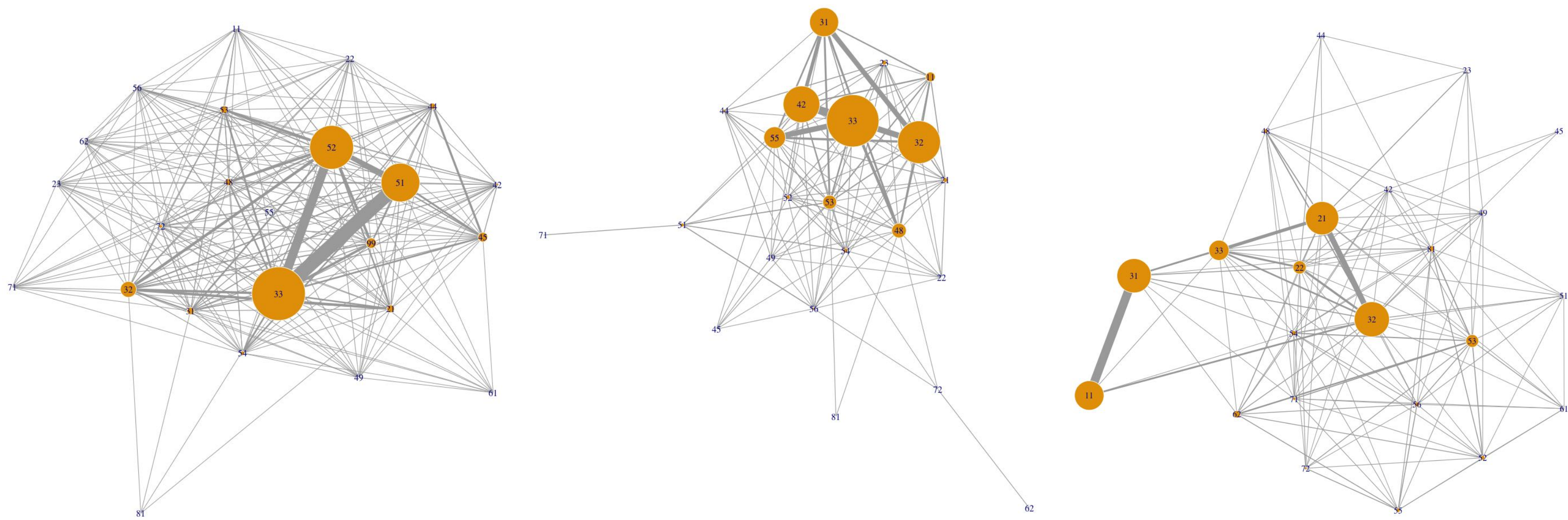
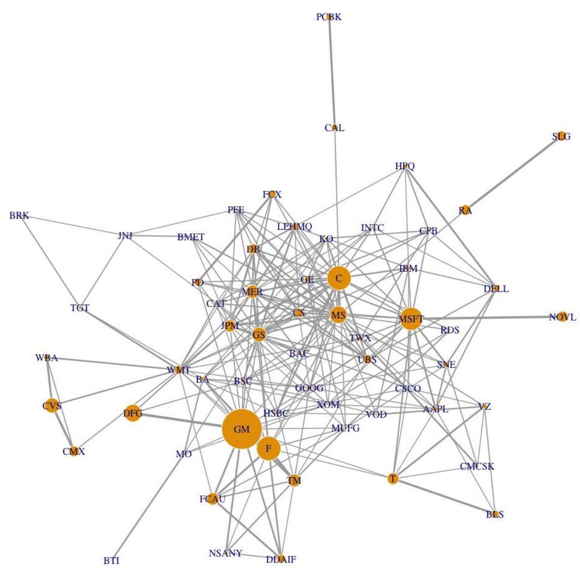


Figure 3: Network of firms implied by the full news data sample covering the years 2006 through 2013. We only plot the largest 50 firm nodes in our network. The size of a node is proportional to the number of times that firm is identified to be connected to another firm. The width of a link between two firms is proportional to the number of times that link is identified in our data.

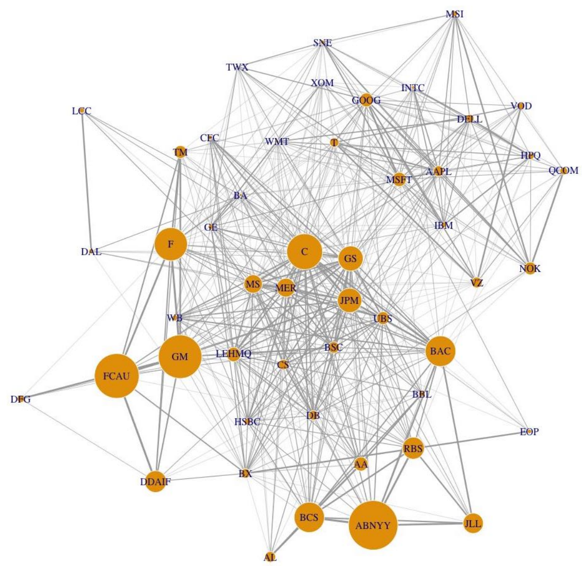


(a) News-implied network. (b) BEA input network. (c) BEA output network.

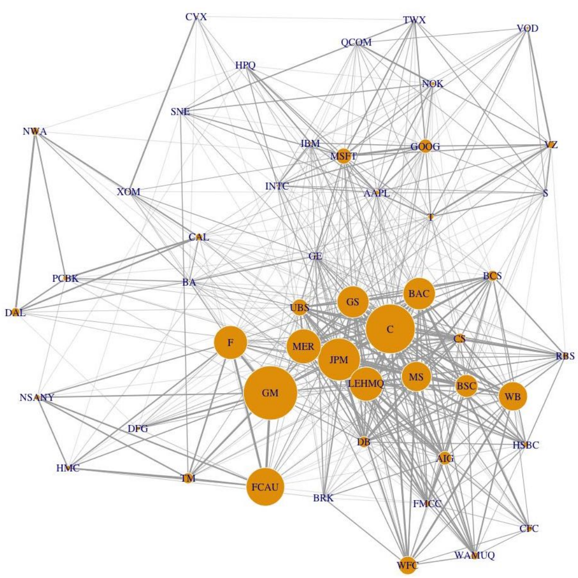
Figure 4: Intersectoral networks implied by our news data and by the 2012 BEA input-output matrix. We obtain data on NAICS codes from the CRSP / Compustat database. We then aggregate firms by the first two-digits of their NAICS sector codes. The size of a node is proportional to the number of times that sector is identified to be connected to another sector. The width of a link between two sectors is proportional to the number of times that intersectoral link is identified in the text data. We build the BEA input (output) network from the upper (lower) triangular matrix of the BEA data. We use the 2012 Industry by Industry/After Redefinitions/Producer Value Total Requirements table. The size of a node in the BEA input network is equal to the net value of the input from other sectors required to produce one dollar of output. The width of a link is proportional to the net input shared across sectors. In the BEA output network, the size of a node is equal to the net value of the output of that sector which is used by other sectors as input. The width of a link is proportional to the net value of the output shared between two sectors.



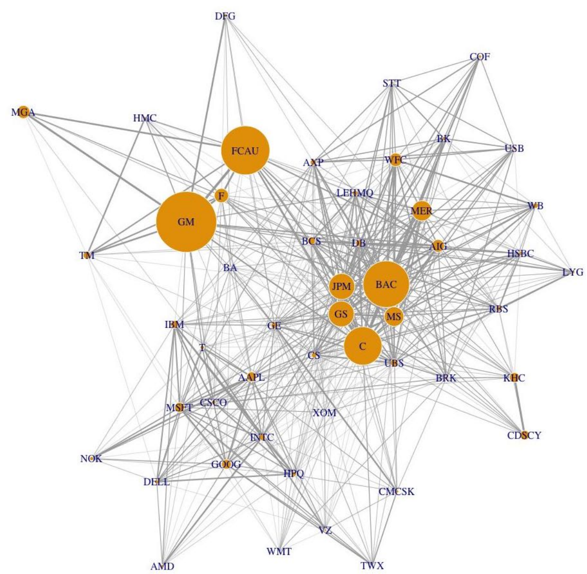
(a) Year 2006.



(b) Year 2007.

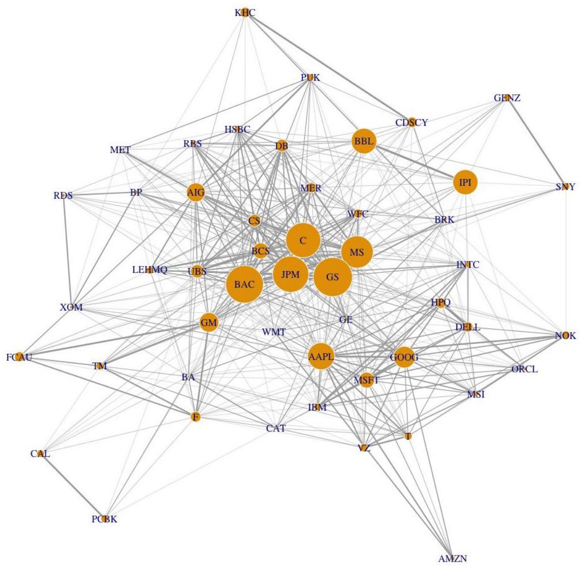


(c) Year 2008.

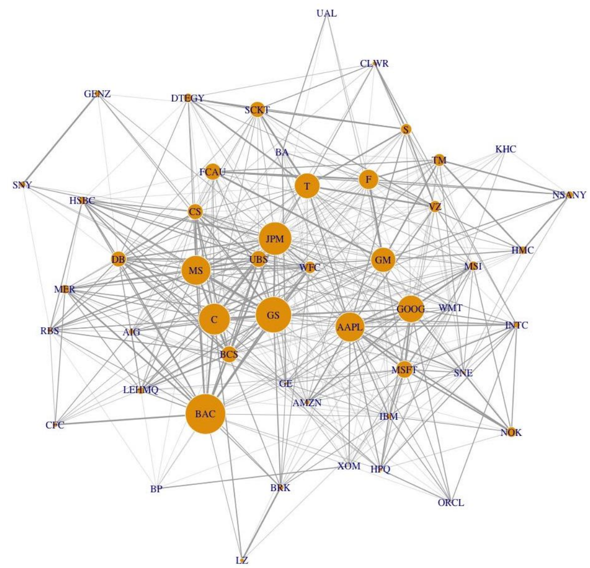


(d) Year 2009.

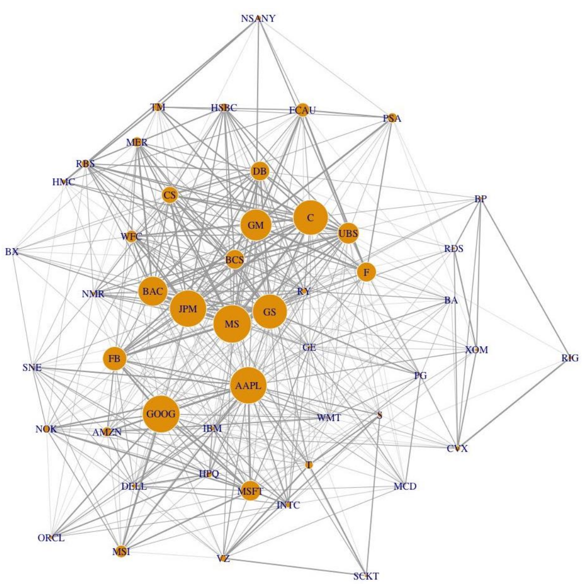
Figure 5: Time series of news-implied networks in our data sample for the years 2006 through 2009. For any given year, we collect the links identified in news articles published in that year and aggregate to a network. The size of a node is proportional to the number of times that firm is identified to be connected to another firmin a year's data. The width of a link between two firms is proportional to the number of times that link is identified in a year's data.



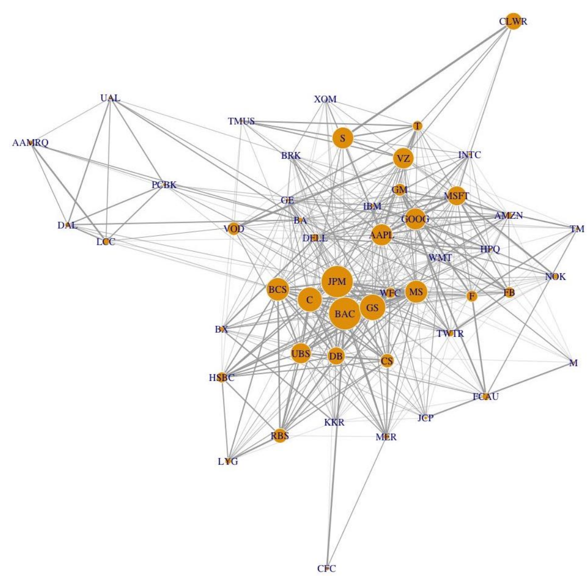
(a) Year 2010.



(b) Year 2011.



(c) Year 2012.



(d) Year 2013.

Figure 6: Time series of news-implied networks in our data sample for the years 2010 through 2013. For any given year, we collect the links identified in news articles published in that year and aggregate to a network. The size of a node is proportional to the number of times that firm is identified to be connected to another firmin a year's data. The width of a link between two firms is proportional to the number of times that link is identified in a year's data.

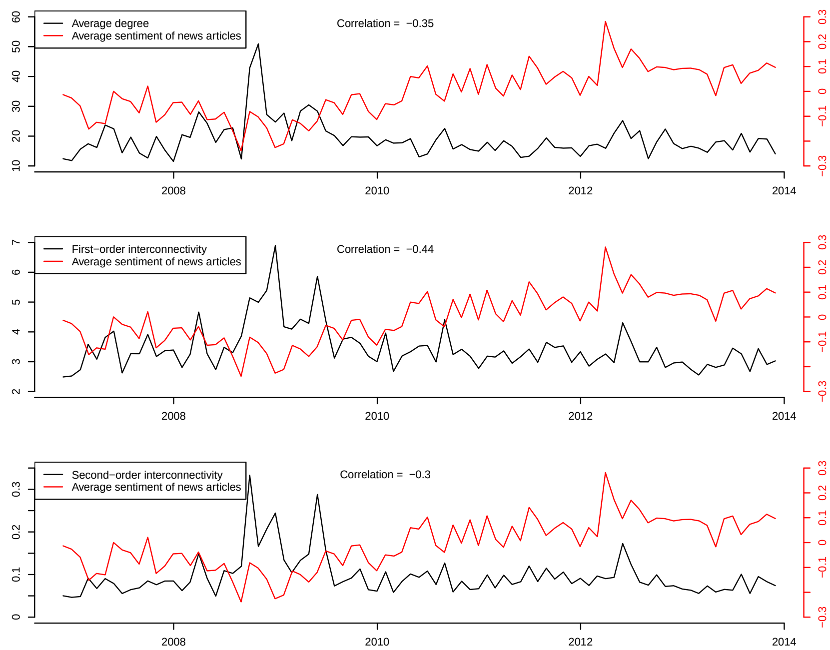
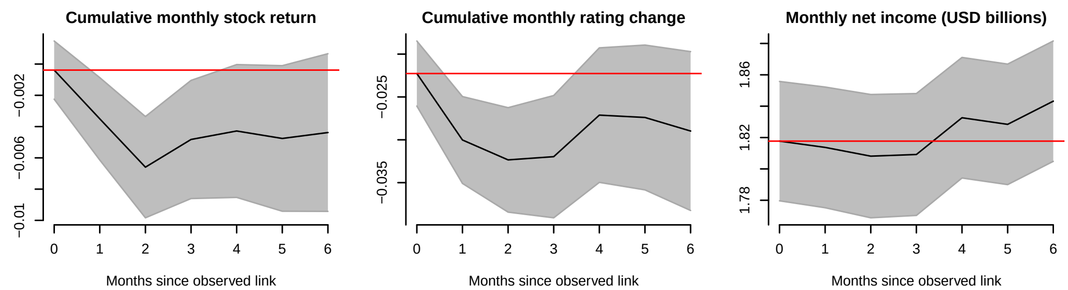
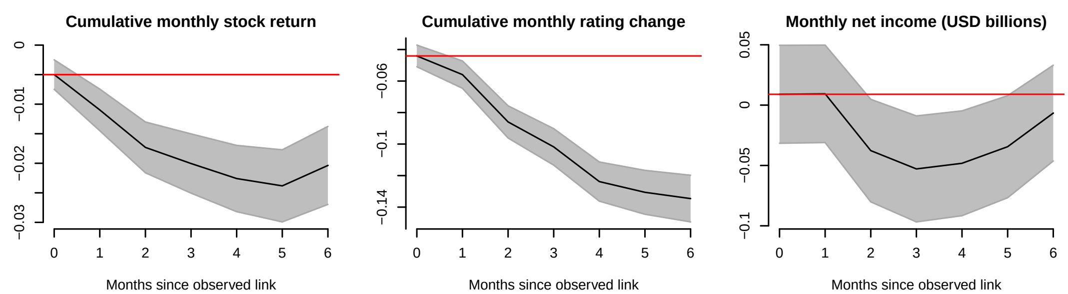


Figure 7: Time series of the average degree, first-order interconnectivity, and second-order interconnectivity for the networks implied by our news data. For any given month, we collect all news article published in that month and extract firm links using the methodology of Section 2. Given a monthly network, we evaluate the average degree as $\bar{d}_t = \frac{1}{N_t}\sum_{n=1}^{N_t} d_t^n$, where N_t is the number of nodes in the network of month t and $d_t^n = \sum_{j=1}^{N} w_t^{j,n}$ is the degree of node n is the number of links of node n on month t ($w_t^{j,n}$ is the number of links that connect nodes j and n on month t). We also evaluate the first-order interconnectivity measure as $\frac{1}{\bar{d}_t}\left(\frac{1}{N-1}\sum_{n=1}^{N_t}(d_t^n - \bar{d}_t)^2\right)^{1/2}$ and the second-order interconnectivity as $\sum_{n=1}^{N}\sum_{j\neq n}\sum_{k\neq j,n} \frac{w_t^{j,n}}{N_t \bar{d}_t}\frac{w_t^{k,n}}{N_t \bar{d}_t}\frac{d_t^j}{N_t \bar{d}_t}\frac{d_t^k}{N_t \bar{d}_t}$. In any given month, we consider only the largest 200 nodes with size measured by the degree of a node.

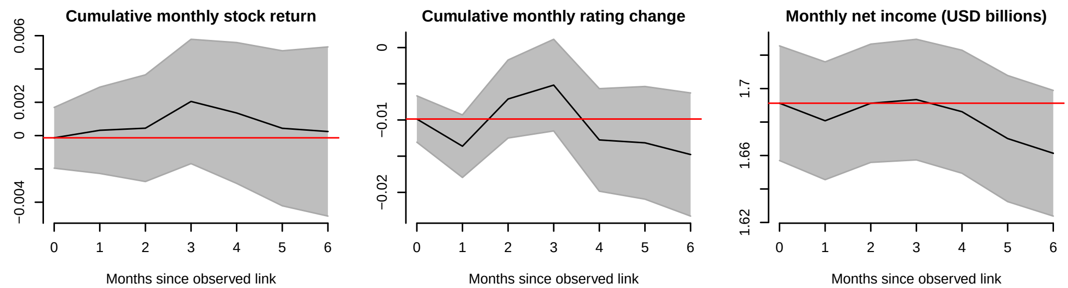


(a) Post-link performance of the larger of the two linked firms.

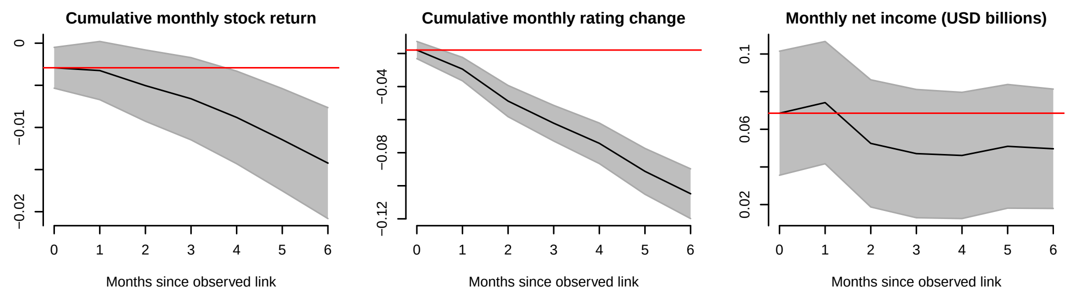


(b) Post-link performance of the smaller of the two linked firms.

Figure 8: Post-link performance of linked firms. Each time we observe a link between two firms, we collect for each of the firms their realized stock returns, credit rating changes, and reported net incomes over the subsequent six months after the link is observed in the data. The above plots show the cross-sectional averages (black lines) for each of the variables in the months following the link month, which we denote month 0. The grey areas show 95% asymptotic confidence intervals for the cross-sectional averages.



(a) Post-link performance of the larger of the two linked firms.



(b) Post-link performance of the smaller of the two linked firms.

Figure 9: Post-link performance of linked firms, where the link is drawn at random if one of the two firms experiences financial distress. Here, we consider a firm to be distressed is one of the following conditions applies: *(i)* its monthly stock return lies in the bottom quintile of the observed distribution of monthly stock returns in our sample, *(ii)* its monthly realized volatility lies in the top quintile of the observed distribution of monthly realized volatilities in our sample, *(iii)* its net income lies in the bottom quintile of the observed distribution of monthly net incomes in our sample, or *(iv)* it experiences a credit downgrade. We draw links among the set of firms that satisfy these properties at random from a Bernoulli distribution, where the probability of success is equal to the unconditional probability of observing a link in our data. Each time we draw a link, we collect for each of the firms their realized stock returns, credit rating changes, and reported net incomes over the subsequent six months after the link is drawn. The above plots show the cross-sectional averages (black lines) for each of the variables in the months following the link month, which we denote month 0. The grey areas show 95% asymptotic confidence intervals for the cross-sectional averages.

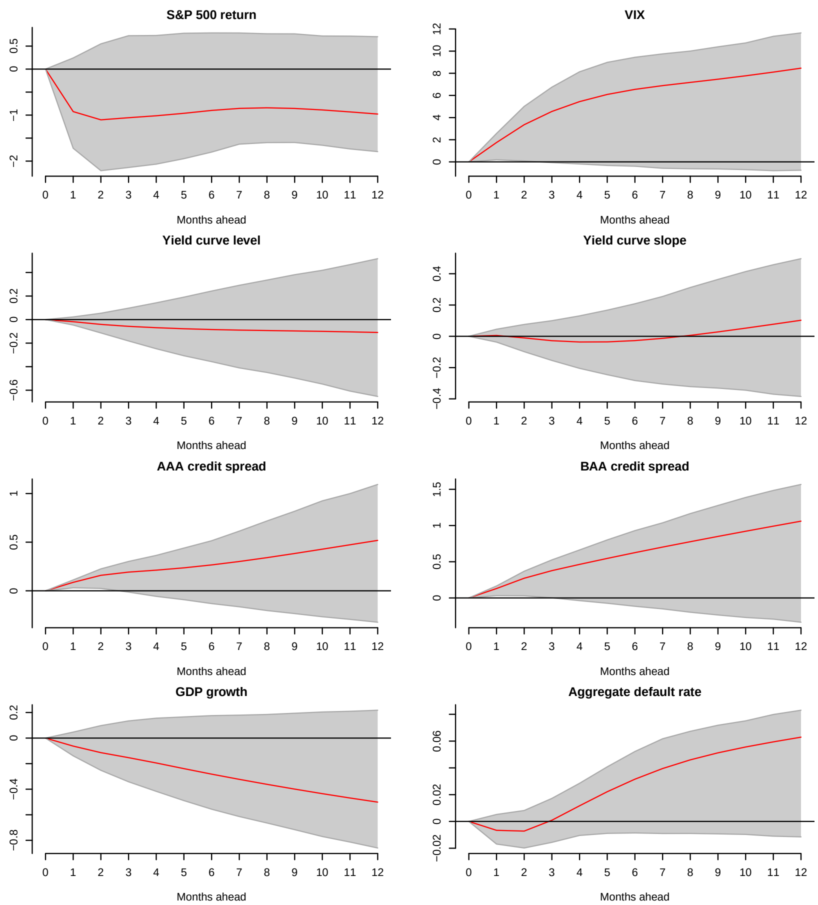


Figure 10: Cumulative impulse response functions to orthogonal shocks to the average degree measure of a news-implied network for a one-lag vector autoregressive model of the joint dynamics of GDP growth, S&P 500 returns, the VIX, the AAA and BAA corporate credit spreads, the level and slope of the yield curve, the aggregate default rate in the U.S., and news-implied connectivity. The red line give the impulse responses and the grey areas give 95% confidence bands computed via bootstrap with 1000 bootstrap samples. We consider one-standard deviation shocks to the orthogonal component of the impulse variable. We base our identification on a Cholesky decomposition of the residual variance-covariance matrix, where the variables are ordered as follows: (1) GDP growth, (2) S&P 500 returns, (3) VIX, (4) Yield curve level, (5) Yield curve slope, (6) AAA credit spread, (7) BAA credit spread, (8) Aggregate default rate, (9) Average degree, (10), The residual of a regression of first-order interconnectivity on average degree, and (11) the residual of a regression of second-order interconnectivity on average degree and first-order interconnectivity. Table 11 provides summary statistics of the factors considered in this analysis.

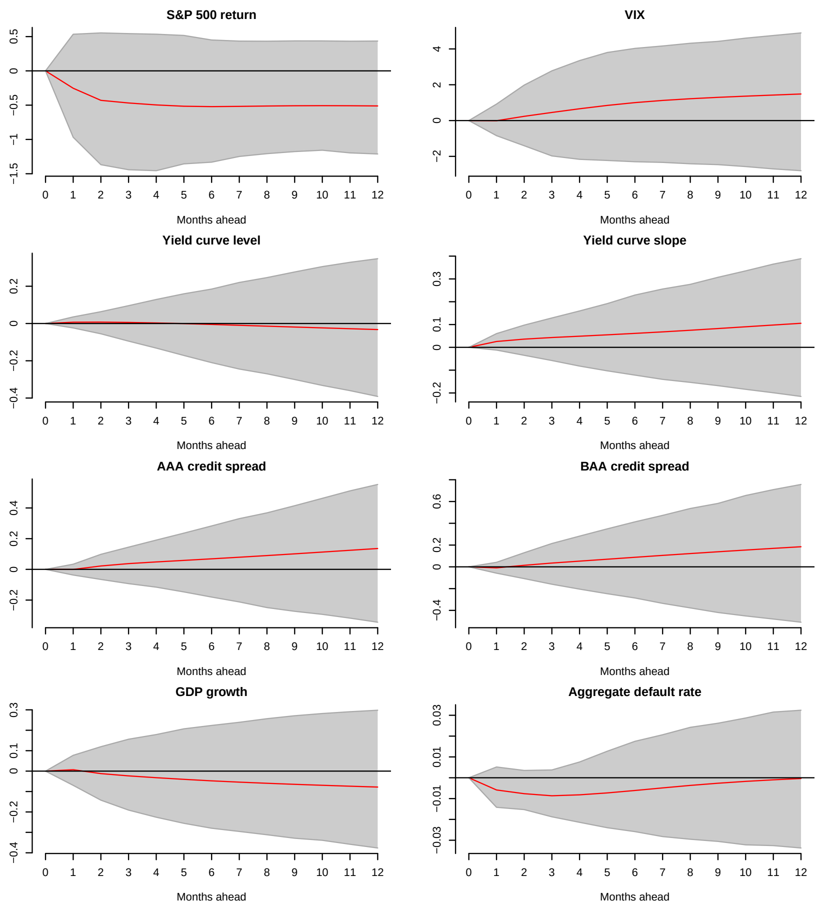


Figure 11: Cumulative impulse response functions to orthogonal shocks to the first-order interconnectivity measure of a news-implied network for a one-lag vector autoregressive model of the joint dynamics of GDP growth, S&P 500 returns, the VIX, the AAA and BAA corporate credit spreads, the level and slope of the yield curve, the aggregate default rate in the U.S., and news-implied connectivity. The red line give the impulse responses and the grey areas give 95% confidence bands computed via bootstrap with 1000 bootstrap samples. We consider one-standard deviation shocks to the orthogonal component of the impulse variable. We base our identification on a Cholesky decomposition of the residual variance-covariance matrix, where the variables are ordered as follows: (1) GDP growth, (2) S&P 500 returns, (3) VIX, (4) Yield curve level, (5) Yield curve slope, (6) AAA credit spread, (7) BAA credit spread, (8) Aggregate default rate, (9) Average degree, (10), The residual of a regression of first-order interconnectivity on average degree, and (11) the residual of a regression of second-order interconnectivity on average degree and first-order interconnectivity. Table 11 provides summary statistics of the factors considered in this analysis.

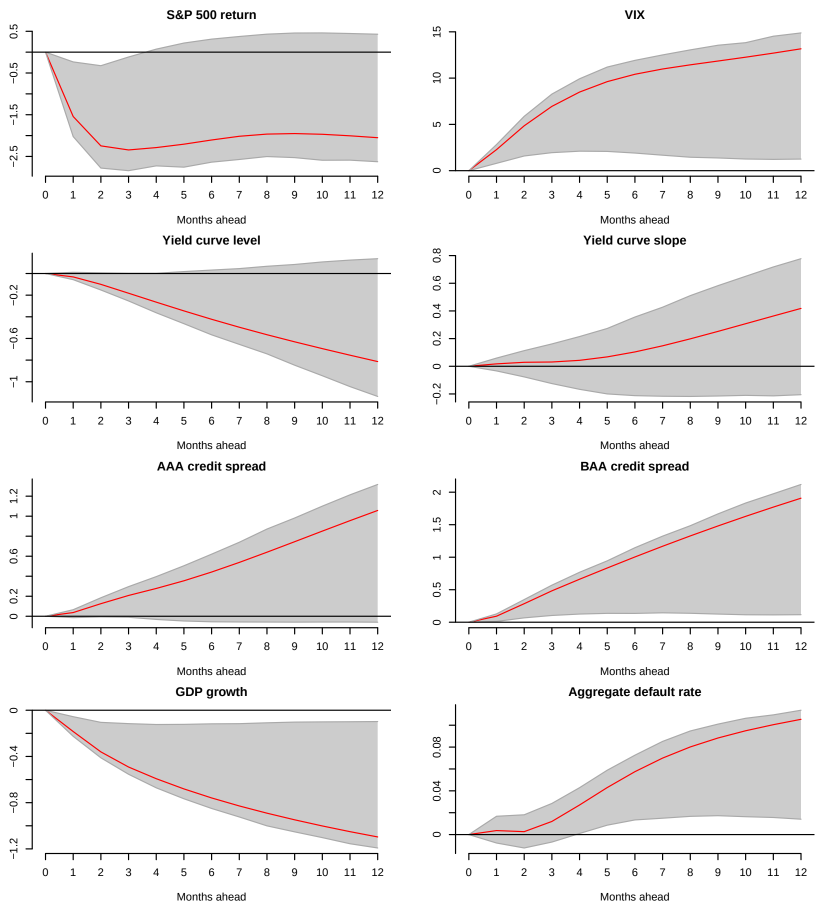
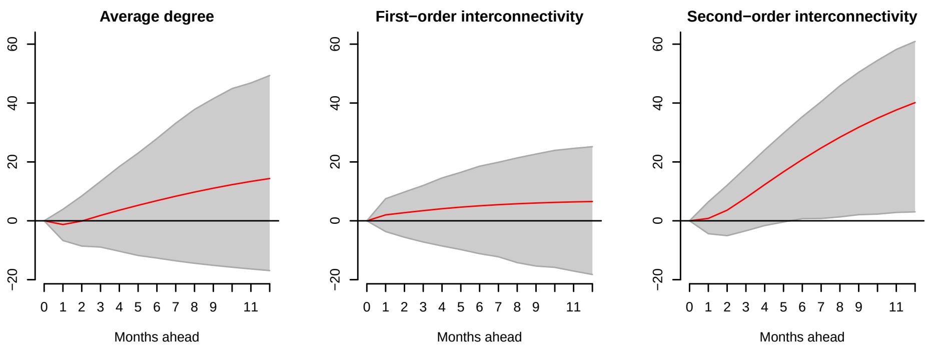
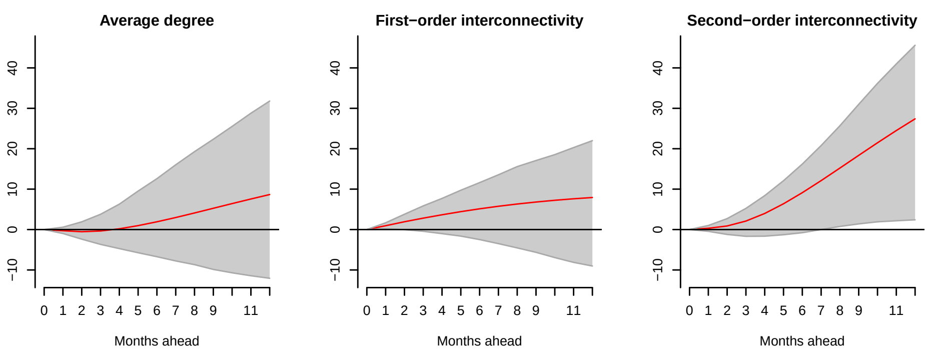


Figure 12: Cumulative impulse response functions to orthogonal shocks to the second-order interconnectivity measure of a news-implied network for a one-lag vector autoregressive model of the joint dynamics of GDP growth, S&P 500 returns, the VIX, the AAA and BAA corporate credit spreads, the level and slope of the yield curve, the aggregate default rate in the U.S., and news-implied connectivity. The red line give the impulse responses and the grey areas give 95% confidence bands computed via bootstrap with 1000 bootstrap samples. We consider one-standard deviation shocks to the orthogonal component of the impulse variable. We base our identification on a Cholesky decomposition of the residual variance-covariance matrix, where the variables are ordered as follows: (1) GDP growth, (2) S&P 500 returns, (3) VIX, (4) Yield curve level, (5) Yield curve slope, (6) AAA credit spread, (7) BAA credit spread, (8) Aggregate default rate, (9) Average degree, (10), The residual of a regression of first-order interconnectivity on average degree, and (11) the residual of a regression of second-order interconnectivity on average degree and first-order interconnectivity. Table 11 provides summary statistics of the factors considered in this analysis.

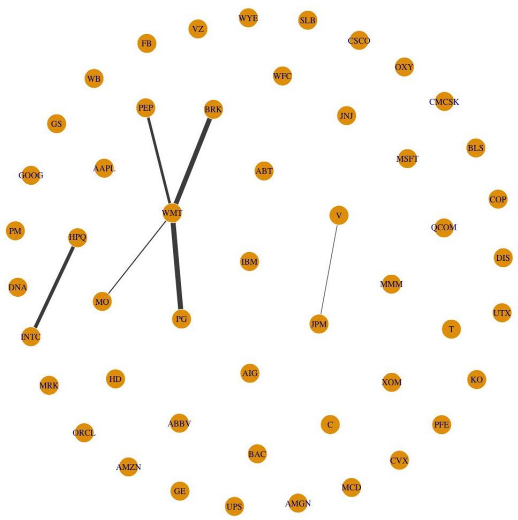


(a) Impulse response functions of the aggregate default rate to orthogonal shocks in the connectivity measures.

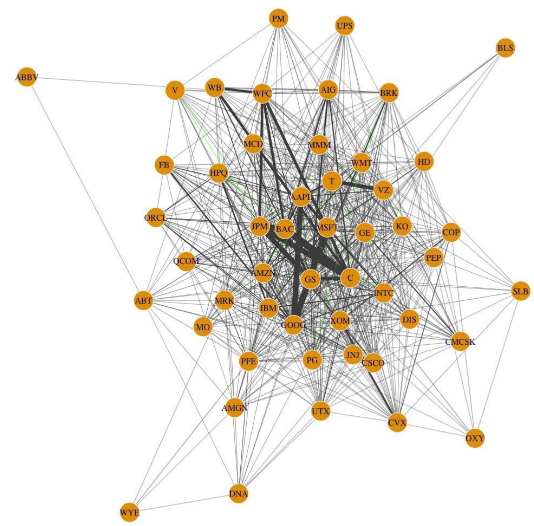


(b) Impulse response functions of the contagion factor to orthogonal shocks in the connectivity measures.

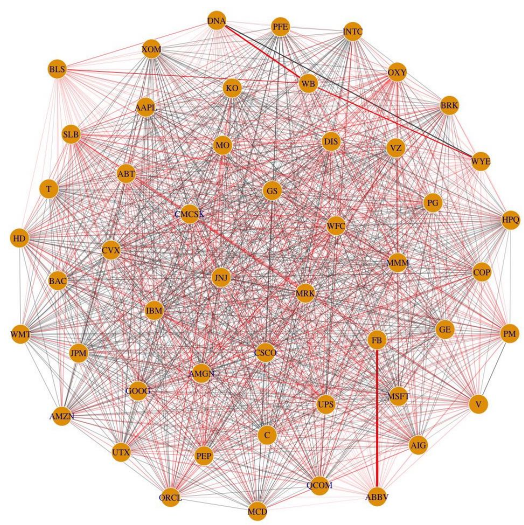
Figure 13: Impulse response functions for the aggregate default rate in the U.S. and the contagion factor of Azizpour et al. (2018) based on one-standard deviation orthogonal shocks to the connectivity measures of the news-implied network. The impulse response functions are implied by a monthly one-lag VAR model for the joint dynamics of GDP growth, the aggregate default rate, the contagion factor, and the three connectivity measures. We estimate the VAR model using default timing data obtained from Moody's Default Risk Service; Table 11 provide summary statistics. The red lines in the impulse response functions give the average responses and the grey areas give 95% confidence bands computed via bootstrap with 1000 bootstrap samples. Our identification is based on a Cholesky decomposition of the residual variance-covariance matrix, where the variables are ordered as follows: (1) GDP growth, (2) Aggregate default rate, (3) Contagion factor of Azizpour et al. (2018), (4) Average degree, (5), The residual of a regression of first-order interconnectivity on average degree, and (6) the residual of a regression of second-order interconnectivity on average degree and first-order interconnectivity.



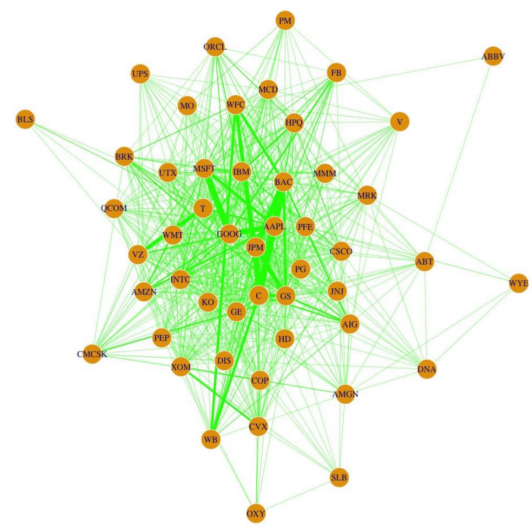
(a) Compustat segments network.



(b) News-implied network.

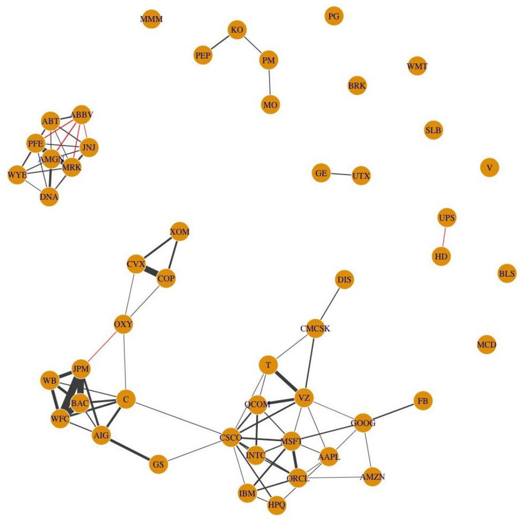


(c) Variance decomposition network (Demirer et al. (2018)).

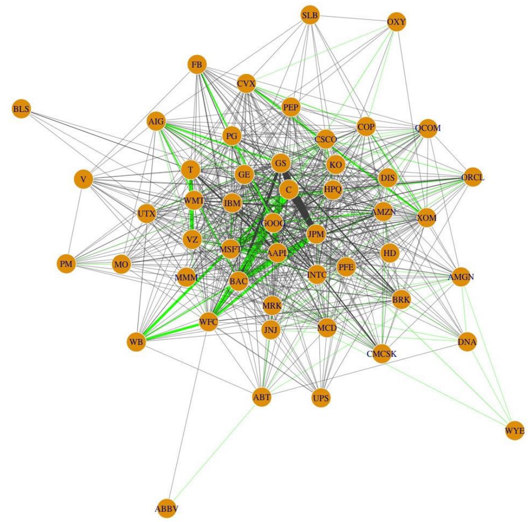


(d) News-implied network.

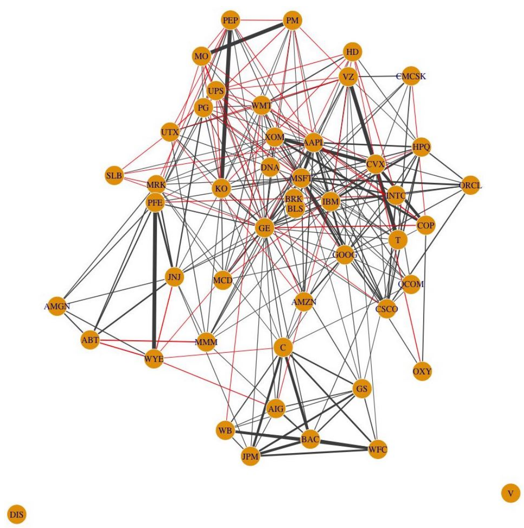
Figure 14: Comparison of alternative networks. Green marks a link that is included in our news-implied network and the corresponding competing network. Red marks a link that is in a competing network but not in our news-implied network. The competing networks are paired row-wise.



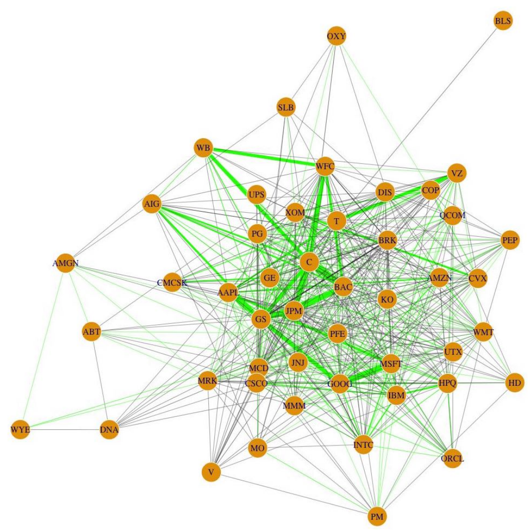
(a) 10-K similarity network (Hoberg and Phillips (2016)).



(b) News-implied network.



(c) EDGAR co-search network (Lee et al. (2015)).



(d) News-implied network.

Figure 15: Comparison of alternative networks. Green marks a link that is included in our news-implied network and the corresponding competing network. Red marks a link that is in a competing network but not in our news-implied network. The competing networks are paired row-wise.