ESSAYS ON THE FINANCIAL IMPLICATIONS OF WEB TRAFFIC

by

Logan P. Emery

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Krannert School of Management

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Huseyin Gulen, Chair

Krannert School of Management, Purdue University

Dr. Mara Faccio

Krannert School of Management, Purdue University

Dr. Zhi Da

Mendoza College of Business, Notre Dame

Dr. Sergey Chernenko

Krannert School of Management, Purdue University

Dr. Adem Atmaz

Krannert School of Management, Purdue University

Approved by:

Dr. Yanjun Li

To my friends, my family, and my wife.

ACKNOWLEDGMENTS

I thank my advisor Huseyin Gulen, as well as my committee members Adem Atmaz, Sergey Chernenko, Zhi Da, and Mara Faccio for their invaluable comments. I also thank Gennaro Bernile, Craig Brown, Doug Emery, Dimas Fazio, Mitch Johnston, Chan Lim, Ping Liu, Fangcheng Ruan, Deniz Yavuz, Haiqing Zhao, and seminar participants at the FMA Doctoral Consortium, University of Dayton, Marquette University, California State University - Fullerton, University of Connecticut, Rotterdam School of Management, and Purdue University for helpful feedback.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In the first chapter, I document that online feedback loops, such as search engines, drive customers and revenue to prominent firms, contributing to rising industry concentration. To identify prominent firms online, I measure centrality in a network of firm websites covering more than 100,000 public and private firms. Industries with firms that are more central become more concentrated and central firms increase their market share during the sample period. This appears to be due to firms' ability to meet earnings expectations. Central firms become more profitable and peripheral firms earn negative risk-adjusted returns and underperform earnings forecasts. Evidence from the COVID-19 shutdown, which drove economic activity to the Internet, supports these conclusions. Central firms received the vast majority of the influx of web traffic and had significantly higher returns during the shutdown.

In the second chapter, I create a novel definition of peer groups (web-based peers) for over 100,000 public *and* private firms by extracting clusters from a network of firm websites. The network is built from unique data on overlapping web traffic. Peer firms are therefore more likely to have similar website users, and by extension, provide similar products or services. Web-based peer groups are related to traditional industry classifications, the preferred choice when defining industries for private firms, but outperform them in standard benchmarking tests. To further demonstrate the benefits of web-based peer groups, I examine IPO waves. IPO activity is closely related to peer-firm IPO activity in the past 12 months, controlling for the overall IPO market and waves within traditional industries. IPO followers earn lower post-IPO returns up to three months after going public, consistent with these firms being lower quality and attempting to benefit from the successful IPOs of their peers.

1. MARKET DOMINANCE IN THE DIGITAL AGE: ONLINE FEEDBACK LOOPS AND RISING INDUSTRY CONCENTRATION

1.1 Introduction

The U.S. economy has become increasingly concentrated in fewer firms since the 1990's (Grullon, Larkin, & Michaely, 2019). At the same time, these firms appear to have increased their market power, creating concerns about the state of competition in the economy (Covarrubias, Gutiérrez, & Philippon, 2019).[1] Evidence points to the possibility that new technologies play an important role in this phenomenon. Industries with more extensive use of information and communications technology appear to exhibit stronger concentration effects, although the empirical link is inconclusive.[2] Autor, Dorn, Katz, Patterson, & Van Reenen (2020) discuss how technological changes have the potential to give rise to "winner-take-most" markets, where only the most efficient firms are able to survive.

The "winner-take-most" aspect of these markets is created by feedback loops inherent in new technology, like the Internet. The digital revolution has increased the value of intangible capital, which tends to have low marginal costs relative to fixed costs and therefore benefits established firms. These firms also benefit from increased network effects, where consumer utility is a function of the number of other consumers also using the platform. Even the infrastructure of the Internet naturally benefits established websites. Web traffic is driven to sites with the most links, leading to the creation of even more links to those sites (Barabási & Albert, 1999). In addition, search engines report results as a function of popularity, making it more difficult for small websites to gain web traffic (Cho & Roy, 2004). This effect is further strengthened by consumer biases, as consumers do not scroll down for more search

[1]↑A number of related macroeconomic effects have also been documented: markups have increased (De Loecker, Eeckhout, & Unger, 2020), labor share of output has declined (Karabarbounis & Neiman, 2014; Barkai, 2020), firm entry rate has declined (Decker, Haltiwanger, Jarmin, & Miranda, 2016), the relationship between investment and profitability and weakened (Covarrubias, Gutiérrez, & Philippon, 2019), and patent concentration has increased (Akcigit & Ates, 2021). One prominent explanation is the decrease in anti-trust enforcement, which has received significant support in the literature (Bessen, 2016; Covarrubias, Gutiérrez, & Philippon, 2019; Grullon, Larkin, & Michaely, 2019). Given the magnitude and breadth of the effects documented, it is likely that several mechanisms are contributing to this phenomenon.

[2]↑See Bessen (2020), Kurz (2017), and Calligaris, Criscuolo, & Marcolin (2018).

results, agree to use default applications and settings, and single-home on platforms. Finally, artificial intelligence algorithms have increased the value of consumer data, which is itself a function of economic activity and therefore easiest for established firms to obtain. While the welfare implication of these online feedback loops is still uncertain (Stigler Committee on Digital Platforms, 2019), there is reason to suspect that the digitization of the economy has significantly contributed to rising industry concentration.

In this paper, I investigate how online feedback loops drive customers, and consequently revenue, to prominent firms, increasing industry concentration. To identify prominent firms online, I measure the centrality of firms in a network of firm websites. The network is constructed using a unique dataset detailing website audience overlap for more than 100,000 public and private firm websites, available semiannually from June 2017 to December 2019. Website audiences have a larger overlap when they are more similar, so a firm's centrality reflects how much its website audience is similar to, or representative of, the aggregate audience of firm websites (which is empirically dominated by consumers). A more representative audience provides stronger network effects, more benefits from search engines, and a stronger position to collect consumer data.

I provide empirical support for the connection between centrality and online feedback loops by documenting that central firms are more likely to appear in search engine results, receive more web traffic, and have higher web traffic growth rates in the future. This demonstrates that online feedback loops push customers to central websites. The network approach also offers important advantages. The inclusion of private firms allows for a holistic view of the economy, which significantly improves the measure of centrality for *public* firms. Furthermore, the data naturally encode market peers (i.e., industries) into the network structure, allowing centrality to account for firms operating in different markets.

Using this measure of centrality, I first document a strong correlation between centrality and concentration in the future. Central industries are more exposed to the effects of online feedback loops, which appear to exacerbate concentration within the industry. During the sample period, I find that central industries become more concentrated in the following year by an additional 7%, controlling for other industry characteristics. Results at the firm level tell a similar story, with central firms increasing their market share by an additional 2.1%

in the following year. Furthermore, I find that central firms become more profitable in the following year, consistent with online feedback loops influencing revenue.

While these results suggest a significant relationship between centrality and concentration, it is also possible that this relationship is endogenous. Central firms tend to be large and more productive, so it is possible that these characteristics determine centrality while also driving outcome variables like growth in market share. Several control variables are included in the regressions to control directly for the correlation between centrality and size or productivity, but nonetheless there could remain some omitted component. I therefore employ three techniques to mitigate these concerns.

First, I decompose centrality to isolate the component that is a function of the centrality of a firm's neighbor. Centrality is a function of the number of connections a firm has as well as the centrality of its neighbors (i.e., firms to which it is connected). Firm characteristics like size play an important role in determining the number of connections a firm has, but they have a much smaller effect on the centrality of its neighbors. I therefore decompose centrality to identify firms that are central not simply because they are large and have many connections, but because they are well positioned in the network. Results remain significant when using this isolated component.[3]

Second, I investigate the relationship between centrality and stock returns. To the extent that investors are already accounting for the fact that central firms are large and productive, these quantities can tell us more about the specific effects of centrality. Forming centrality-sorted portfolios each month from 2014 to 2019,[4] I find that central firms earn significantly higher risk-adjusted returns compared to peripheral firms at approximately 0.70% per month. Several robustness tests confirm the relationship between centrality and stock returns, including double-sorted portfolios and examining firms with a stronger relationship between

[3]↑I also exploit mergers and acquisitions between public firms as a disruption to prior industry structure and a shock to firms' access to online feedback loops in Appendix C.3. In the year following the completed merger or acquisition, I find a relatively larger increase in industry concentration if the target firm was more central, i.e., the acquiring firm gained more access to online feedback loops. This supports the conclusion that online feedback loops can exacerbate industry concentration.

[4]↑Audience overlap data is first available in June 2017, but the beginning of the sample period is extended in order to provide a wider time frame. I use alternative web traffic data, first available in 2014, to investigate potential look-ahead bias in Appendix C.4.1.

sales and web traffic. The return difference, however, is primarily a function of the negative alphas earned by peripheral firms.

I then examine earnings surprises to assess the source of the return differential. Peripheral firms also significantly underperform earnings forecasts and have more negative cumulative abnormal returns (CARs) around their earnings announcements, suggesting that the difference in returns reflects peripheral firms' difficulty in meeting expected cash flows. It seems that market participants, namely investors and analysts, are aware of the advantages of central firms, but are not fully accounting for the significant negative effects on peripheral firms. This is plausible due to the relative difficulty of sorting through peripheral firms, which are smaller and less visible, whereas central firms are relatively easier to identify ex-ante.[5] Furthermore, this demonstrates that the effects of centrality extend beyond the growth of central firms that are already large and productive. Peripheral firms are significantly affected as well – but negatively.

Finally, to observe the effects of online feedback loops more directly, I use the Coronavirus Disease 2019 (COVID-19) shutdown as an exogenous shock to total amount of web traffic. As many brick-and-mortar businesses closed in-person operations in March 2020, a significant amount of economic activity moved online. Indeed, I find a large increase in total web traffic for firm websites. The implication of online feedback loops, however, is that they create a more extreme distribution of consumer activity across firm websites, and I find that this was especially true during the COVID-19 shutdown. The increase in web traffic was almost entirely concentrated in central firms, even after controlling for how different industries were affected by the shock. Central firms also earned significantly higher stock returns during the shutdown, indicating that their position online allowed them to benefit from the influx of web traffic. Overall, the COVID-19 shutdown offers a unique view into what a digitized economy may look like, and the evidence points toward further concentration.

This paper contributes to the literature on the causes of rising concentration in the U.S. economy. From a broader perspective, it provides empirical evidence connecting information

[5]↑Grullon, Larkin, & Michaely (2019) find that firms in industries with a greater increase in concentration earn higher risk-adjusted stock returns, and argue that the recency of rising industry concentration explains why investors continue to underreact.

and communications technology (ICT) to industry concentration. Several papers (Kurz, 2017; Van Reenen, 2018; Grullon, Larkin, & Michaely, 2019; Autor, Dorn, Katz, Patterson, & Van Reenen, 2020) offer conceptual or theoretical arguments for this connection. Empirical evidence, however, has mainly relied on industry-level data, such as the industry-level use of ICT (Bessen, 2020). The use of more granular data has been focused on firm-level production (Autor, Dorn, Katz, Patterson, & Van Reenen, 2020) or financial (Grullon, Larkin, & Michaely, 2019) data, in part due to the limited availability of ICT-related data. Furthermore, consumer behavior has received relatively little attention, also due to a lack of data. I employ a unique dataset that is able to capture the interactions of consumers and firms online, and therefore provides more direct firm-level evidence documenting the relationship between new technology and industry concentration.

This paper also sheds light on the *channels* through which new technology, such as the Internet, impacts concentration. There is a debate in the literature as to whether increased concentration is a function of highly productive "superstar" firms, reduced antitrust enforcement, or growing intangible capital (Autor, Dorn, Katz, Patterson, & Van Reenen, 2020; Covarrubias, Gutiérrez, & Philippon, 2019; Crouzet & Eberly, 2019). Conversations surrounding tech giants such as Google, Amazon, Facebook, and Apple, who benefit greatly from online feedback loops, have involved all three. This paper focuses on online feedback loops, which are born from the characteristics of intangible capital and facilitate the rise of "superstar" firms, and finds evidence that these feedback loops contribute to concentration beyond simply increasing the productivity of central firms.

Finally, this paper connects the literatures on industry concentration and how network structure impacts the economy. Acemoglu, Carvalho, Ozdaglar, & Tahbaz Salehi (2012) model the propagation of idiosyncratic shocks through an intersectoral input-output network. They find that if the network is sufficiently disaggregated (i.e., some sectors are significantly more central than others), then sector-specific shocks can produce significant aggregate fluctuations. Herskovic, Kelly, Lustig, & Van Nieuwerburgh (2020) build on this and investigate how skewness in the distribution of firm sizes, or "granularity" (Gabaix, 2011), interacts with firm networks and affects firm-level volatility. In their model, a more concentrated distribution of firm sizes means that firms are less able to diversify shocks, which results in increased volatility. Herskovic (2018) examines the asset pricing implications of a network economy and finds that the network structure captures firms' exposure to aggregate risk. As the network becomes more concentrated, the lower productivity of large sectors decreases aggregate consumption. While the papers cited here do not discuss it explicitly, the effects they document raise serious concerns as more economic activity is moved online. If digitization continues to concentrate the economy, it may lead to larger aggregate fluctuations, increased firm-level volatility, and lower consumption.

1.2 Web Traffic Data and Centrality

The primary data used for this paper come from Alexa Internet, Inc. (henceforth Alexa).[6] Alexa is an Amazon.com, Inc. subsidiary that provides commercial web traffic data and analytics for the vast majority of active websites. They report a wide range of statistics covering web traffic aggregated at the domain level,[7] including the extent to which two website audiences overlap.

Alexa describes their procedure for collecting the data used to estimate these web traffic statistics as such:

> "Alexa's traffic estimates are based on data from our global traffic panel, which is a sample of millions of Internet users using one of many different browser extensions." *They also gather data from* "a large number of 3rd party providers, representing a diverse panel of web surfers with a broad mix of interests. In addition, we gather much of our traffic data from direct sources in the form of sites that have chosen to install the Alexa script on their site and certify their metrics. As people in our panel visit websites, we count their visits and pageviews for each site and apply data science to estimate what the total traffic and engagement for each site might be."[8]

[6]↑See www.alexa.com.

[7]↑For example, web traffic from example.com, example.com/page1, and example.com/page2 are aggregated together and reported under example.com.

[8]↑See www.alexa.com/about.

Finally, Alexa uses search engine results to measure website popularity and exposure to key words, i.e., the likelihood of a website being one of the first links returned when certain key words are put into a search engine.

The overlap of website audiences is defined by an overlap score, which is estimated from the previous six months of web traffic data and defined at the website-pair level. It is calculated as the Jaccard similarity of the two website audiences, i.e., the number of people who visit website A *and* website B divided by the number of people who visit website A *or* website B:

$$\text{Overlap}(A, B) = \frac{|A \cap B|}{|A \cup B|}. \tag{1.1}$$

Based on a queried website, Alexa reports overlapping websites with an overlap score between 1 and 100, with 100 being two websites that have identical audiences. I obtain overlap score data for over 10 million websites at the end of June and December for each year, starting in 2017.

To identify firm websites, I combine the websites reported by firms in Compustat and Capital IQ as of the month in which the overlap data is collected.[9] This results in 411,584 unique websites for 354,083 public and private firms headquartered in the U.S., of which 231,561 appear in the overlap data.[10]

To provide intuition for the audience overlap of firm websites, Figure B.1 shows four illustrative examples. The figure plots the top overlapping firm websites for Walmart, Alcoa, Microsoft, and Bank of America. Walmart, a large retailer, is closely related to other retailers, such as Target, Best Buy, and Home Depot. Alcoa, an industrial company that produces aluminum, overlaps with other aluminum companies such as Alumina Limited, Clinton Aluminum, and Arconic (which was spun off by Alcoa in 2016). Microsoft, an information technology company, overlaps with technology companies Apple, Amazon, and Adobe. Bank of America, a financial company, overlaps with other financial companies such

[9]↑Some of the websites reported are corporate websites (e.g., aboutmcdonalds.com) while others are conglomerate websites (e.g., Yum! Brands, which owns KFC, Pizza Hut, and Taco Bell, reports their website as yum.com). Starting with websites reported in Compustat, I use Clearbit, Inc.'s Data Enrichment API to identify alternative websites associated with the firm. I then add websites for private firms provided by *Capital IQ*.

[10]↑This corresponds to 6,879 public firms and 224,682 private firms.

as Chase Bank, Wells Fargo, and US Bank. These connections are intuitive, although they should not be taken for granted. Target and Walmart may significantly overlap, but they are classified in separate Global Industry Classification Standard (GICS) industries: Multiline Retail (255030) and Food, Staples, and Retailing (301010), respectively.

The pairwise audience overlap scores can be used to create a network, where the connections between firms represent the similarity of their website audiences. A visual representation of the graph for public firms, corresponding to the June 2017 data, can be seen in Figure B.2. Given the size of the full network used in this paper, it is difficult to visually represent the network structure, so I present a sub-network of only public firm websites with connections to other public firm websites. It contains 3,850 public firm websites and 21,980 overlap scores.[11] The average firm has between 11 and 12 direct connections to other firms, although the median firm only has five direct connections and the most connected firm has 144 connections. A Fruchterman & Reingold (1991) algorithm is used to move websites toward other websites they are more closely connected to. Firms within the same cluster, or group of closely connected websites extracted from the network, are similarly colored. These groups identify firms that are closely related to each other, as represented by their overlapping website audiences.

1.2.1 Centrality

I calculate the centrality of firms within the network using eigenvector centrality. Network centrality is a commonly used characteristic in the network literature, and is often used in the finance literature concerning networks.[12] There are a number of alternative ways to

[11]↑Public firms with connections only to private firms are removed from the graph. The difference between the number of public firms in the full network (5,134) and the number of public firms in the public-firm-only network (3,850) highlights the importance of including private firms in the network.

[12]↑Ozsoylev, Walden, Yavuz, & Bildik (2013) calculate centrality within a network of investors to capture investor informativeness, and find that more central investors earn higher returns. El-Khatib, Fogel, & Jandik (2015) calculate centrality within a network of CEOs and document that more central CEOs participate more frequently in M&A activity but carry greater value losses while being better able to avoid discipline. Bajo, Chemmanur, Simonyan, & Tehranian (2016) calculate centrality within a network of IPO underwriters and show more central underwriters produce better IPO performance. Hollifield, Neklyudov, & Spatt (2017) use transaction data to measure centrality in an interdealer network, and find that central dealers receive lower and less dispersed spreads. Rossi, Blake, Timmermann, Tonks, & Wermers (2018) find that more central investment managers take more risk, receive higher investment flows, and have better portfolio performance. Ahern (2013) explores stock returns and industry centrality as calculated from a network of input/output

calculate network centrality, the most common being degree, closeness, betweeness, and Katz centrality.[13] The hallmark of eigenvector centrality is that a firm's centrality is not only based on the number and strength of connections it has, but also the centrality of the firms it is connected to. This means that a firm can be central in the network even if it has relatively few connections, provided it is well positioned in the network such that those connections are with other central firms.

In essence, a firm's eigenvector centrality in the network of firm websites reflects how well its similarities with other firms explain the overall similarity of firms in the network. Centrality is calculated relative to the entire network, so it is driven by the aggregate flow of Internet users to firm websites. Therefore, a firm's centrality represents how much their website audience is similar to, or representative of, the aggregate audience of firm websites. Internet users may visit a particular firm website for a variety of reasons, but consumers represent the largest portion of visitors to firm websites, and so are the primary drivers of centrality.

In robustness tests, I exploit the fact that eigenvector centrality is a function of both a firm's connections and the centrality of its neighbors to help identify the relationship between centrality and outcome variables such as market share. One issue in identifying this relationship is that some unobservable firm characteristic may be driving both centrality and the outcome variables. A firm's characteristics will have a strong impact on the number and strength of connections it has in the network, but a much weaker impact on the centrality of the firm's neighbors. I therefore isolate this second component of eigenvector centrality by calculating the degree centrality of firms in the network and taking the difference of the natural logarithms of the two measures of centrality. Degree centrality captures the centrality of firms based on their first-order connections, and so removes the component of eigenvector centrality that relates to the number and strength of connections a firm has in the network. Appendix C.2 describes the process of estimating eigenvector centrality (henceforth centrality) and degree centrality in more detail.

linkages. He finds that firms in central industries earn higher stock returns, and argues they arise from central industries being more exposed to economic shocks.

[13]↑Valente, Coronges, Lakon, & Costenbader (2008) show that all of these measures of centrality are usually highly correlated with each other in empirical networks.

Table A.1 presents summary statistics for firm centrality calculated from the full network of public and private firms, averaged across the time periods. Centrality is defined between zero and one, with a mean of 1.09×10^{-3} and a median of 3.31×10^{-9} as shown in the first row of Panel A. The significant difference between the mean and median centrality highlights the skewed nature of the measure. Due to this skewness, I use the natural log of centrality in regressions. Panel A also reports the mean and median centrality of firms by industry, sorted by descending mean centrality, as well as the number of firms and the most central firm in each industry. The data support the notion that consumers are the primary drivers of the network, as the most central firms in each industry are firms that would be familiar to most consumers. The most central industry on average is Communication Services (most central firm Yelp), followed by Consumer Discretionary (most central firm Target). Panel B of Table A.1 reports the 10 most central firms in the network. The fact that consumer driven firms like Yelp and Zillow top this list, along with large retail firms like Target, Walmart, and The Home Depot, once again supports the notion that consumers are the primary drivers of centrality.

It is also interesting to highlight how effectively centrality in the network of *firm* websites distinguishes between firms central to consumers and firms central to Internet users at large. While the average Internet user would likely describe social networking firms like Twitter and Facebook as some of the most central in their network, the network of firm websites places them as the 8th and 17th most central firms in the Communication Services industry, respectively. This is behind more economy-driven firms like Yelp (1), Zillow (2), and TripAdvisor (3). Even LinkedIn, a social networking website focused on connecting professionals, is more central than the biggest players in online friends. This underscores two important aspects of the network. First, restricting the set of websites to firm websites focuses the measure of centrality on the activity of Internet users who visit firm websites, i.e., predominantly consumers, as opposed to the general Internet audience. Second, the inclusion of private firms in the network creates a holistic view of the aggregate firm website audience, as opposed to just the aggregate *public* firm website audience.

1.2.2 Characteristics of Central Firms

To better understand the characteristics of central firms, Panel A of Table A.2 reports a number of financial and web traffic statistics for decile portfolios of firms sorted on centrality.[14] In general, central firms tend to be larger, advertise more, and have more institutional ownership. They also tend to be more profitable on average. To the extent that central firms are the primary beneficiaries of increased concentration, this is consistent with Grullon, Larkin, & Michaely (2019) who find that firms in more concentrated industries tend to be more profitable. The average industry HHI is largest for central firms, providing empirical motivation for the possibility that central industries are becoming more concentrated.

Central firms are also more likely to appear in search engine results (Search Power) and receive more web traffic on average. These statistics offer support for the connection between centrality and online feedback loops, such as search engines, that provide greater benefits to firms with more web traffic. To test this relationship more rigorously, I run regressions of Search Power and Web Traffic on centrality using the following regression specification:

$$\begin{aligned} \Delta Y_{\mathrm{i},t+\tau} &= \beta \ln(\text{Centrality}_{\mathrm{i},t}) + \lambda \boldsymbol{X}_{\mathrm{i},t} + \mu_{k,t} + \eta_{\mathrm{i}}, \\ \Delta Y_{\mathrm{i},t+\tau} &\equiv Y_{\mathrm{i},t+\tau} - Y_{\mathrm{i},t}. \end{aligned} \tag{1.2}$$

where $Y_{\mathrm{i},t}$ is either the Search Power or Web Traffic for firm i in month t. Search Power is an estimate of how competitive a website is for key words on search engines, provided by Alexa for one cross-section in January 2020. Being more competitive for key words means that a web site is more likely to appear higher up on the list of search results when Internet users search for phrases, or key words.[15] Web Traffic is the estimated percent of all Internet users who visited the website, provided daily by Alexa from 2014 to 2019.

$\boldsymbol{X}_{\mathrm{i},t}$ is a vector of control variables including traffic in month t and a measure of website engagement provided by Alexa reflecting how much time visitors spent on the website (Time on Site), as well as market capitalization, sales, advertising expenditure, a dummy variable

[14]↑A description of each variable and how it is calculated is given in Appendix C.1.

[15]↑For example, if a user searches for the phrase "groceries" in Google, the first few links may be from companies like Walmart or Target. These companies have a high Search Power, while other, more local, grocery stores may be pushed to the second or third page of the search results.

for firms that do not report advertising expenditure, absolute value returns, and trading volume. Each variable and the respective data sources are discussed in Appendix C.1. $\mu_{k,t}$ are industry-time fixed effects and η_i are firm fixed effects. Standard errors are clustered by firm and variables are standardized to facilitate interpretation.

The results are reported in Panel B of Table A.2. Column (1) reports a regression of Search Power in January 2020 on centrality and other control variables as of December 2019. While the regression only includes one cross-section, and Search Power is available for only half as many firms as centrality, there is a strong positive relationship between centrality and Search Power. This suggests that central firms are more likely to appear in search engine results, even when controlling for the level of web traffic, website engagement, and firm characteristics.

Columns (2) through (4) report monthly regressions of Traffic on centrality from July 2017 to December 2019. Traffic is measured as a monthly average, while the most recent observation of centrality, measured semiannually, is applied to each month. Column (2) examines web traffic in month $t+1$, and indicates a strong positive relationship between centrality and the level of web traffic next month. Columns (3) and (4) examine the change in web traffic over the next six and twelve months, respectively, which again demonstrate a strong positive relationship with centrality. These results indicate that central firms not only receive higher levels of web traffic, but also grow their web traffic at higher rates. A one standard deviation change in centrality corresponds to approximately a 6% higher growth rate in web traffic over the following year. Overall, this provides strong empirical support for the idea that central firms benefit more from online feedback loops: they get more benefits from search engines, they receive more web traffic, and they have higher web traffic growth rates.

1.3 Centrality and Industry Concentration

I use centrality in the network of firm websites to investigate the relationship between industry concentration and the benefits firms receive from online feedback loops. Central firms appear to receive benefits in the form of higher growth in web traffic, prioritization on search engines, and a position to benefit from network effects and collect more consumer data. If these benefits translate to increased economic activity, then industry concentration will increase.

To investigate this hypothesis, I first examine industry concentration during the sample period. Grullon, Larkin, & Michaely (2019) find that the majority of industries in the U.S. have become more concentrated since the late 1990's, a period that coincides directly with the rise of the Internet. Following them, I measure industry concentration using the Herfindahl-Hirschman Index (HHI) for each industry:

$$HHI_{k,t} = \sum_{\mathrm{i}=1}^{N} \left(\frac{Sales_{\mathrm{i},t}}{\sum_{\mathrm{j}=1}^{N} Sales_{\mathrm{j},t}} \right)^2, \tag{1.3}$$

where N is the number of firms in industry k and $Sales_{\mathrm{i},t}$ is the sales for firm i in quarter t as reported in the Compustat Segments database. Firms with a larger HHI are considered more concentrated, with the squared market share in Equation 1.3 placing an emphasis on industries with a greater skewness in sales.

I then run quarterly regressions from July 2017 through 2019 using the following regression specification:

$$\Delta HHI_{k,t+4} = \beta \ \mathrm{avg}_k \left(\ln(\mathrm{Centrality}_{\mathrm{i},t}) \right) + \lambda \boldsymbol{X}_{k,t} + \gamma_t. \tag{1.4}$$

The main variable of interest, $\mathrm{avg}_k \left(\ln(\mathrm{Centrality}_{\mathrm{i},t}) \right)$, is the average log centrality of firms in industry k in quarter t. $\boldsymbol{X}_{k,t}$ is a vector of control variables that includes traffic, market capitalization, book-to-market ratio, sales, advertising expenditure, intangible share, industry HHI in quarter t, the number of firms in the industry, the number of mergers and acquisitions by firms in the industry, industry returns, turnover, and trading volume. γ_t are time fixed effects, standard errors are clustered by industry, and all variables are standardized to facilitate interpretation.

Traffic measures the total web traffic received by firms in the industry, which I control for to demonstrate that centrality does not simply capture the level of web traffic. Intangible Share measures the industry average of the portion of a firm's capital that is intangible, which Crouzet & Eberly (2019) find has coincided with changes in market share over the past two decades. N M&A is the number of mergers and acquisitions that took place in the acquirer's industry over the past four quarters, controlling for the mechanical increase in HHI that results from M&A activity. The overall number of public firms has been declining simultaneously with the increase in industry concentration (Doidge, Karolyi, & Stulz, 2017), which can also result in increased HHI, so I control for the natural log of the number of firms in the industry. Finally, I include industry HHI as of quarter t to control for the possibility that concentration itself allows industries to become even more concentrated in the future.[16] The remaining control variables are discussed in Appendix C.1.

The results are reported in Panel A Table A.3. Column (1) examines HHI in the following quarter. The coefficient on centrality indicates that more central industries are more concentrated on average, with the coefficient corresponding to a 10% more concentrated industry. Columns (2) through (4) then examine the change in industry HHI over the following two, four, and eight quarters, respectively. These columns report a significant correlation between industry centrality and change in HHI, demonstrating that central industries became more concentrated at a higher rate over the sample period. Economically, the coefficient reflects an increase in HHI over the following year that is 7% larger for more central industries. Central industries are more exposed to the online feedback loops that dominate digital markets, and the fact that industry concentration is increasing at a faster rate for these industries points toward online feedback loops playing a significant role.

As discussed in Section 1.2.1, one concern with identifying a relationship between centrality and concentration is that an unobservable firm characteristic may be driving both variables. To alleviate some of this concern, I exploit the fact that eigenvector centrality is a function of both the connections a firm has and the centrality of its neighbors. A firm's characteristics are likely to have a significant impact on its connections, but a much weaker impact on the centrality of its neighbors. To isolate this second component of eigenvector centrality, I measure degree centrality in the network to capture the first component of

[16]↑Given that HHI is measured at the industry level, the lagged HHI is available for all observations, i.e., no observations are omitted due to the requirement for a lagged HHI. Results are similar when excluding lagged HHI as a control variable, suggesting that any survivorship bias created at the firm level is marginal.

eigenvector centrality. I then subtract the natural logarithm of degree centrality from the natural logarithm of eigenvector centrality and calculate the average difference for firms in each industry.

Panel B of Table A.3 examines the relationship between industry concentration and the second component of eigenvector centrality using the regression specification described in Equation 1.4. The panel follows the same structure as Panel A, and reports similar results. Industries with firms that are more central on average become more concentrated in the future. The key difference is that $(Eig - Deg)\ Centrality$ is only a function of the centrality of a firm's neighbors, and so is less likely to be driven by firm-specific characteristics.

A more specific concern with identifying a relationship between centrality and concentration is that both variables are outcomes of a long-run equilibrium reflecting industry structure. To investigate this concern, I examine mergers and acquisitions as shocks to firms' access to online feedback loops. Acquisitions represent a shock to industry structure, as well as a shock to firms' market position online, breaking this potential equilibrium. The identifying assumption is then that acquisitions of more central firms reflect a larger shock to firms' access to online feedback loops. I find that acquirers' industries become more concentrated over the year following a more central target firm being acquired. The results are reported in Table C.2, and discussed in more detail in Appendix C.3.

1.3.1 Market Share and Profitability

I next investigate the relationship between online feedback loops and industry concentration at the firm level. As online feedback loops provide benefits to the central firms, central firms may be able to capture a larger portion of the market and increase their profitability. Grullon, Larkin, & Michaely (2019) find that firms in more concentrated industries are more profitable on average. To investigate these possibilities, I examine the growth of market share and profitability for central firms.

I calculate market share as the ratio of sales to total industry sales, where industries are defined at the 2-digit SIC level. For profitability, I follow Grullon, Larkin, & Michaely (2019) and measure return-on-assets (ROA), which is calculated as net income divided by total assets. Each of these variables is taken from the Compustat Segments database. I then run quarterly regressions from July 2017 through 2019 using the following regression specification:

$$\Delta Y_{\mathrm{i},t+4} = \beta \ln(\text{Centrality}_{\mathrm{i},t}) + \lambda \boldsymbol{X}_{\mathrm{i},t} + \mu_{k,t}. \tag{1.5}$$

$Y_{\mathrm{i},t}$ is the market share or return-on-assets of firm i in quarter t and $\boldsymbol{X}_{\mathrm{i},t}$ is a vector of control variables including web traffic, market capitalization, book-to-market ratio, sales, intangible share, advertising expenditure, a dummy variable for firms that do not report advertising expense, returns, turnover, and trading volume. $\mu_{k,t}$ are industry-time fixed effects. Standard errors are clustered by industry when examining market share and by firm when examining profitability. All variables are standardized to facilitate interpretation.

The results are reported in Table A.4. Columns (1) and (2) examine market share, while columns (3) and (4) examine profitability. Column (1) uses eigenvector centrality as the independent variable of interest and reports that firms that are one standard deviation more central than average increase their market share by an additional 2.1% over the following year. Once again, this could be a result of unobserved firm characteristics driving both centrality and market share. Column (2) therefore uses the component of eigenvector centrality that is a function of neighbor centrality as the independent variable of interest. The relationship between centrality and market share remains significant. This demonstrates that central firms are able to capture a larger portion of their market, consistent with online feedback loops concentrating economic activity.

The growing market share may also allow central firms to become more profitable. In column (3), the coefficient on centrality implies that a one standard deviation increase in centrality corresponds to an additional 9 basis point increase in ROA over the following year. This means that central firms are increasing their profitability at a higher rate than peripheral firms. The relationship remains significant in column (4), which measures centrality as only the component that is a function of neighbor centrality. The result therefore does not appear to be entirely driven by unobserved firm characteristics.

Overall, the evidence in this section demonstrates a strong correlation between online feedback loops and rising industry concentration. Central industries become more concentrated and central firms gain additional market share over the sample period. Moreover, central firms become more profitable over the following year, pointing toward a possible relationship between online feedback loops and revenue. As the unique characteristics of digital markets drive customers to the central firms, this naturally leads to a greater concentration in economic activity.

1.4 Market Outcomes

To further mitigate concerns that the previous results are driven by omitted firm characteristics, I next investigate whether centrality provides information about future market outcomes, namely stock returns and earnings surprises. If investors and analysts are already accounting for the fact that central firms are large and productive, then these quantities can tell us more about the specific effects of centrality.

1.4.1 Stock Returns

I first examine monthly stock returns to central and peripheral firms from January 2014 to December 2019. Stock return data is taken from the CRSP. To ensure a representative and tradable set of securities, I consider only common shares traded on the NYSE, Amex, or NASDAQ. To reduce concerns of liquidity bias, I remove observations where the stock price is below $1.00 at the beginning of the month. These restrictions leave 3,841 firms in the analyzed sample.

Although the first network is based on data from June of 2017, I apply centrality measured in this network to all monthly observations from January 2014 to December 2017. This is done to increase the number of monthly observations from 30 to 72, although results are also reported for the 30-month period. Extending the sample period also implicitly assumes that centrality is sufficiently stable over this time period such that look-ahead bias does not significantly affect the results. While the summary statistics suggest that this assumption is reasonable, this issue is more thoroughly investigated in Appendix C.4.1.

Centrality portfolios are formed by sorting firms into deciles at the beginning of each month. Average monthly returns for value-weighted centrality decile portfolios from 2014 to 2019 are given in the first column in Panel A of Table A.5. Each row represents the average monthly return for that portfolio, with the last two rows being the spread portfolio (equivalent to buying portfolio 10, the most central firms, and selling portfolio 1, the least central firms) and the t-statistic for the spread portfolio. Each column represents excess returns (alphas) for the listed asset-pricing model. In all columns, central firms earn significantly higher stock returns than peripheral firms, with the decile-spread portfolio earning approximately 0.70% per month. Panel B of Table A.5 reports a similar analysis of returns from July 2017 to December 2019, the 30-month period for which there is no look-ahead bias in centrality. Spread portfolio returns are similar in economic magnitude to the 2014-2019 period, although with smaller t-statistics due to the reduced sample period.

To control for various potential risk factors and other cross-sectional determinants, I regress the centrality-portfolio returns on the CAPM, Fama-French 3-factor, and Fama-French 5-factor models.[17] I also include a model adding a momentum factor to the Fama-French 3-factor model (Carhart, 1997).[18] The alphas resulting from these regressions represent the return that cannot be explained by the factors in each respective model. Alphas are generally increasing across centrality portfolios, and the t-statistics indicate that the spread-portfolio alphas are significantly larger than zero. This means that the spread in returns between central and peripheral firms cannot be explained by commonly used asset-pricing models.

A series of robustness tests, discussed in Appendix C.4.2, confirm the relationship between centrality and stock returns. I use web traffic statistics to estimate centrality going back to 2014, and find that the estimates are remarkably stable over the time period. The spread portfolio returns are significant when sorting firms based on estimated centrality, and they are also significant when excluding firms with a low correlation between website traffic and sales. Double-sorted portfolios, sorting first on either sales or market capitalization, demonstrate a significant difference between central and peripheral firms, and centrality is a significant

[17]↑The CAPM regressions control for covariance with excess market returns, the Fama-French 3-factor model controls for excess market returns, size, and book-to-market, and the Fama-French 5-factor model controls for excess market returns, size, book-to-market, profitability, and investment. See Fama & French (1993) and Fama & French (2015).

[18]↑Data for the factors were obtained from Ken French's website.

determinant of returns in Fama & MacBeth (1973) regressions controlling for other common determinants.

Across all tests, the difference in returns appears to be primarily driven by the peripheral firms, which earn sizable negative alphas over the sample period compared to the more moderately positive alphas earned by the central firms. This points to the possible negative effects that online feedback loops can have on peripheral firms. It is also generally inconsistent with omitted risk factors explaining the results, because riskier firms should earn positive alphas due to compensating investors for the risk.[19] Instead, it would appear as if peripheral firms are underperforming investors' expectations. It seems likely that it is difficult for investors to determine just how peripheral the peripheral firms are. Once the more clearly central firms are identified, sorting among the remaining firms is less obvious.

1.4.2 Earnings

The stock return evidence suggests that centrality provides information about either future revenue or investors' discount rates. Given the relationship between centrality and profitability, it seems more plausible that the stock return results are a function of revenue. If this is the case, then the difference between central and peripheral firms, and the relatively larger surprise for peripheral firms in particular, should be reflected in firms' earnings.

To investigate this, I compare the reported earnings of portfolios of firms sorted by centrality to the earnings forecasts of those portfolios. Data on earnings announcements and analyst earnings forecasts are taken from the Institutional Brokers Estimate System (IBES) database. For each firm, I multiply earnings-per-share by the number of shares outstanding in that month, found in CRSP, to obtain the total earnings. At the end of each March, June, September, and December, I sort firms into deciles based on centrality. I then

[19]↑Further discussion of potential risk explanations can be found in Appendix C.4.2.

aggregate earnings and earnings forecasts by deciles for all firm-fiscal quarters that ended in the previous three months.[20] The portfolio-level earnings surprise is then calculated as

$$Surp_{p,t} = \frac{Earnings_{p,t} - Forecast_{p,t}}{Forecast_{p,t}}, \tag{1.6}$$

where $Surp_{p,t}$ is the earnings surprise for portfolio p in quarter t, $Earnings_{p,t}$ is the cumulative reported earnings for the portfolio, and $Forecast_{p,t}$ is the cumulative mean analyst forecast for the portfolio.

I also employ a more direct measure of the reaction of investors to earnings announcements. I measure the three-day cumulative abnormal return (CAR), controlling for the 3-factor Fama-French model, around the earnings announcement, following La Porta, Lakonishok, Shleifer, & Vishny (1997). This value reflects the stock price reaction to firms announcing their earnings, controlling for common risk factors. To create a portfolio-level measure, I value-weight the CARs within portfolios.

The average earnings surprise and earnings announcement CAR for centrality-sorted portfolios are reported in Panel A of Table A.6. Columns (1) and (3) cover the full sample period, 2014-2019, and columns (2) and (4) cover only the period after centrality is first observed, July 2017 through 2019. First, it seems that the earnings of central firms do not differ significantly from analyst forecasts, consistent with the notion that the revenue for these firms is easier for market participants to account for. The peripheral firms, on the other hand, significantly underperform analyst expectations in both sample periods. Moreover, peripheral firms have significantly negative announcement CARs in both periods, indicating that under-performance in earnings results in significant market reactions. In general, this suggests that a firm's position in the network provides information about revenue.

[20]↑Aggregating total earnings reduces the possibility that earnings are negative, which can affect the scaling of the eventual earnings surprise measure. In this specific case, no portfolio has negative earnings over the time period.

To control for other confounding variables, I also use firm-level measures and regress them on log centrality along with control variables. Firm-level earnings surprise is calculated following Shue & Townsend (2021),

$$Surp_{\mathrm{i},t} = \frac{Earnings_{\mathrm{i},t} - Forecast_{\mathrm{i},t}}{std(Forecast_{p,t})}, \tag{1.7}$$

where $Surp_{\mathrm{i},t}$ is the earnings surprise for firm i in quarter t, $Earnings_{\mathrm{i},t}$ is the reported earnings, $Forecast_{\mathrm{i},t}$ is the mean analyst forecast, and $std(Forecast_{\mathrm{i},t})$ is the standard deviation of analyst forecasts. These results are reported in Panel B of Table A.6. The control variables include market capitalization, market-to-book ratio, sales, number of shareholders, advertising expense, and number of analyst estimates. A description of each variable and its data source is given in Appendix C.1.

Once again, columns (1) and (3) report regressions for the full sample period, and columns (2) and (4) report regressions for the period after centrality is first observed. Even when controlling for several other variables, it seems that central firms have more positive earnings surprises and announcement CARs. This is consistent with the results in Panel A and suggests that they are not driven by other variables that are correlated with centrality.

Overall, it appears that centrality provides information about firms' ability to generate revenue, and specifically peripheral firms' inability to meet market participants' expectations. This evidence supports the conclusion that centrality impacts firms' revenue. Online feedback loops naturally drive traffic toward central firms and away from peripheral firms, making it difficult for peripheral firms to attract customers to their websites. Moreover, these results demonstrate that the relationship between online feedback loops and concentration goes beyond central firms being the largest and most productive firms in the economy. Peripheral firms are being significantly affected as well, although negatively.

1.5 COVID-19

In early 2020, the Coronavirus Disease 2019 (COVID-19) pandemic caused much of the brick-and-mortar economy to shut down. This created a surge in web traffic as economic activity moved online. As consumers sought online substitutes for their consumption needs, their web browsing was filtered through the online feedback loops that naturally drive traffic to the most central websites. The COVID-19 shutdown therefore presents a compelling setting in which to examine the effects of online feedback loops on the distribution of web traffic and economic activity.

1.5.1 Web Traffic

I first examine how web traffic was distributed during the COVID-19 shutdown. Figure B.3a) reports total weekly web traffic for all firm websites from January 2020 through May 2020. Beginning in late January and through February, web traffic began to marginally increase as the virus was beginning to spread across the U.S. (the first reported case in the U.S. was on January 20, 2020) and many other parts of the world, most notably China. This trend increased more dramatically through early March, with the first state of emergency declared in Florida on March 1st, 2020. Web traffic then reaches it's highest point in mid-to-late March, coinciding with the first stay-at-home orders issued by states, led by California on March 19th, 2020. There is a slight decline in web traffic from this point through April, and then a slight increase in May, consistent with the reported 8.2% increase in consumer spending that took place that month.[21]

A major concern with respect to online feedback loops is that as economic activity moved online, a disproportionate amount of the increased activity may be pushed to the central firms' websites. Indeed, this is what happened. Figures B.3b) and B.3c) split the sample into central and non-central firms, respectively, and examine web traffic over this period. Central firms are defined as the most central decile of firms within each industry as of December 2019. Web traffic is then cumulated for central and non-central firms each week. The figures show that virtually all of the increase in web traffic took place on the central firms' websites. Web traffic increased for non-central firms through February, but then decreased significantly through March, only beginning to recover by the end of May. The significant difference between central and non-central firms underscores the fundamental aspect of digital markets

[21]↑Bureau of Economic Analysis, https://www.bea.gov/data/consumer-spending/main.

that makes them so prone to concentration. High returns to scale, network effects, search engines, and consumer biases all drive consumer activity to only a handful of platforms.

To control for other possible determinants of the difference in web traffic between central and non-central firms, I perform a difference-in-difference test of web traffic around the COVID-19 shutdown. I first sort firms into centrality deciles within each industry, and then combine web traffic each week for each decile. I then define the *Post* period as any week that began after March 1st, 2020. I create a dummy variable equal to one if the decile of firms is the most central decile in that industry ($Central$), and interact $Central$ with the *Post* dummy variable to measure the difference in web traffic between central and non-central firms before and after the COVID-19 shutdown. I also include several control variables, measured as of the end of 2019, and interact them with the *Post* dummy variable as well. These controls include market capitalization, book-to-market ratio, sales, advertising expenditure, and intangible share. Accounting variables are winsorized at the 1% and 99% levels, standard errors are clustered by industry-decile, and all variables are standardized to facilitate interpretation.

The results are reported in Panel A of Table A.7. Column (1) reports the regression without any fixed effects. In general, central firms receive more web traffic than non-central firms, and the significant negative coefficient on the *Post* dummy variable indicates that non-central firms saw approximately a 12% drop in web traffic during the COVID-19 shutdown. The coefficient on the interaction between $Central$ and *Post* is positive and statistically significant, demonstrating that central firms received significantly more web traffic than non-central firms after the COVID-19 shutdown. Column (2) adds industry-decile and industry-time fixed effects to the regression, which absorb the level differences among industry-deciles as well as the overall changes in web traffic for each industry each week. The industry-time fixed effects are especially important to control for how different industries were exposed to the economic shocks of the COVID-19 shutdown. The coefficient on the interaction between $Central$ and *Post*, however, remains positive and statistically significant. Finally, web traffic increased significantly for Amazon, Walmart, and Target during the COVID-19 shutdown as consumers were stocking up on essential goods. While the fact that web traffic was especially concentrated in only a few firms is not inconsistent with online feedback loops, column (3) removes these three firms and demonstrates that they are not driving the effect.

1.5.2 Stock Returns

I next examine the stock returns of firms during the COVID-19 shutdown. I obtain market capitalization data for each week in 2020 from Capital IQ and calculate weekly returns. Figure B.4a) reports the weekly stock returns for value-weighted portfolios of central and non-central firms, cumulated from January 2020 through May 2020. The returns of both portfolios declined dramatically in late-February, and then again in early March, consistent with news of the virus's spread across the U.S. The recovery begins almost immediately, with positive returns in early April, bringing the portfolio of central firms back to similar levels as it was at the beginning of March. The portfolio of central firms also made slight gains during the rest of April, and both portfolios saw an up-tick at the end of May, once again consistent with increased consumption that month. Holistically, central firms earned significantly higher returns than non-central firms over the period. Figure B.4b) reports the difference between the portfolios of central and non-central firms from January 2020 through May 2020. The difference begins to grow substantially in early March, and then continues to increase in April, peaking in the middle of May. This suggests that central firms earned significantly higher stock returns during the COVID-19 shutdown.

To test this possibility more rigorously, I run a regression comparing weekly returns before and after the shutdown began. Once again, I define the *Post* period as any week that began after March 1st, 2020, and interact it with firm centrality to examine how the relationship between centrality and stock returns changed during the COVID-19 shutdown. I include the same control variables previously discussed for web traffic and also interact them with the *Post* dummy variable. I include industry-time fixed effects to control for how different industries responded to the shutdown, cluster standard errors by firm, and standardize all variables to facilitate interpretation.

The results are reported in Panel B of Table A.7. The coefficient on the interaction between $Centrality$ and *Post* is positive and significant, demonstrating that central firms earned significantly higher stock returns during the COVID-19 shutdown. Economically, firms that were one standard deviation more central earned nearly 20 basis points of additional return each week on average. Column (2) then excludes Amazon, Walmart, and Target from the regression to ensure these companies are not driving the effect, and the result is virtually unchanged.

Overall, these results indicate that central firms were the primary beneficiaries of economic activity being moved online. The COVID-19 shutdown was an unprecedented moment for the modern world, and placed many economic processes on a fast track to digitization. As more of the economy continues to digitize, online feedback loops will have an even greater effect on the distribution of economic activity across firms. The shutdown gives us a glimpse into what that digital economy may eventually look like, and the evidence presented here indicates that we should expect further increasing concentration.

1.6 Advantages of the Network

The data and network used in this paper provide important advantages when considering a firm's position in the economy. One such advantage is the inclusion of private firms. This is especially beneficial when considering a statistic that is measured relative to the aggregate, such as centrality, or relative to a particular group, such as market share. Centrality measured in a network with missing firms, e.g., a network of only public firms, could potentially be biased, or at least omitting valuable information.

To demonstrate this, I create a network of public-firm websites using data for the audience overlap between public firms. I then measure centrality within the public-firm-website network and reproduce the returns analysis from Section 1.4.1. The results are reported in Panel A of Table A.8. The spread-portfolio returns are noticeably weaker than those reported in Panel A of Table A.5, both economically and statistically. It seems that the inclusion of private firms in the network improves the informativeness of centrality about public-firm returns. This underscores the importance of a holistic view of the economy that takes into account *all* firms competing for market share.

Another important feature of the data is its ability to encode information about market peers into the network structure. Identifying market peers is traditionally done using industry classifications such as SIC, NAICS, or GICS. However, recent literature has emerged that provides alternative industry definitions (Hoberg & Phillips, 2016; Lee, Ma, & Wang, 2015; Kaustia & Rantala, 2020). While traditional industries can capture many important aspects of firm similarity, the definitions are inflexible and neglect connections across industries that are potentially informative. In contrast, the network of firm websites provides the opportunity for local competitive environments, or industries, to arise naturally from the data. Centrality in the overall network is primarily a function of firms' centrality within their local environments, meaning it inherently reflects the industries that revealed by the network structure. Moreover, the inclusion of private firms allows not only the classification of private firms into alternative industries, a novel contribution to the alternative industries literature, but a more complete representation of competitive environments for public firms.

I demonstrate this advantage by measuring centrality within sub-networks of firm websites and investigating the relation between stock returns and these alternative measures of centrality. The first set of sub-networks I consider are clusters, or groups of related firms, extracted from the network of firm websites. These clusters are based on how dense the connections are in a group of firms and more naturally reflect the network structure that determines overall centrality. To extract these clusters, I employ a hierarchical grouping algorithm based on modularity optimization, proposed by Blondel, Guillaume, Lambiotte, & Lefebvre (2008). This results in 655 clusters of public and private firms, which average 157 firms in size and the largest of which contains 6,913 firms. A more in-depth discussion of the procedure, the resulting clusters, and a comparison of their performance to traditional industry definitions can be found in Chapter 2. In short, the clusters are related to traditional industry classifications but provide superior benchmarking for firms. I then create sub-networks based on each cluster, including only the websites in that cluster and only the connections between cluster residents. Finally, I calculate centrality within each cluster network, which I call Cluster Centrality (CC).

To analyze the relationship between returns and cluster centrality, I employ Fama & MacBeth (1973) regressions controlling for a number of potential return determinants. Starting with those discussed in Section 1.4.2, I add four more variables. Volatility is calculated as the standard deviation of daily returns for the stock in the previous month. Turnover is calculated as the trading volume in the prior month divided by the number of shares outstanding in the prior month. Finally, Returns (t-1) and Returns (t-7 to t-2) are the prior month and lagged prior six-month returns, respectively. They are included to control for potential short-term reversal or momentum effects. All of these variables are calculated from the CRSP database. Standard errors are Newey & West (1987) adjusted for three lags.

Panel B of Table A.8 reports the results of monthly stock returns regressed on these variables. Column (1) focuses on the natural log of cluster centrality, which produces a statistically significant coefficient. This indicates that firms that are central within their cluster tend to earn higher stock returns. Importantly, I include cluster fixed effects to examine the effect of cluster centrality within clusters. I then orthogonalize the log of centrality from the full network (overall centrality) to the log of cluster centrality, capturing the variation in a firm's overall centrality that is not due to being central in its cluster. Column (2) includes cluster centrality and orthogonalized centrality in the regression. While cluster centrality remains significant, orthogonalized centrality is not, indicating that cluster centrality is able to capture the majority of information contained in overall centrality. This is because the primary determinants of a firm's centrality are the firms closely connected to it, and these connections also determine the cluster's residents. In this way, overall centrality naturally reflects the industries that are encoded in the network structure.

I contrast these results with those from the second set of sub-networks: traditional industries, defined at the 2-digit SIC level. I create a network for each industry including only the firms in that industry and only the connections between industry residents. I then calculate centrality within each industry network, which I call Industry Centrality (IC). Similar to the analysis of cluster centrality, column (3) in Table A.8 includes log industry centrality and the resulting coefficient is statistically significant. Again, this suggests that centrality within local environments, this time expressed as traditional industries, contains information about stock returns. The inclusion of industry fixed effects is again important, as industry centrality is calculated relative to the other industry residents. Next, I orthogonalize log overall centrality to log industry centrality, and include it in column (4). Unlike with cluster centrality, the remaining variation left in overall centrality is still significantly related to returns, indicating that industry centrality is unable to fully capture the information contained in overall centrality. This highlights the value of incorporating connections across industries, as well as allowing the data to reveal market peers.

1.7 Conclusion

Rising industry concentration is one of the most significant and wide-spread changes in the U.S. economy over recent decades. Given how pervasive the change seems to be, there are likely several factors contributing to this empirical fact. This paper investigates how online feedback loops, which drive customers to the already most popular websites, contribute to rising industry concentration.

To identify which firms benefit from online feedback loops, I measure centrality in a network of firm websites. The network is built from a unique data set of website audience overlap covering more than 100,000 public and private firms. This approach offers important advantages when considering market concentration. The inclusion of private firms provides a holistic view of the economy, and the data naturally encode market peers (i.e., industries) into the network structure. Centrality in the network reflects how much a website audience is similar to, or representative of, the aggregate audience of firm websites, which is dominated by consumers. Central firms receive more benefits from online feedback loops, as evidenced by their dominance in search engine results, higher levels of web traffic, and larger web traffic growth rates.

I investigate how centrality contributes to industry concentration by affecting firms' abilities to generate revenue. I find a significant correlation between centrality and industry concentration. Central industries become more concentrated and central firms gain additional market share and become more profitable over the sample period. These results are robust to isolating the component of centrality that is determined by the centrality of other firms, mitigating concerns that the relationship is driven by omitted firm characteristics. I also find that centrality provides information about future revenue, however, it is primarily for peripheral firms that earn negative risk-adjusted returns and consistently underperform earnings forecasts during the sample period. This further mitigates concerns that omitted variables like the productivity of central firms are driving the results. Finally, using the COVID-19 shutdown as an exogenous shock to total web traffic, I find that the increase in web traffic was predominantly on central websites, and central firms earned significantly higher stock returns during the shutdown. In general, these results tell a story of online feedback loops benefiting the central firms and hurting the peripheral firms, and thus contributing to rising industry concentration.

The evidence presented in this paper reinforces a worrying trend in the current economy. Incumbent firms have been allowed to expand their control of markets and restrict competition. A decrease in antitrust enforcement likely contributes to this effect, as does the increased returns to scale brought about by new technologies. However, the underlying dynamics of the economy have also undergone significant changes. The Internet has provided unprecedented access to more content than can be consumed in a lifetime. This naturally creates a need for infrastructure to index, navigate, and prioritize content, processes that determine the underlying equations dictating the flow of consumers across the economy. The behavior of other consumers provides a valuable signal when deciding which content to prioritize, in turn influencing the behavior of consumers, and producing an inherently endogenous mechanism. An even more recent example of this is machine learning algorithms, which are trained by observing enormous amounts of consumer data. As these algorithms become more influential in consumer life, they reflect back our most common tendencies.

All of this points toward an economy that is increasingly dominated by feedback loops, which act as barriers to entry and exacerbate concentration. I demonstrate that the Internet, and the feedback loops it creates, make it significantly more difficult for peripheral firms to compete in the economy, and the effects are only becoming more significant as markets continue to digitize. This creates a difficult problem for regulators as they grapple with how to promote competition in the digital age.

2. WEB-BASED PEERS: PEER GROUPS FOR PUBLIC AND PRIVATE FIRMS

2.1 Introduction

The identification of peer firms continues to be an important issue in finance. Financial analysis relies critically on the comparison or benchmarking of firms to their peers. Such comparisons provide context for interpreting statistics and assessing the financial health, productivity, or market position of a firm. Traditional industry classifications, such as the Standard Industry Classification (SIC), the North American Industry Classification System (NAICS), or the Global Industry Classification standard (GICS), were created to classify groups of firms based on the products they produce, thus identifying peer firms. However, these classifications are not always consistent (Guenther & Rosman, 1994; Kahle & Walkling, 1996), and with firms becoming increasingly multidimensional, classification has at times been more of an art than a science.

In response to these shortcomings, a recent literature has developed that uses alternative data sources and advanced analysis techniques to identify peer firms. Hoberg & Phillips (2016) use textual analysis of 10-K filings to measure the similarity of business descriptions among firms. Lee, Ma, & Wang (2015) measure co-searches on the SEC's EDGAR database to identify firms that users tend to search for simultaneously. Kaustia & Rantala (2020) use analyst co-coverage to identify peer firms, as analysts tend to cover similar firms to reduce costs. Each paper analyzes rich datasets to uncover latent information contained in economic processes, and have proven useful in settings such as mergers and acquisitions (Hoberg & Phillips, 2010), manager evaluation (Ma, Shin, & Wang, 2018), and momentum spillovers (Ali & Hirshleifer, 2020). The reliance on financial data in each case, however, restricts these classification schemes to only public firms.

In this paper, I use an alternative data source, namely web traffic, to identify groups of peer firms for over 100,000 public *and* private firms. The data measures the extent to which two website audiences overlap, thus revealing the similarity between the products and/or services provided by the two firms as determined by website users. I use this data to create a large network of firm websites and extract clusters, or groups of closely linked firms, from the network. I refer to these clusters as "web-based peer" groups.

Conceptually, web-based peer groups have several characteristics that may provide advantages in identifying peer firms. First, web-based peer groups are constructed via a data-driven approach. It therefore relies on the "wisdom of crowds"[1] rather than individuals, such as analysts, or the companies that construct traditional industry classifications like GICS. Second, web-based peer groups are determined by website users as opposed to the firm itself (as in, for example, a 10-K filing), and so are less subject to endogenous definitions. Finally, web-based peer groups can include both public *and* private firms because they do not rely on financial data that is only available for public firms. To my knowledge, "web-based peers" is the first alternative industry classification that accommodates private firms.

I begin the analysis of web-based peer groups by demonstrating that they bare some relation to traditional industry classifications. Overall, 57.1% of a firm's 10 closest web-based peers share the same 2-digit GICS industry as the firm, and 39.3% share the same 8-digit GICS industry as the firm. The similarity between web-based peer groups and traditional industries is largest among financial firms (78.1%) and smallest among real estate firms (39.5%). These statistics show that web-based peer groups capture some of the same relations identified by traditional industry classifications, but there remains significant deviation between the two classification schemes. Whether or not this deviation is informative, however, is ultimately an empirical question.

Answering this question requires an accepted methodology by which all relevant industry classification schemes can be compared. Bhojraj, Lee, & Oler (2003) provide just such a methodology, which has become standard in this literature. The methodology compares the ability of a portfolio of peer firms to explain, or benchmark, a set of 11 firm characteristics, such as return on equity, asset turnover, and sales growth. In their paper, the authors compare SIC, NAICS, GICS, and Fama & French (1997) industry classifications, and find that GICS industries provide superior benchmarking. Since the more recent industry classifications, such as Hoberg & Phillips (2016), are not available for private firms, GICS industries

[1]↑See Surowiecki (2004); Kremer, Mansour, & Perry (2014); Chen, De, Hu, & Hwang (2014); Budescu & Chen (2015); Mollick & Nanda (2016).

appear to be the best available industry classification for private firms. I therefore compare web-based peer groups to GICS industries.

I find that equally weighted portfolios of web-based peer firms significantly outperform portfolios of GICS peer firms in benchmarking all six of the firm characteristics that can be calculated for both public and private firms, and fall short of statistical significance in only two of the five firm characteristics that can only be calculated for public firms. Portfolios of web-based peer firms weighted by their website overlap score do even better, with the improvement failing to reach statistical significance in only one public-firm variable (price-to-earnings ratio). Taken holistically, the results demonstrate that web-based peer groups provide significant advantages in identifying peer firms, especially for private firms.

To further demonstrate the benefits of web-based peer groups, I next apply them to a setting in which they are uniquely qualified: identifying IPO waves. There is considerable evidence that IPOs cluster both over time and within groups (Ibbotson & Jaffe, 1975; Ritter, 1984), a phenomenon referred to as IPO waves. Traditionally, the group of firms affected by an IPO wave has been identified using industry definitions like SIC. Given that web-based peer groups outperform traditional industry definitions in many aspects, web-based peer groups may also provide additional information in identifying which firms are influenced by IPO waves. Additionally, it is the only alternative industry definition that can be used in this setting as firms considering an IPO are necessarily private.

I find that firms are significantly more likely to complete an IPO if more of their web-based peers have completed an IPO in the past 12 months. This result persists with the inclusion of GICS $\times$ Time fixed effects, and therefore is distinct from previously documented IPO clustering over time and within traditional industries. I then investigate the post-IPO performance of IPO followers compared to IPO leaders. I find that IPO followers earn 5.7% lower returns in the first month after the IPO, and 12.4% lower returns over the three months after the IPO. The lower returns in the first month are potentially consistent with a "first-mover advantage" (Tufano, 1989; Chemmanur & He, 2011) or IPO followers benefiting from lower information production costs (Benveniste, Ljungqvist, Wilhelm Jr., & Yu, 2003). The lower returns over the following months, however, are more consistent with IPO followers being lower quality firms, possibly attempting to benefit from increased valuations following the successful IPOs of their peers (Lowry & Schwert, 2002).

This paper contributes to the literature on identifying firm peers using alternative data sources by providing the first alternative definition of peer firms for both public and private firms.[2] It also contributes to the literature harnessing the "wisdom of crowds", especially through the use of Internet data. Chen, De, Hu, & Hwang (2014) find that investor opinions shared via social media can predict future stock returns. Huang (2018) finds similar return predictability extracting consumer opinions from product reviews on Amazon.com. Da & Huang (2019) investigate earnings estimates from users of an online platform, and find that individuals expressing their private information is crucial to improving the collective estimate, i.e., the wisdom of crowds. In a similar vein, this paper uses the collective web browsing activity of Internet users to reveal new information about the connections among firms.

2.2 Classifying Web-Based Peers

2.2.1 Data

To identify peers for both public and private firms, I employ a unique dataset of website audience overlap from Alexa Internet, Inc. (henceforth Alexa).[3] Alexa is a wholly owned subsidiary of Amazon.com, Inc. and offers commercial web traffic and analytics for the large majority of active websites. Alexa's data are obtained from multiple sources, including search-engine results, third-party providers, and data provided by millions of individuals and websites. The data are used to estimate a wide range of statistics for web traffic aggregated at the domain level,[4] including a measure of the overlap between two website audiences.

The measure of overlap is given by an overlap score, defined at the website-pair level, and estimated from web traffic during the previous six months. The measure is the Jaccard

[2]↑See Hoberg & Phillips (2016); Lee, Ma, & Wang (2015); Kaustia & Rantala (2020); Lewellen (2015); Leung, Agarwal, Konana, & Kumar (2017).
[3]↑See www.alexa.com.
[4]↑For example, web traffic from example.com, example.com/page1, and example.com/page2 are aggregated together and reported under example.com.

similarity of each website audience pair, which is the number of visitors to website A *and* website B divided by the number of visitors to website A *or* B.

$$\text{Overlap}(A, B) = \frac{|A \cap B|}{|A \cup B|}. \tag{2.1}$$

Alexa provides an overlap score between 1 and 100 for each website pair, with 100 indicating an identical audience for the two websites. My data include scores for more than 10 million websites at the end of each June and December, for 2017-2019.

To identify business websites, I start with corporate websites that are reported in Compustat by public firms in the month the data is collected. Websites are not reported in a consistent manner: Some are corporate (e.g., aboutmcdonalds.com), whereas others are conglomerate websites (e.g., Yum! Brands, which owns KFC, Pizza Hut, and Taco Bell, among others, reports their website as yum.com). I then use Clearbit, Inc.'s Data Enrichment API, which identifies firm websites along with other websites that are associated with the particular firm, to enlarge the number of firm websites. To these, I add private firm websites identified by Capital IQ. This results in 411,584 unique websites associated with 354,083 public and private firms headquartered in the U.S., for which 231,561 have overlap data.[5]

2.2.2 Creating the Network and Extracting Peer Groups

I create a network of firm websites using Alexa's pairwise audience overlap scores, where the firm-pair network connections are the similarities of website audiences. Figure B.2 provides a visual representation of the network of public firms for the June 2017 data. The overwhelming size of the full network used in this paper prohibits a meaningful presentation of it, which is why the sub-set of public firms is presented. The subset is 3,850 public firm websites with 21,980 overlap scores.[6] The median firm has five direct connections to other firms and the average firm has between 11 and 12. The firm with the largest number of

[5]↑This corresponds to 6,879 public firms and 224,682 private firms.

[6]↑Public firms with connections only to private firms are removed from the graph. The difference between the number of public firms in the full network (6,879) and the number of public firms in the public-firm-only network (3,850) highlights the importance of including private firms in the network.

direct connections has 144. I use a Fruchterman & Reingold (1991) algorithm to determine a websites position relative to other websites to which they are most closely connected.

I then extract clusters, or groups of similar firms, from the network using a hierarchical grouping algorithm based on modularity optimization, proposed by Blondel, Guillaume, Lambiotte, & Lefebvre (2008). The "hierarchical" aspect refers to the fact that the algorithm does not attempt to create a given number of clusters, but rather chooses the number of clusters to maximize a particular statistic of network structure: in this case, the modularity score. Modularity measures the effectiveness of a particular division of the network into groups. In essence, it creates a random distribution of connections and then calculates the difference between the fraction of actual connections and the fraction of randomly distributed connections that are within the potential groups. By measuring this statistic across a wide set of possible groupings, the algorithm can identify the set of groups that are most meaningful to the specific network. I refer to these groups as "web-based peer" groups.

The resulting web-based peer groups for the network of public firms can be seen in Figure B.2, where firms in the same peer group are similarly colored. When identifying peer groups, the hierarchical grouping algorithm can be tuned to produce larger or smaller groups by adjusting a resolution parameter. I use a resolution of 0.1 when identifying groups in the network of public and private firms, as it provides a balance between the size of the peer groups and the computational burden when calculating network statistics.[7] This produces 449 web-based peer groups. The average number of firms in each group is 222, the median number of firms is 169, and the largest group contains 1,247 firms. For comparison, there are 158 8-digit GICS industries, with an average of 620 firms and a median of 314 firms in each industry based on the set of firms used in my analysis. The largest 8-digit GICS industry contains 6,776 firms.

2.2.3 Calculating Network Statistics

Once the web-based peer groups are identified, I create a subnetwork for each group, comprised only of firms in that group and their connections to other firms in that group. I

[7] ↑Results are similar when using resolutions of 0.01 or 1.

then calculate two network statistics. The first is the effective overlap score for each pair of firms in each subnetwork, i.e., peer group. For firms that have a direct connection, i.e., a reported overlap score, the effective overlap score is simply the reported overlap score. Many pairs of firms do not have a direct connection, however, so I use a Dijkstra (1959) algorithm to find the shortest $Distance$ between any two unconnected firms in the peer group. The algorithm works by finding the path through the subnetwork, moving from firm to firm, that minimizes the sum of the weighted connections required to connect the two firms. When employing this algorithm, I define network weights as the inverse overlap score, such that

$$\text{Distance}_{a,b} = \min\left[\sum_{[\mathrm{i,j}]\in P} \frac{1}{\text{Overlap Score}_{\mathrm{i,j}}}\right], \tag{2.2}$$

$$\text{Effective Overlap Score}_{a,b} = \frac{1}{\text{Distance}_{a,b}}, \tag{2.3}$$

where P is a path between firm a and firm b and $[\mathrm{i}, \mathrm{j}]$ is a connection between firm i and firm j along path P. The effective overlap score for two unconnected firms in the same peer group is then the inverse of this $Distance$ measure.

The second network statistic I calculate is the eigenvector centrality of each firm within the peer group. There are a number of alternative ways to calculate network centrality, the most common being eigenvector, degree, closeness, betweeness, and Katz centrality.[8] The hallmark of eigenvector centrality is that a firm's centrality is not only based on the number and strength of connections it has, but also the centrality of the firms it is connected to. This means that a firm can be central to its peer group even if it has relatively few connections, provided it is well positioned in the peer group such that those connections are with other central firms.

I calculate eigenvector centrality by extracting the principal eigenvector from the adjacency matrix following Bonacich (1972). To see this mathematically, let C_{i} denote the

[8]↑Valente, Coronges, Lakon, & Costenbader (2008) show that all of these measures of centrality are usually highly correlated with each other in empirical networks.

eigenvector centrality of firm i in the network G defined by adjacency matrix $\mathbf{A}$. Defining C_{i} as a function of the centrality of the other firms in the network gives

$$C_{\mathrm{i}} = \frac{1}{\lambda} \sum_{\mathrm{j} \in G} A_{\mathrm{i,j}} C_{\mathrm{j}}, \tag{2.4}$$

where λ is a normalizing constant and $A_{\mathrm{i,j}}$ is the connection between firm i and firm j. In vector form, this becomes

$$\mathbf{AC} = \lambda \mathbf{C}, \tag{2.5}$$

which is the familiar eigenvector equation where λ is the eigenvalue corresponding to eigenvector $\mathbf{C}$ of the adjacency matrix. In essence, a firm's eigenvector centrality within the peer group reflects how well its similarities with peer firms explain the overall similarity of firms in the peer group. Therefore, a firm's centrality represents how much their website audience is similar to, or representative of, the typical website audience for firms in the peer group.

Finally, I calculate the clustering coefficient of each peer group. This statistic measures the density of connections within each group, revealing the extent to which the peer group as a whole shares the same website audience. Mathematically, it is calculated as the fraction of triplets in the subnetwork that are "open." A triplet is defined as a set of three nodes that have either two ("open") or three ("closed") connections. Peer groups with a higher clustering coefficient are therefore comprised of firms that have more overlapping website audiences.

2.3 Comparing Web-Based Peers and Traditional Industries

Generally speaking, the goal of identifying peer firms is to provide a benchmark for a firm's characteristics, performance, and business decisions. When comparing definitions of peer firms, such as industries, the superior definition should provide superior benchmarking for firms along several dimensions. Bhojraj, Lee, & Oler (2003) provide a formalized method of comparing industry classifications, including a standard set of dimensions (variables) along which to measure how firms compare to their peers. Lee, Ma, & Wang (2015) and Kaustia & Rantala (2020) use this method to compare traditional industries to their definitions of peer firms, which are based on co-searches in the SEC's EDGAR system and analyst co-coverage, respectively. In this section, I use this method to compare web-based peer groups to GICS industries, which Bhojraj, Lee, & Oler (2003) find to be the best performing of the traditional industry classifications.[9]

I start by providing intuitive support for web-based peers, using the ten most overlapping firm websites for Bank of America, Walmart, Microsoft, and Alcoa as illustrative examples, shown in Figure B.1. Bank of America overlaps with other financial companies: Chase, Wellsfargo, and USBank; Walmart is closely related to other retailers: Target, Best Buy, and Home Depot; Microsoft overlaps with other information technology companies: Apple, Amazon, and Adobe; and Alcoa is related to other aluminum companies: Arconic (which was spun off from Alcoa in 2016), Alumina Limited, and Clinton Aluminum. Although overlapping firms are generally as you might expect, they cannot be assumed. For example, Walmart and Target significantly overlap, however, Walmart's GICS classification is Food, Staples, and Retailing (301010), whereas Target's is Multiline Retail (255030).

This divergence between traditional industry classifications, such as GICS, and a more data-driven connection, such as audience overlap, is a key feature of the web-based peer groups. Table A.9 summarizes the similarities between web-based peers and GICS industry classifications. The table reports averages by peer firm rank, i.e., the most overlapping firm through the tenth most overlapping firm, as well as the average of the full sample. The first column shows the average overlap for each group, which decreases as firms become more distant. The next four columns represent the percentage of firms that have the same GICS industry classification at the 2-, 4-, 6-, and 8-digit level. As expected, a smaller percentage of firms share more specific industry classifications, and the percentage is decreasing across peer firm rank for all levels of classification. Overall, 57.1% of the top 10 closest web-based peers share a GICS 2-digit industry code with a given firm, suggesting that although web-based peers are related to traditional industry measures, there is significant divergence.

Table A.10 reports similar statistics for the 10 closest web-based peers aggregated by the given firm's GICS 2-digit industry. As evidenced by the average overlap, Financial firms overlap the most with each other, followed by Consumer Discretionary firms and Utilities

[9]↑These include SIC, NAICS, GICS, and Fama & French (1997) industries.

firms. Financial firms are also the most likely to share the same GICS 2-digit industry with their web-based peers, while Real Estate firms are the least likely. In fact, Real Estate and Communication Services are the only two industries in which less than half of firms share the same industry as their web-based peers. Within each group, firms are less likely to share the same GICS industry with their web-based peers as the industry becomes more specific (e.g., GICS 8-digit industry), as in the full sample. The fall-off is largest for firms in the Health Care and Information Technology industries, and smallest for firms in Consumer Discretionary.

2.3.1 Benchmarking Peers

Finally, I compare web-based peers to GICS 8-digit industries in their ability to explain the cross-section of several financial variables. To compare the two measures of peer firms, I gather financial data from Capital IQ, Compustat, CRSP, and IBES databases. I use these data to calculate: Return on Net Operating Assets, Return on Equity, Asset Turnover, Profit Margin, Leverage, Sales Growth, Returns, Median Analyst Estimate, Price-to-Book Ratio, Enterprise Value, and Price-to-Earnings Ratio. Each variable is described in Appendix D.1, including how it is calculated and the relevant data sources. Of these variables, the first six are available for a subset of both public and private firms, while the last five variables are only available for public firms. I then run monthly or quarterly regressions, depending on the frequency of data availability, of firm-level variables on a portfolio of 10 peer firms, and average the R^2's across all regressions.

For the web-based peers measure, the portfolio consists of the 10 most overlapping firms. I run separate tests for an equal-weighted portfolio of web-based peers (EW) and a portfolio that is weighted by the overlap scores (OW). For the GICS 8-digit industries, the portfolio consists of 10 equally weighted, randomly selected firms in the same GICS 8-digit industry following Lee, Ma, & Wang (2015). I repeat this process 1,000 times to ensure that the results are not a function of the random selection. The regressions are run from 2011 to 2019, applying the measure of web-based peers calculated in June 2017 to all observations from 2011 to 2017. It is important to acknowledge that this assumes web-based peer groups are sufficiently stable over this period so that any potential look-ahead bias does not significantly bias results. While this may not be true of all time periods, it is more reasonable for 2011 to 2017, which saw steady economic growth in the middle of the longest bull market in U.S. history.

Table A.11 reports the average R^2's of the cross-sectional regressions for each variable. The last two columns report the t-statistic for the difference between the R^2's for the portfolio of GICS 8-digit peers and the equal-weighted or overlap-weighted portfolio of web-based peers, respectively. Across all variables, the portfolio of web-based peers produces larger R^2's than the portfolio of GICS peers, with the difference being statistically significant in all cases except for Price-to-Earnings ratio. For example, the portfolio of GICS peers explains 5.15% of the variation in firms' Return on Net Operating Assets, while the portfolio of web-based peers explains between 9.53% and 10.28% of the variation. Moreover, the overlap-weighted portfolio of web-based peers outperforms the equal-weighted portfolio of web-based peers for all variables, demonstrating that overlap scores contain valuable information for benchmarking firms.

Overall, the results in this section show that the data on website-audience overlap are economically relevant for identifying peer firms. Furthermore, the peer groups extracted from this data provide a superior definition of peer firms when compared to traditional industry classifications using standard methodologies. Building these groups from web traffic data also allows for private firms to be classified into web-based peer groups. This is a unique feature of web-based peer groups as an alternative industry classification, as other alternatives typically require financial data that is only available for public firms (Hoberg & Phillips, 2016; Lee, Ma, & Wang, 2015; Kaustia & Rantala, 2020).

2.4 IPO Waves

Web-based peer groups are also economically relevant to many problems that require peer group definitions, especially those involving private firms. To demonstrate this, I next examine IPO waves. These waves, or "hot IPO markets," refer to the strong autocorrelation in IPO volume and returns, and have been studied for some time (Ibbotson & Jaffe, 1975;

Ritter, 1984). The correlation between IPO volume and initial returns is positive for the market at-large (Lowry & Schwert, 2002); however, Benveniste, Ljungqvist, Wilhelm Jr., & Yu (2003) document that the correlation is negative for firms in the same industry that go public around the same time. The authors argue that firms in the same industry are subject to a "common valuation factor," and are therefore implicitly bundled by underwriters to share the costs of information production.

Firms that have similar website audiences produce related products or services, and so website overlap can reveal information about this "common valuation factor" beyond what is captured by traditional industry definitions. To investigate whether web-based peer groups can help identify which firms will go public in an IPO wave, I examine the correlation between the probability that a firm goes public and the firm's peer IPO activity in the prior 12 months. The sample is comprised of all private firms that can be classified into web-based peer groups containing at least 10 firms. This leaves 135,511 firms in the sample. I then obtain data on completed IPOs between 2011 and 2019 from Thomson ONE, of which 760 can be matched to a firm in the sample. I construct a monthly dummy variable that equals one if the firm completes an IPO in that month, and zero in all prior months. Once a firm goes public, it leaves the sample.

I measure peer IPO activity using three variables. The first variable, *Peer IPO Dummy*, is a dummy variable equal to one if any of the firm's web-based peers completed an IPO in the past 12 months. The second variable, *N Peer IPOs*, measures the number of web-based peers that completed an IPO in the past 12 months, which accounts for the fact that an IPO wave may be stronger if more peers exhibit the behavior (i.e., decide to go public). While both of these variables capture peer IPO activity, they are potentially subject to bias due to heterogeneity in peer-group size. The third variable, *Peer IPO %*, adjusts for this potential bias by measuring the percent of a firm's web-based peers that completed an IPO in the past 12 months.

Using these measures, I estimate the correlation between firm IPO activity and peer IPO activity using a linear probability model. I control for peer-group characteristics using several control variables, including the number and average age of firms in the peer group and the clustering coefficient of the peer group. Unfortunately, the lack of wide-spread financial data for private firms limits the available set of control variables at the firm level. However, I do include the age of the firm and the firm's centrality within its peer group. To control for stable differences among firms, I include firm fixed effects in some regressions. Given that my goal is to investigate the information contained in web-based peer groups beyond that in traditional industries, I also include industry-time fixed effects in all regressions, where industry is defined at the GICS 8-digit level and time is defined at the year-month level. Finally, I double cluster standard errors by web-based peer group and time.

The results of these regressions are reported in Table A.12, where variables are standardized to facilitate interpretation. Columns (1) and (2) examine *Peer IPO Dummy*, columns (3) and (4) examine *N Peer IPOs*, and columns (5) and (6) examine *Peer IPO %*. The first column for each variable (columns (1), (3), and (5)) includes industry-time fixed effects, and the second column for each variable (columns (2), (4), and (6)) includes both industry-time and firm fixed effects. Including both sets of fixed effects absorbs the control variable for firm age (*Age*), and so it is dropped in those regressions.

The results show that firms are significantly more likely to complete an IPO if their peers have completed an IPO in the past 12 months. The coefficient on peer IPO activity is statistically significant in all regressions except for *Peer IPO Dummy* when including firm fixed effects. A one-standard deviation increase in *Peer IPO %* is associated with a 0.0055% to 0.0086% increase in the likelihood of going public (columns (5) and (6)). These numbers are actually quite large economically, as only 0.0064% of firms complete an IPO each month on average in the sample period.

It is worth noting that the inclusion of industry-time fixed effects also implies that these results are distinct from defining peer groups by traditional industries. If one were to construct a similar variable measuring the IPO activity of firms in the same industry over the past 12 months, it would be absorbed by these fixed effects. This means the information contained in web-based peers goes beyond that of a simple reshuffling of traditional industries. Moreover, while private firms in the sample prevent the inclusion of most firm-level control variables in the regressions, *Centrality* correlates with many of these variables, such as total assets. I find a significant positive relationship between *Centrality* and firms going public, which is consistent with what one would expect given the correlation between

Centrality and many omitted variables. It also, however, hints at the possibility that firms better positioned within their peer group are more likely to decide to go public.

2.4.1 Post-IPO Performance

The decision to go public following one's peers can have several implications for post-IPO performance. Benveniste et al. (2003) document a negative correlation in post-IPO returns within peer groups, defined as traditional industries. The authors argue that this is a result of IPO followers facing lower information production costs. Information about their peers, gathered by investors during the peer IPOs, may allow IPO followers to require less underpricing to compensate investors for gathering information.

Post-IPO returns for IPO followers may also be affected by other aspects. Firms rushing to go public to take advantage of high valuations, as suggested by Lowry & Schwert (2002), may incur additional costs to do so or cut corners in advertising their IPO to potential investors. Their business may also not be as developed as their peers, and so investors may be less bullish when pricing their equity. There is also a well documented "first-mover advantage" (Tufano, 1989; Chemmanur & He, 2011) wherein the first firm to offer investors the opportunity to invest in a new product or industry is able to raise additional capital.

All of these possibilities would suggest that post-IPO stock returns for IPO followers may be lower than IPO leaders. To investigate this, I examine stock returns in the one, three, and six months following the completion of an IPO. I include *Peer IPO Dummy* to capture the difference in post-IPO returns for IPO leaders and followers. My main variable of interest, however, is *Peer IPO %*. When included with *Peer IPO Dummy*, it measures the extent to which a firm is more of an IPO follower. The resulting coefficient therefore reflects how post-IPO returns vary as a firm follows more of its peers.

I also include several control variables to control for potential differences between the types of firms that are IPO leaders and followers. I include the age, centrality, and offer principal of the IPO firm to control for firms that are older and more established when they go public. I also include the number of peers as well as the clustering coefficient of the IPO firm's cluster to control for the mechanical relationship between the likelihood that a peer went public in the past year and the peer group size or connectedness. Finally, I control for whether the IPO firm was backed by private equity or venture capital to control for the different types of firms typically targeted by these two types of funding.

At the peer-firm level, I include several controls for the type of firms that the IPO followers may be following, all averaged across peers that went public in the past year. This includes peer age, offer principal, and whether the peer firm was backed by private equity or venture capital. I also include the underwriter fees paid by peer firms as a percentage of the offer principal, and the one, three, and six month post-IPO returns for peer firms. These last three variables are important to control for firms that may observe positive post-IPO returns for peer firms as a positive signal about how the market values their product, as discussed in Lowry & Schwert (2002), and so decide to go public while valuations are high.

Table A.13 reports the results of regressing these variables on post-IPO returns from 2011-2019, during which there are 676 IPOs that have non-missing values for all of the variables. I include time fixed effects, measured as the month in which the firm went public, to control for overall fluctuations in the stock market, as well as industry fixed effects to control for time-invariant differences among industries. Standard errors are double-clustered by web-based peer group and time, and all variables are standardized to facilitate interpretation.

Column (1) reports a negative relationship between one month post-IPO returns and *Peer IPO %*. The coefficient implies that IPO followers earn approximately 6% lower returns than IPO leaders in the first month after their IPO. This coefficient more than doubles, however, when examining three month returns in column (2). A one standard deviation increase in the percentage of peers that completed an IPO in the past year corresponds to more than 12% lower three-month post-IPO returns. The difference between IPO leaders and followers levels out beyond three months but still persists, with followers having approximately 11% lower returns after six months.

The inclusion of the *Peer IPO Dummy* variable means that these estimates correspond to the marginal difference in returns for firms that follow more of their peers, not just the effect of being an IPO follower. In fact, controlling for the marginal effect of firms following more of their peers, the difference between IPO followers and leaders is positive but not significant. If results were primarily a function of a strong first-mover advantage, where the first firm to go public received superior returns, then we would expect the coefficient on *Peer IPO Dummy* to be negative and significant. Instead, the positive coefficient is consistent with the positive autocorrelation in post-IPO returns discussed by Lowry & Schwert (2002), although it is not statistically significant in this sample.

It is also possible that the lower post-IPO returns for IPO followers are a function of underpricing that IPO leaders offer to compensate investors for information production costs. This effect is discussed by Benveniste et al. (2003), who also document a negative correlation in post-IPO returns within peer groups, defined as traditional industries. If the results in Table A.13 were primarily a function of information production costs, then we would expect most of the difference in returns to emerge during the first post-IPO month. This could potentially explain why the coefficient on *Peer IPO %* is marginally statistically significant in the first month, however the coefficient doubles from the first month to the third month (6% to 12%), indicating a significant divergence in returns between IPO leaders and followers in the months after the IPO was completed.

Overall, the evidence seems most consistent with IPO followers underperforming investors' expectations compared to IPO leaders. Using web-based peer groups therefore provides a more pessimistic interpretation of the negative within-group correlation in post-IPO returns relative to previously documented evidence.

2.5 Conclusion

Peer comparison is a fundamental part of any standard financial analysis. While traditional industries have been used to identify peer firms in the past, the ad hoc nature of their classification schemes has left room for improvement. Several recent papers have created superior industry classifications utilizing new sources of data and cutting-edge analysis,[10] and have proven useful in a number of contexts.[11] However, these new industry classifications require financial data that is only available for public firms. This means that the old

[10]↑See Hoberg & Phillips (2016); Lee, Ma, & Wang (2015); Kaustia & Rantala (2015); Lewellen (2015); Leung, Agarwal, Konana, & Kumar (2017).

[11]↑See Hoberg & Phillips (2010); Hoberg, Phillips, & Prabhala (2014); Kaustia & Rantala (2015); Israelsen (2016); Ma, Shin, & Wang (2018); Ali & Hirshleifer (2020); Cao, Fang, & Lei (2021).

classification schemes are still the only available option for any analysis that includes private firms.

This paper provides a new industry classification scheme, web-based peers, based on overlapping web traffic that includes both public *and* private firms. Firms in the same web-based peer group are more likely to have the same Internet users visiting their websites, indicating the firms provide related products or services. Using a standard methodology to compare industry classifications, I find that web-based peer groups provide superior benchmarking for firm variables compared to GICS, which is the leading classification scheme otherwise available for private firms. Web-based peer groups also identify firms affected by IPO waves beyond the IPO clustering previously documented within traditional industries, further demonstrating that web-based peer groups reveal connections among firms not captured by traditional industries.

Web-based peer groups can benefit any research involving private firms. This includes analyses of spillover effects in venture capital (Schnitzer & Watzinger, 2020), liquidation events such as mergers and acquisitions or IPOs (Maksimovic, Phillips, & Yang, 2013), and the causes of rising industry concentration (Autor, Dorn, Katz, Patterson, & Van Reenen, 2020). Web-based peer groups may also be helpful in a growing literature that identifies peer effects in firm decisions, such as capital structure (Leary & Roberts, 2014), dividend policy (Grennan, 2019), and IPOs (Aghamolla & Thakor, 2021). Web-based peer groups provide a new perspective on peer firms, and the web-traffic-overlap data allows the web-based peer relationship to be defined continuously. This means the most influential peers to a firm or peer group can be identified, which may help solve the reflection problem in identifying peer effects (Manski, 1993; Gabaix & Koijen, 2021). Overall, web-based peer groups create numerous opportunities for new research into peer firms.

REFERENCES

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, *80*(5), 1,977–2,016.

Aghamolla, C., & Thakor, R. T. (2021). IPO peer effects. *Journal of Financial Economics, forthcoming.*

Ahern, K. R. (2013). Network Centrality and the Cross Section of Stock Returns. *Working Paper*.

Akcigit, U., & Ates, S. T. (2021). Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory. *American Economic Journal: Macroeconomics*, *13*(1), 257–298.

Ali, U., & Hirshleifer, D. (2020). Shared analyst coverage: Unifying momentum spillover effects. *Journal of Financial Economics*, *136*(3), 649–675.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *Quarterly Journal of Economics*, *135*(2), 645–709.

Bajo, E., Chemmanur, T. J., Simonyan, K., & Tehranian, H. (2016). Underwriter networks, investor attention, and initial public offerings. *Journal of Financial Economics*, *122*(2), 376–408.

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5,439), 509–512.

Barkai, S. (2020). Declining Labor and Capital Share. *Journal of Finance*, *75*(5), 2,421–2,463.

Benveniste, L. M., Ljungqvist, A., Wilhelm Jr., W. J., & Yu, X. (2003). Evidence of Information Spillovers in the Production of Investment Banking Services. *Journal of Finance*, *58*(2), 577–608.

Bessen, J. (2016). Accounting for Rising Corporate Profits: Intangibles or Regulatory Rents? *Boston Univ. School of Law, Law and Economics Research Paper No. 16-18*.

Bessen, J. E. (2020). Industry Concentration and Information Technology. *Journal of Law and Economics*, *63*(3), 531–555.

Bhojraj, S., Lee, C. M. C., & Oler, D. K. (2003). What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research. *Journal of Accounting Research*, *41*(5), 745–774.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), 10,008–10,020.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113–120.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, *7*(3), 265–296.

Breeden, D. T., Gibbons, M. R., & Litenberger, R. H. (1989). Empirical Tests of the Consumption-Oriented CAPM. *Journal of Finance*, *44*(2), 231–262.

Budescu, D. V., & Chen, E. (2015). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science*, *61*(2), 267–280.

Calligaris, S., Criscuolo, C., & Marcolin, L. (2018). Mark-ups in the digital era. *OECD Science, Technology and Industry Working Papers*.

Cao, S. S., Fang, V. W., & Lei, L. (2021). Negative peer disclosure. *Journal of Financial Economics*, *140*(3), 815–837.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*, *52*(1), 57–82.

Chemmanur, T. J., & He, J. (2011). IPO waves, product market competition, and the going public decision: Theory and evidence. *Journal of Financial Economics*, *101*(2), 382–412.

Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *Review of Financial Studies*, *27*(5), 1,367–1,403.

Cho, J., & Roy, S. (2004). Impact of Search Engines on Page Popularity. *Proceedings of the 13th International conference on World Wide Web*, (pp. 20–29).

Cochrane, J. H. (1996). A Cross-Sectional Test of an Investment-Based Asset Pricing Model. *Journal of Political Economy*, *104*(3), 572–621.

Covarrubias, M., Gutiérrez, G., & Philippon, T. (2019). From Good to Bad Concentration? U.S. Industries over the past 30 years. *NBER Macroeconomics Annual*, *34*, 1–46.

Crouzet, N., & Eberly, J. (2019). Understanding weak capital investment: The role of market concentration and intangibles. *Changing Market Structure and Implications for Monetary Policy, Jackson Hole Economic Policy Symposium.*

Da, Z., & Huang, X. (2019). Harnessing the Wisdom of Crowds. *Management Science*, *66*(5), 1,783–2,290.

Da, Z., & Warachka, M. C. (2009). Cashflow risk, systematic earnings revisions, and the cross-section of stock returns. *Journal of Financial Economics*, *94*(3), 448–468.

Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, Covariances, and Average Returns: 1929 to 1997. *Journal of Finance*, *55*(1), 389–406.

De Loecker, J., Eeckhout, J., & Unger, G. (2020). The Rise of Market Power and the Macroeconomic Implications. *Quarterly Journal of Economics*, *135*(2), 561–644.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2016). Declining Business Dynamism: What We Know and the Way Forward. *American Economic Review*, *106*(5), 203–207.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*, 269–271.

Doidge, C., Karolyi, G. A., & Stulz, R. M. (2017). The U.S. listing gap. *Journal of Financial Economics*, *123*(3), 464–487.

El-Khatib, R., Fogel, K., & Jandik, T. (2015). CEO network centrality and merger performance. *Journal of Financial Economics*, *116*(2), 349–382.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, *43*(2), 153–193.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical Tests. *Journal of Political Economy*, *81*(3), 607–636.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Journal of Software: Practice and Experience*, *21*(11), 1,129–1,164.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, *79*(3), 733–772.

Gabaix, X., & Koijen, R. S. J. (2021). Granular Instrumental Variables. *Working Paper*.

Grennan, J. (2019). Dividend payments as a response to peer influence. *Journal of Financial Economics*, *131*(3), 549–570.

Grullon, G., Larkin, Y., & Michaely, R. (2019). Are US Industries Becoming More Concentrated? *Review of Finance*, *23*(4), 697–743.

Guenther, D. A., & Rosman, A. J. (1994). Differences between COMPUSTAT and CRSP SIC codes and related effects on research. *Journal of Accounting and Economics*, *18*(1), 115–128.

Herskovic, B. (2018). Networks in Production: Asset Pricing Implications. *Journal of Finance*, *73*(4), 1,785–1,818.

Herskovic, B., Kelly, B., Lustig, H., & Van Nieuwerburgh, S. (2020). Firm Volatility in Granular Networks. *Journal of Political Economy*, *128*(11), 4,097–4,162.

Hoberg, G., & Phillips, G. (2010). Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies*, *23*(10), 3,773–3,811.

Hoberg, G., & Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, *124*(5), 1,423–1,465.

Hoberg, G., Phillips, G., & Prabhala, N. (2014). Product Market Threats, Payouts, and Financial Flexibility. *Journal of Finance*, *69*(1), 292–324.

Hollifield, B., Neklyudov, A., & Spatt, C. (2017). Bid-ask spreads, trading networks, and the pricing of securitizations. *Review of Financial Studies*, *30*(9), 3,048–3,085.

Huang, J. (2018). The Customer Knows Best: The Investment Value of Consumer Opinions. *Journal of Financial Economics*, *128*(1), 164–182.

Ibbotson, R. G., & Jaffe, J. F. (1975). Hot issue. *Journal of Finance*, *30*(4), 1,027–1,042.

Israelsen, R. D. (2016). Does Common Analyst Coverage Explain Excess Comovement? *Journal of Financial and Quantitative Analysis*, *51*(4), 1,193–1,229.

Kahle, K. M., & Walkling, R. A. (1996). The Impact of Industry Classifications on Financial Research. *Journal of Financial and Quantitative Analysis*, *31*(3), 309–335.

Karabarbounis, L., & Neiman, B. (2014). The Global Decline of the Labor Share. *Quarterly Journal of Economics*, *129*(1), 61–103.

Kaustia, M., & Rantala, V. (2015). Social learning and corporate peer effects. *Journal of Financial Economics*, *117*(3), 653–669.

Kaustia, M., & Rantala, V. (2020). Common Analysts-Method for Defining Peer Firms. *Journal of Financial and Quantitative Analysis, forthcoming*.

Kremer, I., Mansour, Y., & Perry, M. (2014). Implementing the "Wisdom of the Crowd". *Journal of Political Economy*, *122*(5), 988–1,012.

Kurz, M. (2017). On the Formation of Capital and Wealth: IT, Monopoly Power and Rising Inequality. *Working Paper*.

La Porta, R., Lakonishok, J., Shleifer, A., & Vishny, R. (1997). Good News for Value Stocks: Further Evidence on Market Efficiency. *Journal of Finance*, *52*(2), 859–874.

Leary, M. T., & Roberts, M. R. (2014). Do Firms Rebalance Their Capital Structures? *Journal of Finance*, *60*(6), 2,575–2,619.

Lee, C. M. C., Ma, P., & Wang, C. (2015). Search-Based Peer Firms: Aggregating Investor Perception through Internet Co-Searches. *Journal of Financial Economics*, *116*(2), 410–431.

Leung, A. C. M., Agarwal, A., Konana, P., & Kumar, A. (2017). Network Analysis of Search Dynamics: The Case of Stock Habitats. *Management Science*, *63*(8), 2,397–2,771.

Lewellen, S. (2015). Executive compensation and industry peer groups. *London Business School Working Paper*.

Lowry, M., & Schwert, G. W. (2002). IPO Market Cycles: Bubbles or Sequential Learning? *Journal of Finance*, *57*(3), 1,171–1,200.

Ma, P., Shin, J.-E., & Wang, C. C. Y. (2018). Relative Performance Benchmarks: Do Boards Follow the Informativeness Principle? *Harvard Business School Accounting & Management Unit Working Paper No. 17-039*.

Maksimovic, V., Phillips, G., & Yang, L. (2013). Private and Public Merger Waves. *Journal of Finance*, *68*(5), 2,177–2,217.

Mankiw, N. G., & Shapiro, M. D. (1986). Risk and Returns: Consumption Beta Versus Market Beta. *Review of Economics and Statistics*, *68*(3), 452–459.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, *60*(3), 531–542.

Mitchell, M. L., & Mulherin, J. H. (1996). The impact of industry shocks on takeover and restructuring activity. *Journal of Financial Economics*, *41*(2), 193–229.

Mollick, E., & Nanda, R. (2016). Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts. *Management Science*, *62*(6), 1,533–1,553.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Ozsoylev, H. N., Walden, J., Yavuz, M. D., & Bildik, R. (2013). Investor Networks in the Stock Market. *Review of Financial Studies*, *27*(5), 1,323–1,366.

Ritter, J. (1984). The "hot issue" market of 1980. *Journal of Business*, *57*(2), 215–240.

Rossi, A. G., Blake, D., Timmermann, A., Tonks, I., & Wermers, R. (2018). Network centrality and delegated investment performance. *Journal of Financial Economics*, *128*(1), 183–206.

Schnitzer, M., & Watzinger, M. (2020). Measuring the Spillovers of Venture Capital. *Review of Economics and Statistics, forthcoming.*

Shue, K., & Townsend, R. R. (2021). Can the Market Multiply and Divide? Non-Proportional Thinking in Financial Markets. *Journal of Finance, forthcoming.*

Stigler Committee on Digital Platforms (2019). Stigler Committee on Digital Platforms: Final Report.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday & Co.

Tufano, P. (1989). Financial innovation and first-mover advantages. *Journal of Financial Economics*, *25*(2), 213–240.

Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How Correlated Are Network Centrality Measures? *Connect (Tor)*, *28*(1), 16–26.

Van Reenen, J. (2018). Increasing Differences between Firms: Market Power and the Macro-Economy. *CEP Discussion Paper No 1576*.

A. TABLES

A.1 Chapter 1 Tables

Table A.1. Centrality Summary Statistics

This table reports summary statistics for eigenvector centrality calculated from a network of firm websites. Panel A reports the mean and median centrality for each GICS 2-digit industry, sorted by mean centrality, as well as for all industries together. These statistics have been multiplied by 1,000 to make them easier to read. Panel A also reports the number of firms in each industry and the most central firm in each industry. Panel B reports the ten most central firms overall, in order, along with their centrality measure.

Panel A: Industry Centrality				
Industry	Mean	Median	N Firms	Most Central Firm
All Industries	1.09	3.31E-06	5,419	
Communication Services	4.34	6.97E-05	226	Yelp Inc
Consumer Discretionary	3.87	7.10E-04	597	Target Corp
Information Technology	1.32	4.19E-05	832	RetailMeNot Inc
Consumer Staples	1.30	1.25E-06	188	Walmart Inc
Industrials	0.845	2.68E-06	621	United Parcel Service Inc
Financials	0.613	4.31E-05	894	Bank of America Corp
Real Estate	0.356	1.23E-06	245	Redfin Corp
Health Care	0.104	8.04E-07	939	CVS Health Corp
Utilities	0.039	5.44E-05	128	PG&E Corp
Materials	0.032	7.82E-07	304	Sherwin-Williams Co (The)
Energy	0.0002	1.04E-06	445	Exxon Mobil Corp

Panel B: Most Central Firms	
Website	Centrality
yelp.com	0.134
zillow.com	0.120
target.com	0.106
homedepot.com	0.099
time.com	0.099
walmart.com	0.097
ups.com	0.085
groupon.com	0.084
bestbuy.com	0.083
bankofamerica.com	0.081

Table A.2. Characteristics of Central Firms

This table reports financial characteristics for firm based on centrality. Panel A separates firms into centrality decile portfolios and reports the median (mean) of Mkt Cap, IO, ROA, B/M, Lev Ratio, Search Power, and Web Traffic (Ad Exp, BRet12, BHRet36, HHI) for each portfolio. Panel B reports regressions of Search Power and Web Traffic on centrality from July 2017 to December 2019, where $\Delta Y_{\mathrm{i},t+\tau} \equiv Y_{\mathrm{i},t+\tau} - Y_{\mathrm{i},t}$. Regressions that include firm fixed effects or industry interacted with time (year-month) fixed effects are indicated with a Y. Industries are defined at the 2-digit SIC level. Each variable and its data source is described in Appendix C.1. Standard errors are clustered at the firm level, and the t-statistic for each estimate is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Panel A: Financial Characteristics of Centrality Deciles

Centrality	Mkt Cap	Ad Exp	IO	ROA	B/M	Lev Ratio	BHRet12	BHRet36	HHI	Search Power	Web Traffic
1	194.5	16.88	56.1%	0.19%	0.468	25.1%	14.7%	7.8%	1,004.8	46	0.00957%
2	397.0	12.20	69.8%	-0.16%	0.472	27.5%	7.0%	-9.6%	903.7	47	0.00961%
3	402.0	10.08	69.1%	-0.73%	0.446	30.1%	11.7%	0.4%	920.0	47	0.01000%
4	806.0	5.72	82.5%	0.43%	0.442	30.0%	16.8%	11.3%	918.1	49	0.01078%
5	861.9	19.61	79.5%	1.63%	0.455	36.9%	20.3%	23.1%	890.9	51	0.01480%
6	673.4	21.50	79.8%	0.90%	0.493	44.8%	24.7%	35.2%	877.4	52	0.01542%
7	751.0	34.15	77.9%	1.05%	0.547	51.2%	25.9%	37.7%	812.9	54	0.02242%
8	1,391.7	20.11	85.8%	1.27%	0.548	50.7%	25.1%	35.8%	886.4	57	0.04486%
9	3,092.8	50.66	86.5%	2.93%	0.437	42.5%	20.7%	30.1%	1,013.8	63	0.16136%
10	8,037.4	230.84	81.9%	4.33%	0.328	32.6%	21.7%	28.3%	1,248.5	72	1.45000%

Panel B: Online Feedback Loops				
	(1)	(2)	(3)	(4)
	Search Power$_{t+1}$	Traffic$_{t+1}$	ΔTraffic$_{t+6}$	ΔTraffic$_{t+12}$
Centrality$_t$	3.3899***	0.0892***	0.1057***	0.0609***
	(16.41)	(9.10)	(6.47)	(2.75)
Traffic$_t$	4.4195***	0.8047***	-1.3734***	-1.6322***
	(21.59)	(38.96)	(-75.06)	(-82.58)
Time on Site$_t$	-0.1203*	0.0132***	0.0033	0.0044
	(-1.68)	(5.55)	(1.14)	(1.25)
Mkt Cap$_t$	0.0482	-0.0355***	-0.0625***	-0.1072***
	(0.35)	(-2.71)	(-2.71)	(-3.70)
Sales$_t$	0.8558***	0.0586***	0.0534**	0.0279
	(7.72)	(4.34)	(2.37)	(1.32)
Ad Expense$_t$	0.2892***	-0.0031	-0.0240	-0.0263
	(3.78)	(-0.26)	(-0.79)	(-0.93)
Ad Expense Missing$_t$	-6.4087***	-0.0312***	0.0176	-0.0032
	(-17.63)	(-3.32)	(1.43)	(-0.30)
$\vert$Returns$_t\vert$	0.1827**	-0.0041*	-0.0045**	-0.0021
	(2.01)	(-1.95)	(-2.19)	(-0.87)
Volume$_t$	0.5014***	-0.0013	0.0113	-0.0056
	(4.89)	(-0.19)	(1.12)	(-0.48)
Observations	2,145	71,402	70,535	59,400
Industry x Time FE		Y	Y	Y
Firm FE		Y	Y	Y

Table A.3. Industry Concentration

This table reports quarterly regressions of measures of the average log centrality of firms within each industry on HHI, where $\Delta HHI_{\mathrm{i},t+\tau} \equiv HHI_{\mathrm{i},t+\tau} - HHI_{\mathrm{i},t}$. Panel A focuses on $Centrality$, which is the natural logarithm of eigenvector centrality, and Panel B focuses on $(Eig - Deg)\ Centrality$, which is the difference of log eigenvector centrality and log degree centrality. The time period is the third quarter of 2017 through the fourth quarter of 2019. Industries are defined at the 2-digit SIC level. Regressions that include time (year-quarter) fixed effects are indicated with a Y. Each variable and its data source is described in Appendix C.1. Standard errors are clustered at the industry level, and the t-statistic for each estimate is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Panel A: Eigenvector Centrality				
	(1)	(2)	(3)	(4)
	HHI_{t+1}	ΔHHI_{t+2}	ΔHHI_{t+4}	ΔHHI_{t+8}
Centrality_t	220.2***	89.9***	125.4***	602.8***
	(3.10)	(4.84)	(2.96)	(4.80)
HHI_t	1,753.3***	-136.0***	-136.7	-811.3***
	(14.17)	(-3.98)	(-1.62)	(-2.79)
Mkt Cap_t	123.7	-15.4	-19.1	-307.9
	(1.03)	(-0.21)	(-0.16)	(-1.08)
Book-to-Market_t	-13.6	-15.4**	-17.1	61.8**
	(-0.67)	(-2.35)	(-1.05)	(2.03)
Sales_t	-13.9	-14.3	-48.2	-134.1
	(-0.19)	(-0.40)	(-0.66)	(-0.69)
Ad Expense_t	-55.7*	-58.0*	-109.6*	-369.2***
	(-1.67)	(-1.87)	(-1.72)	(-2.67)
Intangible Share_t	-56.1	-6.6	-1.7	-81.7
	(-1.50)	(-0.49)	(-0.07)	(-0.83)
N Firms_t	-124.6	-17.6	-2.1	-390.1
	(-1.03)	(-0.44)	(-0.03)	(-1.62)
N $\text{M\&A}_t$	40.1	31.3**	21.2	90.6
	(1.26)	(2.41)	(1.00)	(1.51)
Returns_t	-41.2	-1.7	47.9	-11.2
	(-0.42)	(-0.08)	(1.36)	(-0.13)
Turnover_t	-7.5	-22.8	-7.7	-164.4**
	(-0.24)	(-1.28)	(-0.26)	(-2.53)
Volume_t	-120.6	-1.1	73.4	736.9**
	(-0.90)	(-0.01)	(0.66)	(2.25)
Traffic_t	-97.0*	-49.6*	-87.7*	-221.9
	(-1.95)	(-1.84)	(-1.73)	(-1.64)
Observations	530	476	358	118
Time FE	Y	Y	Y	Y

Panel B: Eigenvector Centrality - Degree Centrality				
	(1)	(2)	(3)	(4)
	HHI_{t+1}	ΔHHI_{t+2}	ΔHHI_{t+4}	ΔHHI_{t+8}
(Eig - Deg) Centrality$_t$	216.6***	90.4***	123.7***	558.2***
	(2.95)	(4.82)	(2.98)	(4.23)
HHI$_t$	1,749.7***	-138.7***	-139.5*	-832.2***
	(13.89)	(-4.16)	(-1.69)	(-2.69)
Mkt Cap$_t$	115.3	-19.7	-26.3	-389.3
	(0.98)	(-0.27)	(-0.22)	(-1.35)
Book-to-Market$_t$	-17.4	-17.0**	-19.4	53.0*
	(-0.88)	(-2.65)	(-1.17)	(1.69)
Sales$_t$	-2.9	-9.6	-41.4	-66.1
	(-0.04)	(-0.26)	(-0.56)	(-0.33)
Ad Expense$_t$	-64.1*	-62.0*	-114.5*	-396.4***
	(-1.93)	(-1.98)	(-1.77)	(-2.76)
Intangible Share$_t$	-62.4	-9.4	-5.3	-86.7
	(-1.60)	(-0.69)	(-0.22)	(-0.87)
N Firms$_t$	-122.4	-14.7	-1.4	-433.4*
	(-0.99)	(-0.37)	(-0.02)	(-1.70)
N M&A$_t$	37.0	30.2**	19.4	77.2
	(1.18)	(2.33)	(0.91)	(1.26)
Returns$_t$	-39.8	-1.3	49.0	6.0
	(-0.41)	(-0.06)	(1.39)	(0.07)
Turnover$_t$	-2.5	-21.2	-5.5	-156.6**
	(-0.08)	(-1.18)	(-0.19)	(-2.31)
Volume$_t$	-118.9	-1.2	76.1	803.7**
	(-0.89)	(-0.02)	(0.70)	(2.39)
Traffic$_t$	-91.6*	-48.5*	-83.8*	-162.6
	(-1.82)	(-1.79)	(-1.69)	(-1.24)
Observations	530	476	358	118
Time FE	Y	Y	Y	Y

Table A.4. Market Share and Profitability

This table reports quarterly regressions of log centrality on log Market Share (columns (1) and (2)) and Return On Assets (columns (3) and (4)), where $\Delta Y_{\mathrm{i},t+\tau} \equiv Y_{\mathrm{i},t+\tau} - Y_{\mathrm{i},t}$. *Centrality* is the natural logarithm of eigenvector centrality, and $(Eig - Deg)$ *Centrality* is the difference of log eigenvector centrality and log degree centrality. The time period is the third quarter of 2017 through the fourth quarter of 2019. Regressions that include industry interacted with time (year-quarter) fixed effects or firm fixed effects are indicated with a Y. Each variable and its data source is described in Appendix C.1. Standard errors are clustered at the industry level in columns (1) and (2) and at the firm level in columns (3) and (4). The t-statistic for each estimate is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

	(1)	(2)	(3)	(4)
	ΔMkt Share$_{t+4}$		ΔROA$_{t+4}$	
Centrality$_t$	2.124***		0.091**	
	(2.96)		(2.37)	
(Eig - Deg) Centrality$_t$		1.348*		0.075**
		(1.78)		(2.01)
Mkt Cap$_t$	7.087***	7.162***	0.158**	0.160**
	(5.69)	(5.73)	(2.54)	(2.56)
Book-to-Market$_t$	-2.132***	-2.089***	0.344	0.308
	(-3.28)	(-3.21)	(1.56)	(1.43)
Sales$_t$	-9.367***	-9.354***	-0.134***	-0.134***
	(-4.54)	(-4.52)	(-3.15)	(-3.14)
Ad Expense$_t$	0.175	0.175	0.036	0.035
	(0.32)	(0.32)	(0.96)	(0.95)
Ad Expense Missing$_t$	0.982	0.919	0.056	0.055
	(1.08)	(1.01)	(0.66)	(0.65)
Intangible Share$_t$	-0.298	-0.300	-0.241**	-0.241**
	(-0.39)	(-0.40)	(-2.13)	(-2.13)
Returns$_t$	0.078	0.210	0.178	0.179
	(0.11)	(0.29)	(1.03)	(1.03)
Turnover$_t$	-0.047	-0.047	-0.000	0.000
	(-0.82)	(-0.81)	(-0.00)	(0.00)
Volume$_t$	-0.891	-0.867	-0.030	-0.030
	(-1.23)	(-1.20)	(-0.63)	(-0.61)
Traffic$_t$	-0.654	-0.118	0.014	0.026
	(-0.93)	(-0.17)	(0.36)	(0.70)
Observations	22,817	22,817	23,514	23,514
Industry x Time FE	Y	Y	Y	Y

Table A.5. Portfolio Returns

This table reports value-weighted monthly returns for portfolios sorted into deciles by centrality in a network of firm websites. Panel A reports results for the full sample from 2014-2019. Panel B reports results for the full sample from July 2017 to December 2019. Raw Return reflects the average return for each portfolio. CAPM is the intercept of a regression of excess portfolio return on the market risk premium. FF 3-Factor is the intercept of a regression of excess portfolio returns on the factor model from Fama & French (1993). FF 3-Factor + Mom is the intercept of a regression of excess portfolio returns on the factor model from Carhart (1997). FF 5-Factor is the intercept of a regression of excess portfolio returns on the factor model from Fama & French (2015). 10-1 is the portfolio equivalent to buying portfolio 10 and selling portfolio 1. The t-statistic for each spread portfolio is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Panel A: Full Sample					
Centrality Portfolio	Raw Return	CAPM	FF 3-Factor	FF 3-Factor + Mom	FF 5-Factor
Least Central	0.51%	-0.46%	-0.48%	-0.48%	-0.49%
2	0.58%	-0.60%	-0.34%	-0.29%	-0.32%
3	0.58%	-0.52%	-0.43%	-0.37%	-0.44%
4	0.48%	-0.59%	-0.38%	-0.32%	-0.39%
5	0.61%	-0.44%	-0.31%	-0.29%	-0.30%
6	0.66%	-0.31%	-0.26%	-0.19%	-0.26%
7	0.99%	0.12%	0.09%	0.09%	0.10%
8	0.96%	0.04%	0.08%	0.08%	0.08%
9	0.70%	-0.22%	-0.21%	-0.21%	-0.22%
Most Central	1.29%	0.30%	0.23%	0.20%	0.23%
10-1	0.77%***	0.76%***	0.70%***	0.68%***	0.72%***
	(3.37)	(3.17)	(3.00)	(2.88)	(3.08)

Panel B: July 2017 - December 2019					
Centrality Portfolio	Raw Return	CAPM	FF 3-Factor	FF 3-Factor + Mom	FF 5-Factor
Least Central	0.60%	-0.55%	-0.49%	-0.49%	-0.71%
2	0.80%	-0.56%	-0.09%	-0.08%	-0.15%
3	0.48%	-0.96%	-0.77%	-0.75%	-0.89%
4	0.39%	-0.99%	-0.64%	-0.62%	-0.84%
5	0.85%	-0.47%	-0.20%	-0.19%	-0.28%
6	0.97%	-0.11%	-0.06%	-0.04%	-0.18%
7	1.18%	0.22%	0.20%	0.20%	0.06%
8	1.06%	-0.04%	0.03%	0.03%	-0.01%
9	0.78%	-0.37%	-0.35%	-0.34%	-0.40%
Most Central	1.58%	0.39%	0.27%	0.26%	0.39%
10-1	0.99%**	0.94%**	0.76%*	0.75%*	1.10%***
	(2.59)	(2.35)	(1.83)	(1.81)	(3.19)

Table A.6. Earnings

This table reports results for the relation between centrality and firm earnings. Panel A sorts firms into centrality quintiles and reports the average portfolio-level earnings surprise, as well as the three-day cumulative abnormal return (CAR) around earnings announcements value-weighted across firms in the portfolio. Panel B reports regressions of log centrality on firm-level earnings surprise and three-day earnings announcement CARs. Portfolio-level earnings surprise is calculated as in Equation 1.6, and firm-level earnings surprise is calculated as in Equation 1.7. In both panels, columns (1) and (2) correspond to earnings surprises, and columns (3) and (4) correspond to CARs, which are adjusted for the Fama-French 3-factor model. Also in both panels, columns (1) and (3) report results for 2014-2019, while columns (2) and (4) report results for July 2017 through December 2019. Each variable and its data source is described in Appendix C.1. Standard errors are clustered at the firm level, and the t-statistic for each estimate is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Panel A: Earnings by Centrality Portfolios

Centrality Portfolio	(1) Earnings Surprise 2014 - 2019	(2) Earnings Surprise July 2017 - 2019	(3) FF3 CAR 2014 - 2019	(4) FF3 CAR July 2017 - 2019
Least Central	-0.86%	-0.42%	-0.42%	-0.91%
2	-0.84%	-0.42%	-0.02%	-0.23%
3	-0.46%	-0.31%	-0.13%	-0.31%
4	-0.19%	0.14%	-0.26%	-0.27%
Most Central	-0.07%	0.21%	0.02%	0.19%

Panel B: Regressions of Earnings on Centrality				
	Earnings Surprise 2014 - 2019	Earnings Surprise July 2017 - 2019	FF3 CAR 2014 - 2019	FF3 CAR July 2017 - 2019
Centrality	0.028***	0.029**	0.033***	0.055***
	(3.50)	(2.45)	(3.29)	(3.25)
Mkt Cap	0.277***	0.355***	-0.187***	-0.250***
	(6.03)	(5.37)	(-3.16)	(-2.42)
M/B	0.268***	0.296***	-0.237***	-0.242**
	(6.14)	(4.57)	(-3.89)	(-2.28)
Sales	0.017	0.025	0.082**	0.101
	(0.61)	(0.65)	(2.03)	(1.58)
Ad Expense	-0.056*	-0.059	0.007	-0.056
	(-1.66)	(-1.13)	(0.19)	(-0.87)
Ad Expense Missing	-0.421***	-0.469**	-0.019	-0.125
	(-3.26)	(-2.39)	(-0.12)	(-0.44)
N Estimates	-0.007	-0.012	0.010	0.015
	(-0.96)	(-1.06)	(1.24)	(1.05)
N Shareholders	-0.135***	-0.163***	-0.050	-0.084
	(-4.72)	(-3.88)	(-1.53)	(-1.53)
Intercept	-4.265***	-5.418***	4.602***	6.024***
	(-8.51)	(-7.35)	(6.29)	(4.83)
Observations	41,350	15,922	47,132	18,021

Table A.7. COVID-19

This table reports regressions of log centrality on weekly web traffic (Panel A) and stock returns (Panel B) from January 2020 through May 2020. In Panel A, firms are sorted into centrality deciles within each industry as of December 2019 and web traffic is cumulated by centrality decile each week. *Central* is a dummy variable equal to one if the decile is the most central decile of firms within that industry. In both panels, *Post* is a dummy variable that equals one if the observation week begins after March 1st, 2020. All other independent variables are measured as of December 2019. Column (3) in Panel A and column (2) in Panel B report regressions excluding Amazon, Walmart, and Target from the sample. Regressions that include industry interacted with time (week) fixed effects and industry interacted with decile fixed effects are indicated with a Y. Each variable and its data source is described in Appendix C.1. Standard errors are clustered at the industry-decile level in Panel A and at the firm level in Panel B. The t-statistic for each estimate is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Panel A: Web Traffic			
	(1)	(2)	(3) Ex. AMZN, WMT, TGT
Central	1.486***		
	(3.15)		
Post	-0.124***		
	(-21.09)		
Central x Post	0.110***	0.118***	0.111***
	(2.79)	(3.39)	(3.30)
Mkt Cap	0.463**		
	(2.00)		
Mkt Cap x Post	-0.011	0.040*	0.043*
	(-0.67)	(1.68)	(1.83)
B/M	-0.149*		
	(-1.82)		
B/M x Post	0.016*	0.010	0.011
	(1.97)	(0.84)	(0.92)
Sales	-0.099		
	(-0.40)		
Sales x Post	0.039**	0.002	-0.002
	(2.01)	(0.10)	(-0.10)
Ad Expense	1.017***		
	(6.62)		
Ad Expense x Post	0.016	-0.015	-0.013
	(1.22)	(-0.78)	(-0.65)
Intangible Share	0.141*		
	(1.68)		
Intangible Share x Post	0.008	-0.002	-0.001
	(1.01)	(-0.29)	(-0.02)
Observations	2,310	2,310	2,310
Industry x Time FE		Y	Y
Industry-Decile FE		Y	Y

Panel B: Returns		
	(1)	(2) Ex. AMZN, WMT, TGT
Centrality	0.0150	0.0137
	(0.26)	(0.24)
Centrality x Post	0.1938**	0.1951**
	(2.32)	(2.34)
Mkt Cap	0.3821***	0.3833***
	(3.20)	(3.21)
Mkt Cap x Post	-0.5888***	-0.5905***
	(-4.87)	(-4.88)
B/M	-0.2827***	-0.2823***
	(-3.68)	(-3.67)
B/M x Post	-0.0377	-0.0416
	(-0.51)	(-0.56)
Sales	-0.2513	-0.2559
	(-1.58)	(-1.61)
Sales x Post	0.0682	0.0694
	(1.33)	(1.35)
Ad Expense	-0.1668	-0.1680
	(-0.96)	(-0.97)
Ad Expense x Post	-0.1411	-0.1396
	(-0.81)	(-0.80)
Ad Exp Missing	0.2408**	0.2405**
	(2.17)	(2.17)
Ad Exp Missing x Post	0.1760*	0.1796*
	(1.67)	(1.70)
Intangible Share	0.4689**	0.4731**
	(2.06)	(2.08)
Intangible Share x Post	-0.0012	-0.0023
	(-0.02)	(-0.03)
Observations	54,357	54,307
Cluster	Firm	Firm
Industry x Time FE	Y	Y

Table A.8. Advantages of Network Approach

This table reports results demonstrating the advantages of the network approach. Panel A reports value-weighted monthly returns from 2014-2019 for portfolios sorted into deciles by centrality in the network of only public firm websites. Raw Return reflects the average return for each portfolio. CAPM is the intercept of a regression of excess portfolio return on the market risk premium. FF 3-Factor is the intercept of a regression of excess portfolio returns on the factor model from Fama & French (1993). FF 3-Factor + Mom is the intercept of a regression of excess portfolio returns on the factor model from Carhart (1997). FF 5-Factor is the intercept of a regression of excess portfolio returns on the factor model from Fama & French (2015). 10-1 is the portfolio equivalent to buying portfolio 10 and selling portfolio 1. Panel B reports Fama & MacBeth (1973) regressions of returns on centrality, calculated in sub-networks, from 2014 to 2018. Cluster Centrality is the log of centrality calculated within sub-networks corresponding to clusters, or groups of related firms, extracted from the network. Centrality $\perp$ CC is the residual from a regression of log centrality on log cluster centrality. Industry Centrality is the log of centrality calculated within sub-networks corresponding to 2-digit SIC industries. Centrality $\perp$ IC is the residual from a regression of log centrality on log industry centrality. Regressions that include cluster fixed effects or industry fixed effects are indicated with a Y. Each variable and its data source is described in Appendix C.1. Standard errors are Newey & West (1987) adjusted for three lags. The t-statistic for each coefficient is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Panel A: Centrality in Public-Firms-Only Network

Centrality Portfolio	Raw Return	CAPM	FF 3-Factor	FF 3-Factor + Mom	FF 5-Factor
Least Central	0.61%	-0.24%	-0.27%	-0.24%	-0.30%
2	0.90%	-0.13%	0.00%	0.05%	-0.03%
3	0.83%	-0.10%	-0.07%	-0.04%	-0.11%
4	0.81%	-0.12%	-0.11%	-0.07%	-0.14%
5	0.32%	-0.73%	-0.62%	-0.58%	-0.61%
6	0.45%	-0.53%	-0.34%	-0.30%	-0.33%
7	0.89%	0.01%	0.00%	0.00%	0.05%
8	0.86%	-0.19%	-0.14%	-0.12%	-0.15%
9	1.22%	0.21%	0.29%	0.25%	0.31%
Most Central	1.28%	0.31%	0.21%	0.18%	0.20%
10-1	0.67%***	0.55%**	0.47%*	0.42%*	0.50%**
	(2.70)	(2.17)	(1.84)	(1.68)	(2.00))

Panel B: Centrality in Sub-Networks				
	(1)	(2)	(3)	(4)
Industry Centrality	0.163**	0.129*		
	(2.32)	(1.95)		
Centrality ⊥ IC		0.164**		
		(2.52)		
Cluster Centrality			0.119**	0.130**
			(2.01)	(2.05)
Centrality ⊥ CC				0.046
				(0.64)
Mkt Cap	-0.135	-0.145	-0.275*	-0.320**
	(-0.82)	(-0.86)	(-1.76)	(-2.02)
M/B	0.020	0.020	-0.013	-0.020
	(0.21)	(0.21)	(-0.15)	(-0.23)
Sales	0.285	0.282	0.536***	0.523***
	(1.58)	(1.57)	(3.04)	(2.93)
N Shareholders	-0.004	-0.003	-0.033	-0.040
	(-0.06)	(-0.05)	(-0.57)	(-0.69)
Ad Exp	-0.097	-0.101	-0.033	-0.066
	(-1.32)	(-1.41)	(-0.42)	(-0.91)
Ad Missing	-0.232	-0.227	-0.354**	-0.347**
	(-1.49)	(-1.46)	(-2.34)	(-2.26)
Volatility	-0.169	-0.170	-0.143	-0.147
	(-1.11)	(-1.11)	(-1.09)	(-1.13)
Turnover	-0.092	-0.091	-0.021	-0.026
	(-0.45)	(-0.44)	(-0.10)	(-0.12)
Returns (t-1)	-0.139	-0.138	-0.147	-0.146
	(-1.48)	(-1.47)	(-1.38)	(-1.36)
Returns (t-7 to t-2)	0.141	0.142	0.171*	0.176*
	(1.42)	(1.41)	(1.71)	(1.74)
Industry FE	Y	Y		
Cluster FE			Y	Y
N	135,220	135,220	126,532	126,532

A.2 Chapter 2 Tables

Table A.9. Similarity Between Web-Based Peers and GICS: by Peer Rank
This table reports the percent of firms that share the same GICS industry as their peer firm, separated by peer firm ranking. The column labeled Overlap Score reports the average overlap score for peer firms with the given ranking. The columns labeled Same GICSX report the percent of firms that share the same GICS X-digit industry code as their peer firm that has the given ranking. The row labeled Total reports these statistics for all firms, while the subsequent rows report these statistics for peer firms with the given ranking (i.e., 1 through 10). A ranking of 1 implies the peer firm has the largest overlap score with the firm in question.

Peer Rank	Overlap Score	Same GICS2	Same GICS4	Same GICS6	Same GICS8
Total	15.8	57.1%	51.1%	44.4%	39.3%
1	44.0	63.7%	58.1%	51.5%	45.9%
2	20.5	60.4%	54.7%	48.0%	42.6%
3	15.7	58.8%	52.8%	46.2%	41.1%
4	13.5	57.5%	51.6%	45.0%	39.8%
5	12.3	56.6%	50.8%	44.1%	39.1%
6	11.3	55.7%	49.6%	42.9%	38.0%
7	10.6	55.1%	49.1%	42.3%	37.3%
8	10.1	54.6%	48.5%	41.7%	36.8%
9	9.6	54.0%	47.7%	40.9%	36.1%
10	9.3	54.1%	47.9%	41.0%	36.1%

Table A.10. Similarity Between Web-Based Peers and GICS: by GICS

This table reports the percent of firms that share the same GICS industry as their peer firm, separated by firm GICS 2-digit industry. The column labeled Overlap Score reports the average overlap score of the 10 peer firms with the highest overlap score for each firm, based on the firm's GICS 2-digit industry. The columns labeled Same GICSX report the percent of firms that share the same GICS X-digit industry code as the 10 peer firms with the highest overlap score for each firm, based on the firm's GICS 2-digit industry. The row labeled Total reports these statistics for all firms, while the subsequent rows report these statistics for firms in each GICS 2-digit industry.

GICS2 Industry	Overlap Score	Same GICS2	Same GICS4	Same GICS6	Same GICS8
Energy	11.4	49.9%	49.9%	45.9%	33.5%
Materials	8.0	50.1%	50.1%	42.7%	33.2%
Industrials	7.1	54.8%	46.6%	35.9%	33.4%
Consumer Discretionary	15.4	57.4%	48.5%	44.7%	42.0%
Consumer Staples	6.0	61.5%	52.0%	49.9%	45.9%
Health Care	8.9	64.8%	59.9%	48.1%	39.3%
Financials	38.2	78.1%	72.2%	69.5%	63.0%
Information Technology	6.5	54.2%	45.2%	34.9%	29.4%
Communication Services	8.3	39.8%	38.4%	29.5%	25.5%
Utilities	13.0	53.4%	53.4%	37.8%	37.5%
Real Estate	9.2	39.5%	39.5%	36.2%	19.5%

Table A.11. Benchmarking Peers

This table reports financial characteristics for firm based on centrality. Regressions that include firm fixed effects or industry interacted with time (year-month) fixed effects are indicated with a Y. Industries are defined at the 2-digit SIC level. Each variable and its data source is described in Appendix D.1. Standard errors are clustered at the firm level, and the t-statistic for each estimate is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Public & Private Variables	GICS	EW WBP	OW WBP	t(EW-GICS)	t(OW-GICS)
Return on Net Op Assets	5.15%	9.53%	10.28%	21.07***	24.56***
Return on Equity	0.64%	0.96%	1.04%	3.78***	4.56***
Asset Turnover	17.45%	24.92%	25.61%	25.56***	27.34***
Profit Margin	8.28%	12.86%	13.34%	10.55***	10.96***
Leverage	3.65%	5.97%	7.52%	10.70***	14.61***
Sales Growth	1.06%	2.48%	2.69%	8.40***	9.72***
Public Variables	**GICS**	**EW WBP**	**OW WBP**	**t(EW-GICS)**	**t(OW-GICS)**
Returns	4.64%	6.29%	7.24%	11.35***	16.09***
Median Analyst Estimate	14.24%	25.38%	26.97%	30.87***	31.96***
Price to Book	2.89%	3.39%	4.00%	3.51***	6.51***
Enterprise Value	16.20%	16.69%	16.85%	1.44	1.84*
Price to Earnings	0.46%	0.51%	0.56%	0.60	1.22

Table A.12. IPO Waves

This table reports monthly regressions of peer-firm IPO activity on firm IPO activity. *IPO Dummy* is a dummy variable that equals one if the firm completes an IPO in the given month. *Peer IPO Dummy* is a dummy variable that equals one if any peer firm completed an IPO in the previous 12 months. *N Peer IPOs* is the number of peer firms that completed an IPO in the previous 12 months. *Peer IPO %* is the percent of peer firms in the web-based peer group that completed an IPO in the previous 12 months. Each variable and its data source is described in Appendix D.1. The time period is 2011-2019. Industries are defined at the GICS 8-digit level. Standard errors are double clustered at the web-based peer group and year-month level. The t-statistic for each coefficient is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

	(1) IPO Dummy	(2) IPO Dummy	(3) IPO Dummy	(4) IPO Dummy	(5) IPO Dummy	(6) IPO Dummy
Peer IPO Dummy	0.0044**	-0.0001				
	(2.54)	(-0.08)				
N Peer IPOs			0.0109***	0.0089***		
			(9.85)	(3.11)		
Peer IPO %					0.0086***	0.0055**
					(4.66)	(2.09)
Age	-0.0008**		-0.0007**		-0.0007**	
	(-2.33)		(-2.33)		(-2.31)	
Centrality	0.0066***	0.0058*	0.0062***	0.0052*	0.0062***	0.0053*
	(4.49)	(1.84)	(4.30)	(1.67)	(4.54)	(1.70)
N Peers	0.0006	0.0009	-0.0012***	-0.0017*	0.0004	0.0002
	(1.10)	(1.06)	(-3.55)	(-1.76)	(1.04)	(0.26)
Avg Peer Age	-0.0023***	-0.0008*	-0.0007**	0.0012	-0.0011***	0.0004
	(-3.82)	(-1.68)	(-2.21)	(1.48)	(-3.49)	(0.49)
Cluster Coefficient	-0.0013**	0.0011	-0.0009**	0.0012	-0.0010***	0.0011
	(-2.29)	(1.45)	(-2.58)	(1.62)	(-3.07)	(1.46)
Observations	11,827,761	11,827,755	11,827,761	11,827,755	11,827,761	11,827,755
Industry x Time FE	Y	Y	Y	Y	Y	Y
Firm FE		Y		Y		Y
SE Clustering	WBP Group, Time	WBP Group, Time	WBP Group, Time	WBP Group, Time	WBP Group, Time	WBP Group, Time

Table A.13. Post-IPO Performance

This table reports regressions of peer-firm IPO activity on post-IPO returns. *X Month Ret* is the X-month post-IPO return. *Peer IPO Dummy* is a dummy variable that equals one if any peer firm completed an IPO in the previous 12 months. *Peer IPO %* is the percent of peer firms in the web-based peer group that completed an IPO in the previous 12 months. Each variable and its data source is described in Appendix D.1. The time period is 2011-2019. Time is defined at the year-month level and industries are defined at the GICS 8-digit level. Standard errors are double clustered at the web-based peer group and year-month level. The t-statistic for each coefficient is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

	(1) 1 Month Ret	(2) 3 Month Ret	(3) 6 Month Ret
Peer IPO Dummy	10.9527	10.9434	22.1471
	(0.74)	(0.53)	(1.25)
Peer IPO %	-5.7237*	-12.3738***	-11.3692**
	(-1.85)	(-2.82)	(-2.33)
Age	-0.3613	-2.4061	0.2840
	(-0.22)	(-0.94)	(0.12)
Centrality	-0.8043	1.0931	-2.8707
	(-0.29)	(0.27)	(-0.70)
N Peers	2.2443	3.3912	2.1476
	(0.82)	(0.79)	(0.62)
Avg Peer Age	0.2835	-2.6301	-0.1164
	(0.15)	(-1.01)	(-0.04)
Cluster Coefficient	4.4854	5.5731	5.9931
	(1.10)	(1.34)	(1.52)
PE Backed	-0.1860	-4.8822	8.9677
	(-0.05)	(-0.74)	(1.23)
Venture Backed	18.8008***	20.8561***	21.8004***
	(4.27)	(2.84)	(2.95)
Offer Principal	8.6530***	8.7165***	6.4643**
	(4.97)	(3.29)	(2.53)
Peer PE Backed	0.8916	-2.7203	0.4050
	(0.47)	(-0.86)	(0.16)
Peer Venture Backed	3.2158	0.7817	2.9641
	(1.09)	(0.18)	(0.78)
Peer Offer Principal	-6.4656	-5.3729	-11.8498
	(-0.71)	(-0.47)	(-1.30)
Peer Fees/Principal	4.2125	9.3907**	12.2694***
	(1.59)	(2.14)	(2.85)
Peer 1 Month Return	-1.2008	-3.7649	-4.4527
	(-0.36)	(-0.83)	(-0.97)
Peer 3 Month Return	0.2081	-1.2942	1.0926
	(0.07)	(-0.26)	(0.22)
Peer 6 Month Return	0.5065	3.6018	-0.4986
	(0.24)	(1.00)	(-0.12)
Observations	676	676	676
Time FE	Y	Y	Y
Industry FE	Y	Y	Y
SE Clustering	WBP Group, Time	WBP Group, Time	WBP Group, Time

B. FIGURES

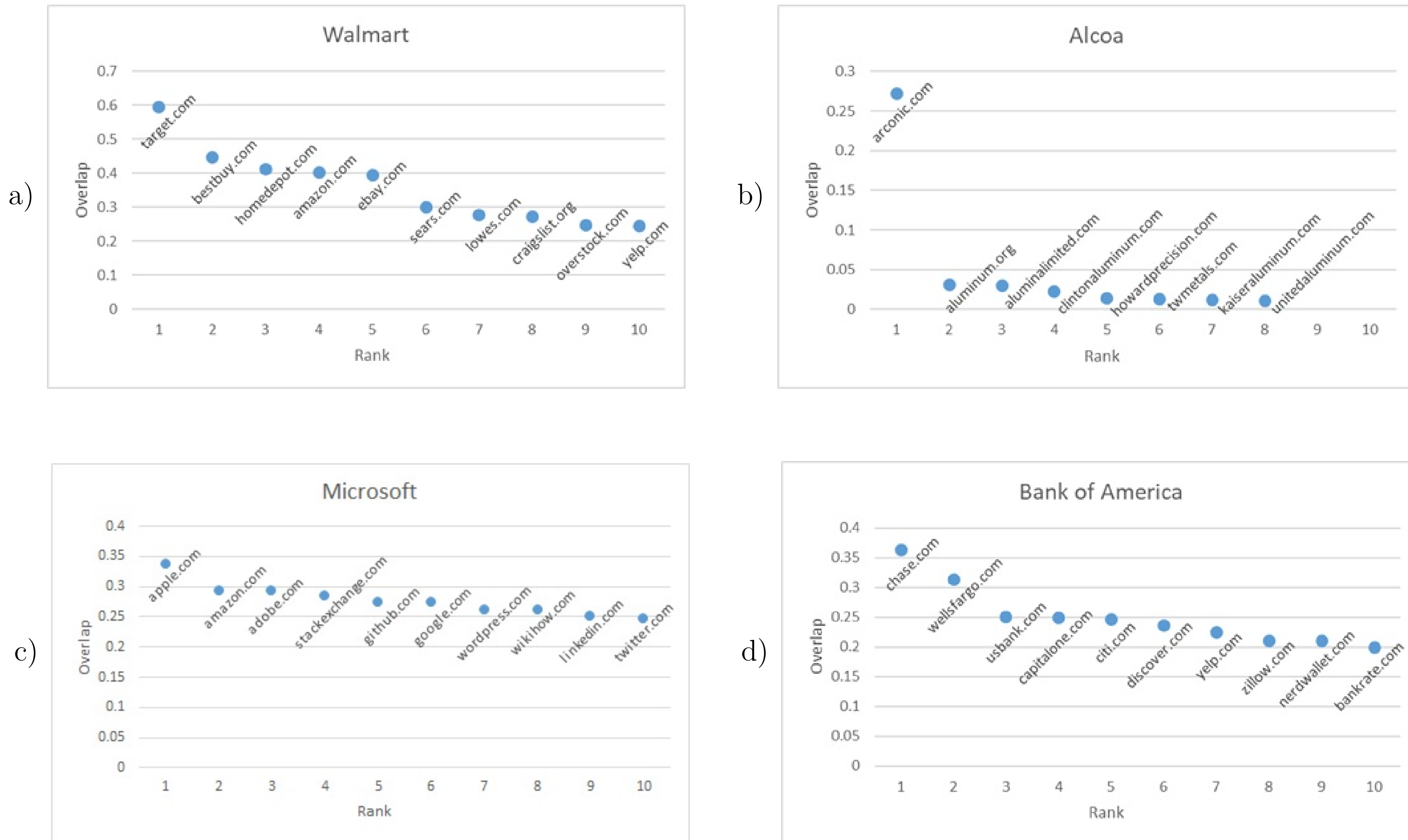


Figure B.1. Overlapping Firms Examples

This figure presents the ten most overlapping firms for Walmart, Alcoa, Microsoft, and Bank of America. In each plot, the vertical axis is the overlap score between the two websites and the horizontal axis is the rank of the website (from 1 to 10).

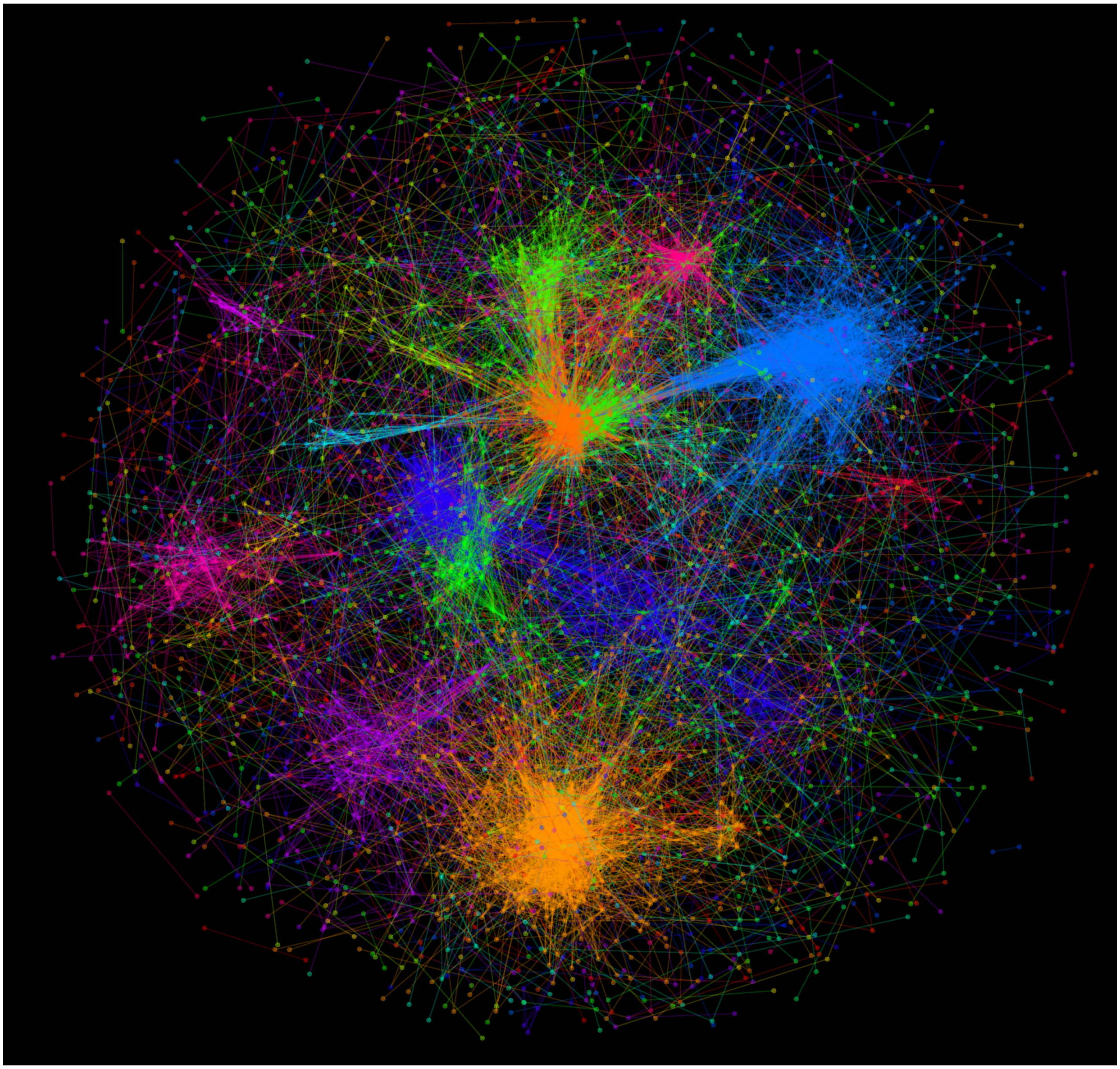


Figure B.2. Network of Public Firm Websites

This figure presents a visual representation of the network of public-firm websites corresponding to overlap data collected at the end of June 2017. It is a sub-network of the private and public firm network used to calculate centrality, as it contains only public firms and connections between public firms. A Fruchterman & Reingold (1991) algorithm is used to move websites toward those they are more connected to and away from those they are less connected to. Clusters, which are extracted from the public and private firm network using a modularity optimizing algorithm, are identified by different colors.

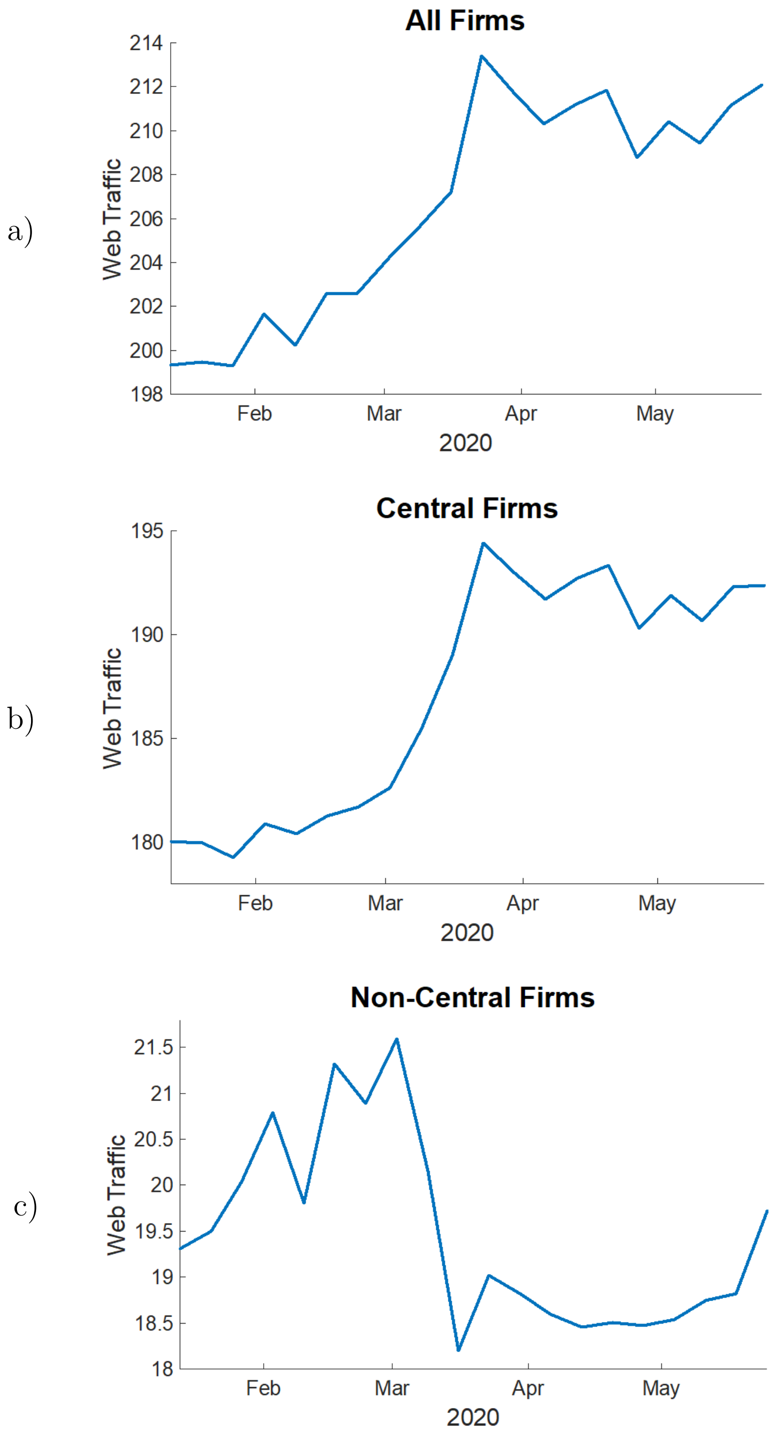
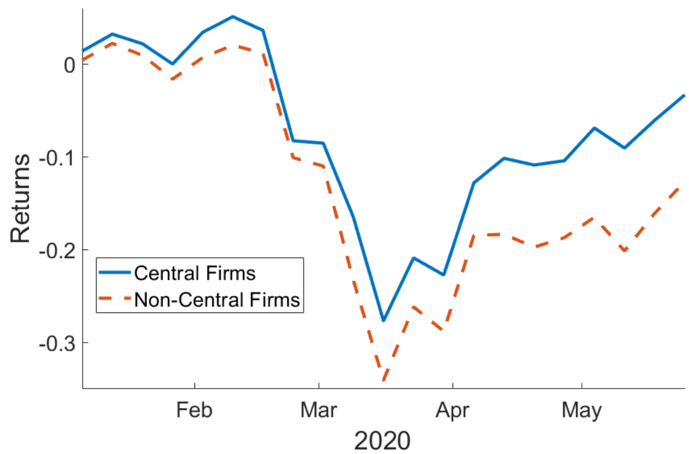


Figure B.3. Web Traffic During COVID-19

This figure presents total weekly web traffic for firm websites from January 2020 to May 2020. Figure a) presents the sum of all web traffic for firm websites. Figure b) presents the sum of all web traffic for central firms. Figure c) presents the sum of all web traffic for non-central firms. Central firms are defined as the most central decile of firms within each industry.

a)



b)

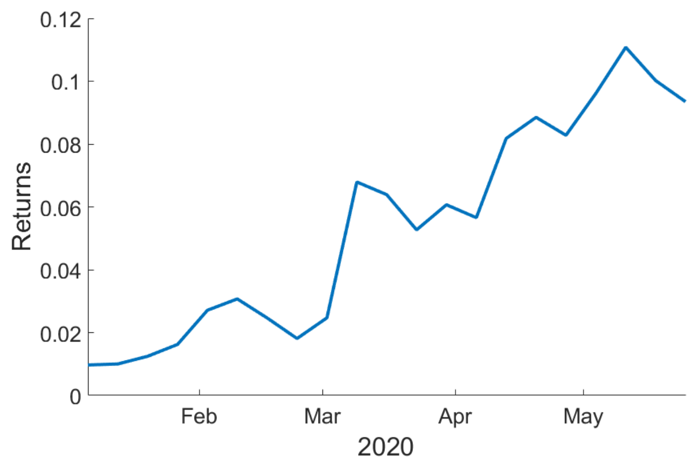


Figure B.4. Stock Returns During COVID-19

This figure presents weekly cumulative stock returns from January 2020 to May 2020. Figure a) presents stock returns for value-weighted portfolios of central and non-central firms. Central firms are defined as the most central decile of firms within each industry. Figure b) presents the difference in returns between the central and non-central portfolios.

C. SUPPLEMENTARY CHAPTER 1 MATERIAL

C.1 Data Appendix

Table C.1. Chapter 1 Variables

Description of variables for Chapter 1 and data sources.

Variable Name	Variable Description (Data Source)
Ad Expense	Advertising Expense, set equal to zero if missing (Compustat)
Ad Expense Missing	Dummy variable equal one if Ad Exp is missing
B/M (M/B)	Book-to-Market (Market-to-Book), calculated as the ratio of book equity to Mkt Cap, where book equity is calculated as in Davis, Fama, & French (2000) (CRSP, Compustat)
BHRet12	Buy-and-hold returns over the previous 12 months (CRSP)
BHRet36	Buy-and-hold returns over the previous 36 months (CRSP)
Centrality (Cluster, Industry)	Measured within the network of firm websites, or within cluster or industry sub-networks (Alexa)
(Eig - Deg) Centrality	Difference between the natural logarithm of eigenvector centrality and the natural logarithm of degree centrality, both measured within the network of firm websites (Alexa)
HHI	Herfindahl-Hirschman Index, see Equation 1.3 (Compustat)
Intangible Share	Calculated as intangible assets divided by intangible assets plus property, plant, and equipment (Compustat)
IO	Percent of Institutional Ownership (Factset Stock Ownership)
Lev Ratio	Ratio of Total Leverage to Total Leverage plus Mkt Cap (Compustat)
Mkt Cap	Market capitalization, calculated as the share price times the number of shares outstanding (CRSP)
Mkt Share	Calculated as Sales divided by total industry Sales (Compustat)
N Estimates	Number of analyst estimates (IBES)
N Firms	Number of firms in the industry (Compustat)

N M&A	Number of mergers or acquisitions in the industry (Thomson ONE)
N Shareholders	Number of shareholders (Compustat)
Returns	Stock returns (CRSP)
ROA	Return-on-assets, calculated as Net Income divided by Total Assets (Compustat)
Sales	Reported sales (Compustat)
Search Power	Estimate of how competitive the website is for key words on search engines (Alexa)
Total Assets	Total assets of the firm, AT (Compustat)
Time on Site (ToS)	Estimate of how many seconds the average visitor spent on the website (Alexa)
Turnover	Calculated as Volume divided by the number of shares outstanding (CRSP)
Volatility	Calculated as the standard deviation of daily returns (CRSP)
Volume	Trading volume (CRSP)
Web Traffic	Estimate of the percent of all Internet users that visited the website (Alexa)

C.2 Networks and Centrality

This section defines basic network terminology and details the calculation of eigenvector centrality and degree centrality. In any network, or "graph," the constituent members are referred to as "nodes," and the connection between any two nodes is called an "edge." Edges can be either "directed," indicating an asymmetric connection that applies in only one direction, or "undirected," indicating a symmetric connection that applies in both directions. Edges can also be "weighted," meaning each connection has a measure of the strength of the connection, or "unweighted," meaning each connection is of equal strength. Mathematically, a graph is denoted by a square "adjacency" matrix, where the element in the ith row and the jth column represents the edge between node i and node j (equal to one or zero for an unweighted graph or the weight for a weighted graph). For the network used in this paper, the nodes are represented by websites and the edges are represented by the overlap data, with the overlap score representing the weight of each edge. Given the symmetric nature of overlap scores, the edges are undirected.

The centrality of node i, C_{i}, is derived by first defining it recursively as a function of the centrality of node i's neighbors,

$$C_{\mathrm{i}} = \frac{1}{\lambda} \sum_{\mathrm{j} \in G} A_{\mathrm{i,j}} C_{\mathrm{j}}, \tag{C.1}$$

where $\mathbf{A}$ is the adjacency matrix defining network G, $A_{\mathrm{i,j}}$ is the weight of the connection between node i and node j, and λ is a normalizing constant. If we consider centrality evolving over time, then this is equivalent to

$$C_{\mathrm{i}}(t) = \frac{1}{\lambda_t} \sum_{\mathrm{j} \in G} A_{\mathrm{i,j}} C_{\mathrm{j}}(t-1), \tag{C.2}$$

where $C_{\mathrm{i}}(0) = 1$ for all i defines the initial centrality of all nodes and λ_t normalizes the vector for time t.[1] Writing this in matrix notation and plugging in for $C_{\mathrm{j}}(t-1)$ recursively gives

$$\mathbf{C}(t) = \frac{1}{\lambda_t} \mathbf{A}^t \mathbf{C}(0). \tag{C.3}$$

[1]↑Conceptually, this is equivalent to shocking the network with a unit vector and calculating how the shocks propagate through the network.

Equation C.3 provides a general formula for both eigenvector centrality and degree centrality. Degree centrality is calculated as $\mathbf{C}(1)$, and therefore only considers the first-order connections of nodes in the graph. In contrast, eigenvector centrality is calculated by taking $t \to \infty$, and so Equation C.3 can be rewritten as

$$\mathbf{AC} = \lambda \mathbf{C}. \tag{C.4}$$

Equation C.4 is the familiar eigenvector equation, where the constant λ is the corresponding eigenvalue to the eigenvector $\mathbf{C}$ of the adjacency matrix $\mathbf{A}$. Eigenvector centrality is therefore the principal eigenvector of the adjacency matrix (Bonacich, 1972). One can extract up to N eigenvectors and corresponding eigenvalues, where N is the cross-sectional dimension (e.g., number of firms), but the principal eigenvector is the one with the largest corresponding eigenvalue.

More conceptually, consider a square matrix $\mathbf{M}$ with N dimensions, each element representing a point of data in the N-dimensional space. The principal eigenvector of $\mathbf{M}$ is the single vector that best fits all of the data, i.e., it "points" in the direction that explains the largest possible amount of $\mathbf{M}$. Considering $\mathbf{M}$ as a whole means that the elements of the principal eigenvector reflect how each dimension contributes to the overall explanatory power of the principal eigenvector relative to other dimensions.

For the adjacency matrix corresponding to the similarity network in this paper, the N dimensions are the firms, and each element in the principal eigenvector represents how much the connections of that firm serve to explain the overall connectivity of the graph. The connections that constitute this adjacency matrix reflect how similar two website audiences are, and so the elements of the principal eigenvector reflect how the similarities of a particular firm explain the overall similarity of firms in the network. Therefore, a firm's centrality represents how much their website audience is similar to, or representative of, the aggregate audience of firm websites.

C.3 Mergers and Acquisitions

One concern with identifying a relationship between centrality and concentration is that both variables are outcomes of a long-run equilibrium reflecting industry structure that is determined by some omitted variable. To investigate this concern, I examine mergers and acquisitions (henceforth just acquisitions) as shocks to firms' access to online feedback loops. Specifically, I examine how the centrality of the target firm relates to changes in HHI over the year following the completion of the acquisition. Acquisitions represent a shock to industry structure, as well as a shock to firms' market position online, breaking this potential equilibrium. The identifying assumption is then that acquisitions of more central firms reflect a larger shock to firms' access to online feedback loops.

I collect data on all acquisitions between public companies for July 2017 through December 2018 from Thomson ONE. I require the acquiring firm to own less than 50% of the target firm prior to the acquisition, seek to purchase more than 50% of ownership, and own more than 90% of the target firm after the acquisition. I then merge this data with CRSP and Compustat to provide financial data on both the target and acquiring firms. This results in 75 acquisitions that took place between public companies during the 18 month period and satisfy all of the data requirements for both target and acquiring firms.

The regression specification compares the change in acquirer industry HHI, measured at the 2-digit SIC level, over the four quarters following the completed acquisition with target and acquirer centrality measured prior to the acquisition announcement. I control for several firm characteristics for both targets and acquirers, including total assets, sales, market share, advertising expenditure, intangible share, and industry centrality to control for the fact that more central industries have become more concentrated over the sample period. I also control for acquirer industry HHI prior to the announcement, the number of firms in the acquirer's industry, and returns for the acquirer's industry over the past 12 months to capture industry shocks that may induce M&A activity. Accounting variables are winsorized at the 1% and 99% levels, standard errors are clustered at the industry level, and variables are standardize to facilitate interpretation.

The results of these regressions are reported in Table C.2. Column (1) reports the regression without any fixed effects. The coefficient on target centrality is positive and statistically significant, indicating that industries became more concentrated following the acquisition of a more central firm. Economically, the estimate reflects an additional 5% increase in concentration when target firms are one standard deviation more central. This regression controls for many industry characteristics, but there may remain some industry heterogeneity contributing to the effect. Column (2) therefore adds industry fixed effects to the regression. Given the limited sample size, however, these fixed effects are included at the highest SIC industry level. Nonetheless, target centrality remains positive and statistical significant.

Finally, acquisitions tend to occur in waves (Mitchell & Mulherin, 1996), and it is possible that acquisitions of central target firms are followed by other acquisitions that mechanically produce increases in concentration. To account for this, I calculate an effective HHI, HHI_{Eff}, assuming that all acquisitions that took place between July 2017 and December 2019 were completed prior to July 2017. This adjusts HHI for any consolidation of sales between targets and acquirers. Column (3) reports regressions using the effective HHI measure. The coefficient on target firm centrality remains positive and statistically significant, indicating that the effect is not driven by a mechanical relationship with M&A waves. Overall, the results in Table C.2 show that industries become more concentrated during the sample period following a larger shock to the acquiring firm's access to online feedback loops, represented by the target firm's centrality.

Table C.2. Industry Concentration and M&A

This table reports regressions of the log centrality of target and acquiring firms on industry HHI following the completion of a merger or acquisition. The regressions include mergers and acquisitions between public companies that took place between July 2017 and December 2018. HHI is calculated using 2-digit SIC industries. $\mathrm{HHI}_{\mathrm{Eff}}$ is the effective HHI, calculated using sales data that is aggregated assuming all mergers and acquisitions that took place between July 2017 and December 2019 were completed before July 2017. Regressions that include industry fixed effects, measured at the highest SIC level, are indicated with a Y. Each variable and its data source is described in Appendix C.1. Standard errors are clustered at the industry level, and the t-statistic for each estimate is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

	(1) Δ HHI$_{t+4}$	(2) Δ HHI$_{t+4}$	(3) Δ HHI$_{\text{Eff},t+4}$
Target Centrality	24.241**	18.631**	17.756*
	(2.22)	(2.15)	(1.80)
Acquirer Centrality	-6.271	-2.122	-5.181
	(-0.70)	(-0.24)	(-0.50)
Target Total Assets	-6.388	33.777***	32.409***
	(-0.39)	(2.81)	(2.94)
Target Sales	-7.463	-65.249***	-51.342***
	(-0.32)	(-3.87)	(-3.32)
Target Mkt Share	16.196	50.052***	35.226**
	(0.86)	(3.09)	(2.35)
Target Ad Expense	-4.335	5.491	0.444
	(-0.72)	(1.39)	(0.06)
Target Ad Expense Missing	2.175	-9.844	-4.312
	(0.25)	(-1.40)	(-0.53)
Target Intangible Share	4.751	7.276**	5.719**
	(1.06)	(2.54)	(2.39)
Target Industry Centrality	-4.244	-1.114	-2.026
	(-0.60)	(-0.23)	(-0.37)
Acquirer Total Assets	-11.847	-12.522	-5.021
	(-0.74)	(-0.79)	(-0.33)
Acquirer Sales	-4.235	-7.911	-28.701
	(-0.18)	(-0.39)	(-1.34)
Acquirer Mkt Share	13.736	-1.631	12.829
	(0.60)	(-0.10)	(0.66)
Acquirer Ad Expense	-0.530	0.081	-0.350
	(-0.23)	(0.05)	(-0.18)
Acquirer Ad Expense Missing	-5.807	-15.943***	-59.558***
	(-1.20)	(-4.85)	(-3.31)
Acquirer Intangible Share	-12.367	-27.435***	-20.968*
	(-1.12)	(-3.53)	(-1.92)
Acquirer Industry HHI	-20.771	-50.227***	-49.529***
	(-1.55)	(-5.56)	(-6.18)
Acquirer Industry Centrality	19.867**	36.472***	33.367***
	(2.53)	(4.27)	(3.22)
Acquirer Industry Returns	3.067	16.094**	13.035*
	(0.42)	(2.11)	(2.01)
Acquirer Industry N Firms	4.947	-15.492**	-16.627**
	(0.52)	(-2.15)	(-2.10)
Observations	75	74	74
Industry FE		Y	Y

C.4 Stock Return Robustness

In this appendix, I run a number of tests to check the robustness of the relation between centrality and stock returns and examine potential alternative explanations.

C.4.1 Predicted Centrality

One concern regarding the relationship between centrality and stock returns is the potential for look-ahead bias in the extended sample period. I investigate this concern by creating an approximation of centrality using daily web traffic data. Centrality is driven by multiple dimensions of each firm's web traffic. For each website, Alexa reports: a web-traffic ranking (Alexa Rank, AR), the percentage of all Internet users who visited the website (Visitors Percent, VP), the percentage of all pageviews across the Internet that took place on the website (Pageviews Percent, PP), the average number of pageviews per visitor (PPV), the average time spent on the website per visitor (Time on Site, ToS), and the percentage of visits that resulted in only one pageview (Bounce Percent, BP). Each variable is reported daily from 2014 to 2019.

I use these variables to create a predicted daily measure of centrality. To first calibrate the model, I regress centrality on the average or median of each web traffic variable in the corresponding month in which centrality was calculated. Not all of the web traffic statistics are always available, so I account for missing data by filling in the missing observation and setting a dummy variable equal to one.[2] I then take the resulting coefficients from the regression and use them to predict centrality using the daily observations of each web traffic variable. The model is summarized below,

$$\begin{aligned} C_{\mathrm{i},m} = \hat{\alpha} &+ \hat{\beta}_1 \ln(AR_{\mathrm{i},m}) + \hat{\beta}_2 \ln(VP_{\mathrm{i},m}) + \hat{\beta}_3 \ln(PP_{\mathrm{i},m}) + \hat{\beta}_4 PPV_{\mathrm{i},m} + \hat{\beta}_5 ToS_{\mathrm{i},m} + \hat{\beta}_6 BP_{\mathrm{i},m} \\ &+ \hat{\beta}_7 I(AR_{\mathrm{i},m} \text{ missing}) + \hat{\beta}_8 I(VP_{\mathrm{i},m} \text{ missing}) + \hat{\beta}_9 I(PP_{\mathrm{i},m} \text{ missing}) \\ &+ \hat{\beta}_{10} I(PPV_{\mathrm{i},m} \text{ missing}) + \hat{\beta}_{11} I(ToS_{\mathrm{i},m} \text{ missing}) + \hat{\beta}_{12} I(BP_{\mathrm{i},m} \text{ missing}) + \hat{\epsilon}_{\mathrm{i},m}, \end{aligned} \tag{C.5}$$

[2] ↑For Alexa Rank, I set missing observations to 1,000,000, as Alexa ranks websites up to the one-millionth website. The remaining variables are set to zero if missing.

$$\hat{C}_{\mathrm{i},d} = \hat{\alpha} + \hat{\beta}_1 \ln(AR_{\mathrm{i},d}) + \hat{\beta}_2 \ln(VP_{\mathrm{i},d}) + \hat{\beta}_3 \ln(PP_{\mathrm{i},d}) + \hat{\beta}_4 PPV_{\mathrm{i},d} + \hat{\beta}_5 ToS_{\mathrm{i},d} + \hat{\beta}_6 BP_{\mathrm{i},d}$$
$$+ \hat{\beta}_7 I(AR_{\mathrm{i},d} \text{ missing}) + \hat{\beta}_8 I(VP_{\mathrm{i},d} \text{ missing}) + \hat{\beta}_9 I(PP_{\mathrm{i},d} \text{ missing})$$
$$+ \hat{\beta}_{10} I(PPV_{\mathrm{i},d} \text{ missing}) + \hat{\beta}_{11} I(ToS_{\mathrm{i},d} \text{ missing}) + \hat{\beta}_{12} I(BP_{\mathrm{i},d} \text{ missing}), \tag{C.6}$$

where $C_{\mathrm{i},m}$ is the centrality for firm i in the corresponding centrality month m (June or December of 2017, 2018, or 2019), and $\hat{C}_{\mathrm{i},d}$ is the predicted centrality for firm i on day d.

Equation C.5 produces an R^2 of 67%, suggesting that web traffic can explain a significant portion of centrality, but there remains a nontrivial network component to the measure that web traffic cannot account for. Nevertheless, predicted centrality provides an alternative measure of centrality that is substantially less subject to look-ahead bias. It also can be used to asses the extent to which look-ahead bias may be present in centrality when applied to the extended sample period. To do this, I first take the average predicted centrality for each firm in each month to get a monthly predicted centrality. I then sort firms into deciles based on monthly predicted centrality in June of 2017 and calculate average monthly predicted centrality for each decile in each month from 2014 to 2019.

The results are plotted in Figure C.1. If centrality varies significantly over time, and look-ahead bias significantly affects the results, then the relative ranking of centrality portfolios should be varying over time. However, there is little overall variation in average monthly predicted centrality for each decile, and especially in relation to other deciles. A decile's ranking for average monthly predicted centrality is unchanged over the entire period, with the only exceptions being the most peripheral firms in deciles 1 and 2. It appears that centrality has been fairly stable over the six year period, and applying centrality as measured in June of 2017 to prior months should induce relatively little look-ahead bias.

To further demonstrate this point, I repeat the returns analysis from Section 1.4.1 using monthly predicted centrality, and report the results in Panel A Table C.3. The spread portfolio returns are consistent with those presented in Panel A of Table A.5, with peripheral firms earning significantly more negative alphas than central firms. The somewhat smaller economic magnitude, as well as lower statistical significance, is likely attributable to the additional information provided by the network aspects of centrality not accounted for by web traffic variables. Although not reported for brevity, spread portfolios using monthly predicted centrality from July of 2017 to December of 2019 (the no-look-ahead bias period for centrality) show a similarly smaller economic magnitude and statistical significance relative to Panel B of Table A.5.

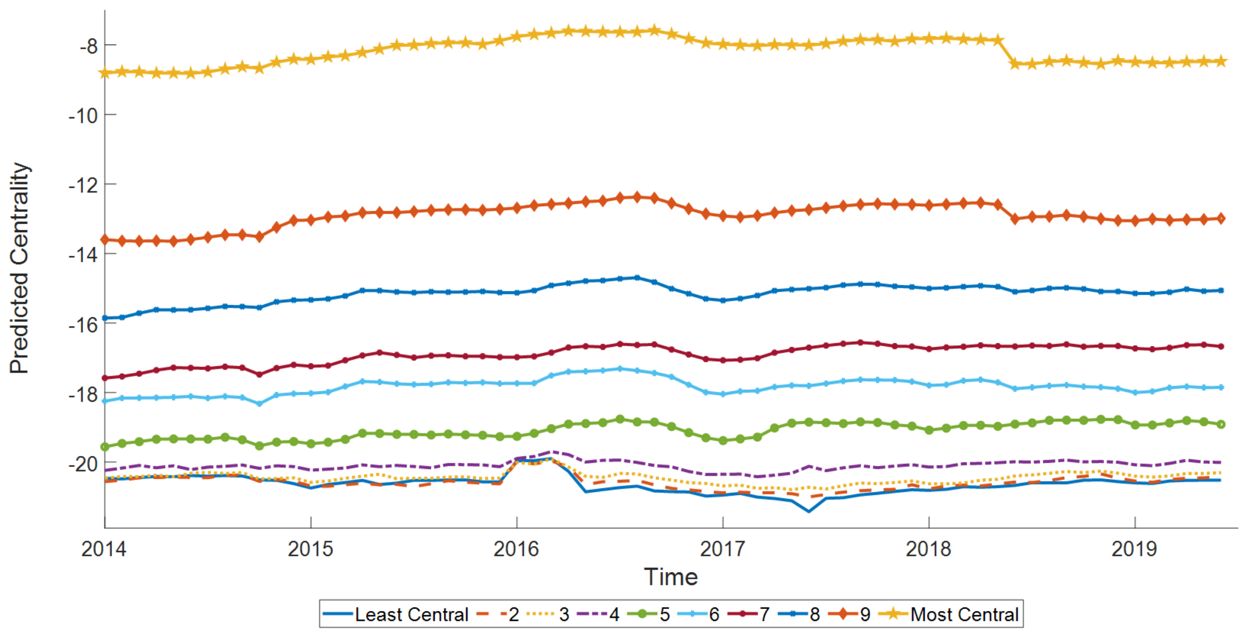


Figure C.1. Predicted Centrality Over Time

This figure plots monthly average predicted centrality for portfolios of firms sorted by predicted centrality in June of 2017. To calculate predicted centrality, centrality from the network of firm websites is regressed on several web traffic statistics to calibrate a model, which is then used to predict centrality each day from 2014-2019. The portfolios of firms are numbered by increasing centrality. The vertical axis is the predicted centrality for the portfolio, and the horizontal axis is the corresponding calendar month ranging from January 2014 to June 2019.

Table C.3. Return Robustness

This table reports robustness tests for the relation between centrality and stock returns. Panel A reports value-weighted returns for portfolios sorted into deciles by predicted centrality. Centrality is predicted using daily web traffic from 2014-2019. Panel B reports value-weighted returns for portfolios sorted into deciles by centrality after removing the bottom 20% of firms based on correlation between web traffic and sales. Raw Return reflects the average return for each portfolio. CAPM is the intercept of a regression of excess portfolio return on the market risk premium. FF 3-Factor is the intercept of a regression of excess portfolio returns on the factor model from Fama & French (1993). FF 3-Factor + Mom is the intercept of a regression of excess portfolio returns on the factor model from Carhart (1997). FF 5-Factor is the intercept of a regression of excess portfolio returns on the factor model from Fama & French (2015). 10-1 is the portfolio equivalent to buying portfolio 10 and selling portfolio 1. The t-statistic for each spread portfolio is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Panel A: Predicted Centrality

Centrality Portfolio	Raw Return	CAPM	FF 3-Factor	FF 3-Factor + Mom	FF 5-Factor
Least Central	0.39%	-0.77%	-0.44%	-0.39%	-0.38%
2	0.41%	-0.65%	-0.37%	-0.33%	-0.29%
3	0.63%	-0.43%	-0.13%	-0.10%	-0.05%
4	0.54%	-0.45%	-0.26%	-0.25%	-0.27%
5	0.39%	-0.61%	-0.40%	-0.35%	-0.38%
6	0.68%	-0.27%	-0.13%	-0.07%	-0.17%
7	0.66%	-0.28%	-0.22%	-0.17%	-0.22%
8	0.84%	-0.11%	-0.05%	-0.03%	-0.06%
9	0.83%	-0.02%	-0.06%	-0.07%	-0.04%
Most Central	1.20%	0.23%	0.14%	0.12%	0.12%
10-1	0.81%**	1.00%***	0.58%**	0.50%*	0.50%*
	(2.19)	(2.68)	(2.05)	(1.82)	(1.83)

Panel B: Excluding firms with low correlation between website visits and sales					
Centrality Portfolio	Raw Return	CAPM	FF 3-Factor	FF 3-Factor + Mom	FF 5-Factor
Least Central	0.16%	-0.60%	-0.59%	-0.58%	-0.59%
2	0.34%	-0.49%	-0.26%	-0.20%	-0.26%
3	0.26%	-0.48%	-0.42%	-0.39%	-0.43%
4	0.29%	-0.49%	-0.32%	-0.26%	-0.30%
5	0.38%	-0.36%	-0.26%	-0.24%	-0.23%
6	0.29%	-0.44%	-0.30%	-0.20%	-0.27%
7	0.74%	0.05%	0.07%	0.12%	0.10%
8	0.62%	-0.05%	0.03%	0.02%	0.02%
9	0.33%	-0.29%	-0.30%	-0.29%	-0.29%
Most Central	1.06%	0.32%	0.24%	0.21%	0.21%
10-1	0.90%***	0.92%***	0.83%***	0.79%***	0.80%***
	(3.08)	(3.07)	(2.80)	(2.66)	(2.69)

C.4.2 Further Robustness

The web traffic that creates overlap scores can come from a variety of sources, not all of which necessarily result in economic activity for the firm. I therefore repeat the returns analysis on a subset of firms, dropping those with the lowest correlation between web traffic and sales. To measure this correlation, I first obtain quarterly sales data from Compustat. I then calculate the average percent of all Internet users that visited the website per day for each firm in each quarter, which is provided by Alexa. Finally, I measure the correlation between web traffic and sales for each firm from 2014 to 2018. The subsample is formed by dropping the bottom 20% of firms based on this correlation,[3] and therefore consists of the firms for which web traffic has a stronger relationship with economic activity.

The results from the returns analysis for this subsample are reported in Panel B of Table C.3. They are remarkably similar to those from the full sample analysis, reported in Panel A of Table A.5. This suggests that the significant difference in returns between central and peripheral firms is not being driven by firms with substantial web traffic that is unrelated to their economic activity. Moreover, it supports the notion that consumers play a large part in determining centrality.

Centrality can also be correlated with a number of other variables, so I investigate whether one of these other variables can explain the relation between stock returns and centrality. In particular, I examine sales and market capitalization (size), both of which are closely related to centrality. To demonstrate the marginal information present in centrality, I first sort firms into quintiles based on sales (size), and then within each sales (size) quintile I sort firms into quintiles based on centrality. This produces 25 value-weighted dependent-sorted sales (size) and centrality portfolios, and the corresponding Fama-French 3-factor alphas are reported in Table C.4.

Panel A first sorts on sales and then on centrality. The marginal information in centrality is especially present in the extreme sales quintiles, although the centrality spread portfolio within all sales quintiles is positive and economically large. Panel B first sorts on size and then on centrality. Here, the centrality spread portfolio within all size quintiles is positive

[3]↑Similar results are obtained dropping the bottom 10% or bottom 40% of firms.

and statistically significant, again especially true for the smallest quintile of firms. In both panels, it seems that centrality is especially important for “small” firms (low sales, low market cap). Small firms that are central are able to survive, while the peripheral small firms suffer the most. Overall, it seems that centrality provides information beyond that of these closely related variables.

To further test the robustness of the relation between centrality and stock returns, I perform Fama & MacBeth (1973) regressions of monthly returns on log centrality, controlling for the same return determinants as in Section 1.6. Table C.5 reports the results of these regressions. Column (1) reports a regression without any control variables, and then variables are individually added in the subsequent columns. Column (11) reports the regression with all control variables included. Standard errors are Newey & West (1987) adjusted for three lags and all independent variables have been standardized to facilitate comparison. Across all regressions, centrality remains a significant determinant of cross-sectional returns after controlling for other variables. It is also economically reasonable, with a one standard deviation change in centrality accounting for somewhere between 0.16% and 0.29% return per month. This result, combined with the double-sorted portfolios discussed above, suggest that the returns observed in centrality-sorted portfolios are not a function of other common return determinants.

Table C.4. Double-Sorted Portfolios

This table reports value-weighted returns for double-sorted portfolios. Panel A first sorts firms on Sales and then on centrality. Panel B first sorts firms on Mkt Cap and then on centrality. Each variable and its data source is described in Appendix C.1. The returns reported in all three panels are intercepts from a regression of excess portfolio returns on the factor model from Fama & French (1993). 5-1 is the portfolio equivalent to buying the portfolio of most central firms and selling the portfolio of least central firms. The t-statistic for each spread portfolio is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

Panel A: Double-Sort by Sales and Centrality

	Lowest Sales	2	3	4	Highest Sales
Least Central	-1.48%	-0.33%	-0.30%	-0.35%	-0.44%
2	-1.15%	-0.39%	-0.19%	-0.41%	-0.08%
3	-1.24%	0.38%	-0.10%	0.00%	-0.18%
4	-0.13%	0.45%	0.07%	0.18%	-0.04%
Most Central	0.37%	0.33%	-0.02%	0.19%	0.25%
5-1	1.85%**	0.66%	0.28%	0.54%*	0.70%**
	(2.44)	(1.10)	(0.82)	(1.65)	(2.13)

Panel B: Double-Sort by Mkt Cap and Centrality

	Smallest	2	3	4	Largest
Least Central	-0.98%	-0.28%	-0.51%	-0.59%	-0.45%
2	-0.78%	-0.26%	-0.28%	-0.31%	-0.07%
3	-0.68%	-0.12%	0.11%	-0.05%	-0.14%
4	0.52%	0.48%	0.33%	-0.03%	-0.01%
Most Central	0.44%	0.34%	0.33%	-0.05%	0.24%
5-1	1.42%***	0.62%*	0.84%***	0.55%*	0.69%**
	(4.30)	(1.91)	(3.05)	(1.98)	(2.28)

Table C.5. Fama-MacBeth Regressions on Centrality

This table reports Fama & MacBeth (1973) regressions of returns on log centrality along with other control variables from 2014 to 2018. Each variable and its data source is described in Appendix C.1. Returns (t-1) is Returns in the previous month. Returns (t-7 to t-2) is the cumulative return from month t-7 to month t-2. Standard errors are Newey & West (1987) adjusted for three lags. The t-statistic for each coefficient is reported below in parentheses. Statistical significance at the 1%, 5%, and 10% levels is denoted by ***, **, and *.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Centrality	0.261**	0.230**	0.265**	0.285***	0.202*	0.253**	0.163*	0.264**	0.264**	0.240**	0.179**
	(2.38)	(2.46)	(2.59)	(3.07)	(1.88)	(2.34)	(1.80)	(2.30)	(2.44)	(2.30)	(2.08)
Mkt Cap		0.063									-0.114
		(0.57)									(-0.67)
M/B			-0.060								0.008
			(-0.39)								(0.06)
Ad Expense				-0.180**							-0.178**
				(-2.21)							(-2.66)
Ad Expense Missing				-0.440**							-0.374**
				(-2.10)							(-2.41)
Sales					0.097						0.239
					(0.72)						(1.22)
N Shareholders						0.032					-0.006
						(0.32)					(-0.08)
Volatility							-0.471***				-0.219
							(-2.88)				(-1.27)
Turnover								-0.431*			-0.161
								(-1.93)			(-0.69)
Returns (t-1)									-0.137		-0.139
									(-1.43)		(-1.32)
Returns (t-7 to t-2)										0.252*	0.204*
										(1.74)	(1.83)
Observations	169,145	169,145	162,292	169,145	164,847	150,114	169,132	168,074	167,412	169,139	138,032

C.4.3 Estimating Consumption Betas

Finally, it is also possible that central firms earn higher stock returns because they are more exposed to a risk factor. The results discussed in Section 1.4.1 control for many risk factors, but perhaps an omitted factor, such as consumption risk, could explain the results. Central firms tend to be firms that receive more consumer attention, and so may be more exposed to changes in aggregate consumption. This would result in central firms having larger consumption betas via the Consumption CAPM (Breeden, 1979), and thus higher stock returns to compensate investors for the increased risk. To test this possibility, I measure the consumption betas for portfolios of firms sorted by centrality. I calculate betas as the covariance between the cash-flow innovations of a portfolio and those of the market, following a procedure similar to Da & Warachka (2009). They find that this measure of cash-flow risk is closely related to more traditional measures of consumption risk, but is superior in explaining the cross-section of stock returns, for which traditional measures of consumption risk have notoriously poor performance (Mankiw & Shapiro, 1986; Breeden, Gibbons, & Litenberger, 1989; Cochrane, 1996).

Consumption, or cash-flow, betas are measured in a regression of the cash-flow innovations of a portfolio of firms on the cash-flow innovations of the market following Da & Warachka (2009). Cash-flow innovations of a portfolio are captured using revisions in analyst forecasts. IBES provides data on analyst forecasts, including each firm's consensus earnings forecasts for the current and subsequent fiscal year, updated monthly. Using the unadjusted summary file, the earnings-per-share (EPS) forecasts are split-adjusted using CRSP adjustment factors. I then form a time-weighted average of the current and subsequent fiscal year forecasts to obtain a 12-month-ahead earnings forecast for every firm each month,

$$\mathbf{E}_t(EPS_{\mathrm{i},t+12}) = w_{\mathrm{i},t}\mathbf{E}_t(EPS_{\mathrm{i},t+\tau_m}) + (1 - w_{\mathrm{i},t})\mathbf{E}_t(EPS_{\mathrm{i},t+\tau_m+12}), \tag{C.7}$$

where $EPS_{\mathrm{i},t}$ is the forecast for the earnings-per-share announced by firm i in month t, τ_m is the time in months until the current fiscal-year earnings announcement, and $w_{\mathrm{i},t} = \tau_d/365$ where τ_d is the time in days until the current fiscal-year earnings announcement. As the current fiscal-year earnings announcement approaches, Equation C.7 reduces the weight of the current fiscal-year forecast and increases the weight of the subsequent fiscal-year forecast.

To capture revisions in forecasts, I use a similar weighting of the monthly changes in the current and subsequent fiscal year forecasts,

$$\begin{aligned}\Delta\mathbf{E}_t(EPS_{\mathrm{i},t+12}) = w_{\mathrm{i},t}\Big(&\mathbf{E}_t(EPS_{\mathrm{i},t+\tau_m}) - \mathbf{E}_{t-1}(EPS_{\mathrm{i},t+\tau_m})\Big)\\ &+ (1-w_{\mathrm{i},t})\Big(\mathbf{E}_t(EPS_{\mathrm{i},t+\tau_m+12}) - \mathbf{E}_{t-1}(EPS_{\mathrm{i},t+\tau_m+12})\Big).\end{aligned} \tag{C.8}$$

By weighting the change in forecasts, rather than measuring the change in weighted forecasts, $\Delta\mathbf{E}_t(EPS_{\mathrm{i},t+12})$ will be primarily driven by revisions in expectations as opposed to changes in weights. The weighted EPS measures are then multiplied by the number of shares outstanding to produce measures of expected earnings, $\mathbf{E}_t(Earn_{\mathrm{i},t+12})$ and $\Delta\mathbf{E}_t(Earn_{\mathrm{i},t+12})$.

Next, I sort firms into deciles based on centrality and aggregate Equations C.7 and C.8 by deciles, as well as the overall market, each month. The portfolio-level forecast revision is then scaled by the earnings forecast in the previous month to control for differences in average firm size across portfolios. Aggregating earnings to a portfolio level significantly reduces the likelihood of a portfolio having negative earnings, which could give the wrong sign to the scaled-earnings revision. Indeed, no portfolio has negative aggregate earnings over the sample period. The final consumption beta for each portfolio, β_p, is measured in a monthly regression of the portfolio-level quantity on the market level quantity from 2014 to 2018,

$$\frac{\Delta\mathbf{E}_t(Earn_{p,t+12})}{\mathbf{E}_{t-1}(Earn_{p,t+12})} = \alpha_p + \beta_p\left(\frac{\Delta\mathbf{E}_t(Earn_{m,t+12})}{\mathbf{E}_{t-1}(Earn_{m,t+12})}\right) + \epsilon_{p,t}. \tag{C.9}$$

Table C.6 reports the consumption betas for each centrality portfolio. The portfolio of the most central firms has a cash-flow beta of 0.72 over this period, which is considerably smaller than the portfolios of the least central firms, which have betas ranging from 1.15 to 2.60. In general, it does not appear as if the cash flows of central firms are more exposed to fluctuations in aggregate cash flows, and so should not be receiving a premium for their consumption risk. This result may be intuitive when considering the types of firms that are central. These tend to be large firms, such as Walmart and Target, which have established themselves as the go-to option for consumers on a wide range of products. The smaller and more specialized firms that are particularly subject to fluctuations in consumption tend to be peripheral, as their size draws fewer website visitors and specialization means less overlap with other firm websites.

Table C.6. Consumption Betas

This table investigates the potential for consumption risk to explain the relation between centrality and stock returns. Firms are sorted in decile portfolios based on centrality. Consumption Beta is calculated in a regression of the cash-flow innovations of the portfolio of firms on the cash-flow innovations of the market.

Centrality Portfolio	Consumption Beta
Least Central	1.15
2	1.46
3	2.60
4	1.67
5	0.98
6	1.41
7	0.74
8	1.02
9	0.84
Most Central	0.72

D. SUPPLEMENTARY CHAPTER 2 MATERIAL

D.1 Data Appendix

Table D.1. Chapter 2 Variables

Description of variables for Chapter 2 and data sources.

Variable Name	Variable Description (Data Source)
1 Month Return	Stock return in the month after the IPO is completed (Thomson ONE)
3 Month Return	Stock return over the three months after the IPO is completed (Thomson ONE)
6 Month Return	Stock return over the six months after the IPO is completed (Thomson ONE)
Age	Year of observation minus year founded (Capital IQ)
Asset Turnover	Total Assets divided by Total Revenue (Capital IQ)
Avg Peer Age	Average *Age* of firms in the web-based peer group
Centrality	Measured within the network of web-based peers, see Section 2.3 (Alexa)
Cluster Coefficient	Measured within the network of web-based peers, see Section 2.3 (Alexa)
Enterprise Value	Market Capitalization plus Long-term Debt, divided by Sales (Compustat)
IPO Dummy	Dummy variable that equals one in the month the firm completes an IPO (Thomson ONE)
Leverage	Long-term Debt divided by Total Equity (Capital IQ)
Median Analyst Estimate	Median analyst long-term growth forecast (IBES)
N Peer IPOs	Number of peer firms that completed an IPO in the prior 12 months (Thomson ONE)
N Peers	Number of firms in the web-based peer group (Alexa)
Offer Principal	Number of shares offered at IPO times the Offer Price (Thomson ONE)

Overlap Score	Measures the extent to which two website audiences overlap, see Section 2.1 (Alexa)
PE Backed	Dummy variable that equals one if the firm is funded by private equity (Thomson ONE)
Peer 1 Month Return	Average *1 Month Return* of peer firms that went public in the prior 12 months
Peer 3 Month Return	Average *3 Month Return* of peer firms that went public in the prior 12 months
Peer 6 Month Return	Average *6 Month Return* of peer firms that went public in the prior 12 months
Peer Fees/Principal	Average IPO fees divided by *Offer Principal* of peer firms that went public in the prior 12 months (Thomson ONE)
Peer IPO %	Percent of firms in the web-based peer group that completed an IPO in the prior 12 months (Thomson ONE)
Peer IPO Dummy	Dummy variable that equals one if a firm in the web-based peer group completed an IPO in the prior 12 months (Thomson ONE)
Peer Offer Principal	Average *Offer Principal* of peer firms that went public in the prior 12 months
Peer PE Backed	Average *PE Backed* of peer firms that went public in the prior 12 months
Peer Venture Backed	Average *Venture Backed* of peer firms that went public in the prior 12 months
Price to Book	Market Capitalization divided by Common Equity (Compustat)
Price to Earnings	Market Capitalization divided by Income Before Extraordinary Items (Compustat)
Profit Margin	Net Income divided by Total Revenue (Capital IQ)
Return on Equity	Net Income divided by Common Equity (Capital IQ)
Return on Net Op Assets	Net Income divided by Total Assets (Capital IQ)
Returns	Stock Returns (CRSP)
Sales Growth	Percent growth in Total Revenue over the subsequent 12 months (Capital IQ)

Venture Backed	Dummy variable that equals one if the firm is funded by venture capital (Thomson ONE)

VITA

Education

AUGUST 2021 Ph.D. in FINANCE
Purdue University
Krannert School of Management
West Lafayette, IN

MAY 2019 M.S. in ECONOMICS
Purdue University
Krannert School of Management
West Lafayette, IN

MAY 2014 B.A. in PHYSICS
Reed College
Portland, OR
Thesis: "Spectroscopic Characterization of Water Adsorption on Calcium Carbonate and its Astrophysical Implications"

Research

- "Market Dominance in the Digital Age: Online Feedback Loops and Rising Industry Concentration."
- "Web-Based Peers: Peer Groups for Public and Private Firms."
- "Exposing The Revolving Door in Executive Branch Agencies," with Mara Faccio.
- "Expanding Horizons: The Effect of Information Access on Geographically Biased Investing," with Huseyin Gulen.
- "What are the Returns to Long-Term Buy-and-Hold Options?" with Douglas Emery and John Finnerty.

Other Publication

- Emery, L. P. and Powell, J. W. "The hydration dependence of $CaCO_3$ absorption lines in the Far IR," *Bulletin of the American Astronomical Society*, 46 (2014).

Teaching

Purdue University

– MGMT 304: Introduction to Financial Management — Spring 2020

– MGMT 411: Investments Management — Fall 2018, Spring 2021

Professional Service

- Journal referee
  - *Journal of Banking & Finance*
- Discussant
  - 2021 Midwest Financial Association Meeting (Chicago, IL)
  - 2020 European Finance Association Meeting (Helsinki, Finland)
  - 2019 Midwest Financial Association Meeting (Chicago, IL)
  - 2018 Financial Management Association Meeting (San Diego, CA)

Awards and Fellowships

- Krannert Certificate for Distinguished Teaching (Fall 2018, Spring 2020, Spring 2021)
- Best in Session - KDSA Research Symposium (2017, 2018, 2019)
- Blake Family Fund for Ethics, Governance, and Leadership Research Award (2018)
- Purdue Research Foundation Research Grant (Spring and Summer 2018)
- Ross Fellowship - Purdue University (Academic Year 2015-2016)
- Student Summer Research Grant - Reed College (Summer 2014)
- Excellence in Scholarship - Reed College (Academic Year 2013-2014)